Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.fw001

State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization

Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.fw001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

ACS SYMPOSIUM SERIES **1202**

State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization

Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques

John E. Schiel, Editor
National Institute of Standards and Technology
Gaithersburg, Maryland

Darryl L. Davis, Editor
Janssen Research and Development, LLC
Spring House, Pennsylvania

Oleg V. Borisov, Editor
Novavax, Inc.
Gaithersburg, Maryland

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.fw001

American Chemical Society, Washington, DC

Distributed in print by Oxford University Press

Library of Congress Cataloging-in-Publication Data

State-of-the-art and emerging technologies for therapeutic monoclonal antibody characterization / John E. Schiel, editor, National Institute of Standards and Technology, Gaithersburg, Maryland, Darryl L. Davis, editor, Janssen Research and Development, LLC, Spring House, Pennsylvania, Oleg V. Borisov, editor, Novavax, Inc., Gaithersburg, Maryland.

volumes cm. -- (ACS symposium series ; 1202)

Includes bibliographical references and index.

Contents: v. 3. defining the next generation of analytical and biophysical techniques

ISBN 978-0-8412-3031-6 (v.3)

1. Monoclonal antibodies. 2. Immunoglobulins--Therapeutic use. I. Schiel, John E., editor. II. Davis, Darryl L., editor. III. Borisov, Oleg V., editor.

QR186.85.S73 2014

616.07′98--dc23

2014040141

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48n1984.

PRINTED IN THE UNITED STATES OF AMERICA

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.fw001

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.fw001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.pr001

Preface

The line between where we are, and where we are going often blur. Development of novel analytical and biophysical technology are described well by this notion, as advances evolve in real time. Definition of "emerging technology", however, is often associated with a continuous uptick in industry acceptance. This may include promising modifications, or in some cases drastic accelerations, of state-of-the-art technology. The following volume of the book series is titled "Defining the Next Generation of Analytical and Biophysical Techniques" and contains 15 original chapters, authored by scientists from the biotechnology industry, academia, government agencies, and instrument-manufacturing firms that span method, technology, and informatics platforms. This volume describes novel and emerging analytical technologies for analysis of proteins with the emphasis on technologies aimed to address characterization "knowledge gaps" and/or improve our ability to measure specified attributes with improved selectivity, sensitivity, resolution, and throughput.

Higher order structure of proteins is a recognized important attribute of mAbs, with potential implications on stability, safety, and biological function of these large molecules. X-ray crystallography, NMR, hydrogen-deuterium exchange mass spectrometry (Chapter 2) and covalent labeling techniques (Chapter 3) are described in light of their application to examine higher order structure of mAbs. Ion mobility mass spectrometry, in Chapter 4, provides structural information by examining the collisional cross-sections of proteins in a gas phase under native ionization conditions, the information being particularly useful for comparability investigations, including development of biosimilars. Chapter 5 summarizes the current knowledge on the nature of protein aggregation (at nanometer-sized scale) of mAb formulations. This chapter further emphasizes the need for more sophisticated and high-resolution techniques to replace conventional lower resolution biophysical approaches for probing structure and molecular interactions. Chapter 6 introduces a novel tool to study protein aggregation simultaneously under multiple conditions by light scattering to enable expedited, controlled, and reliable formulation screening. Chapter 7 discusses specifics of applications of modern bioinformatics tools for the analysis of biotherapeutic proteins, an issue that has been largely underrepresented in the literature. In this regard, Chapter 14 continues the discussion by introducing several new software tools for the analyzing peptide mapping data and enabling trending attributes by comparing multiple data sets. Chapter 8 describes newer nucleic acid-based polymerase chain reaction (PCR) methods for the detection of adventitious agents during biopharmaceutical manufacturing. Microfluidic technologies such as lab-on-a-chip and high-performance liquid chromatography (HPLC)-chip mass

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophy
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

spectrometry tools, in Chapter 9, simplify integration of multiple steps, enabling higher throughput and the ease of use of complex analytical protocols. Analysis of large proteins, such as intact IgG, by state-of-the-art mass spectrometry, with the emphasis on extracting useful sequence information from the top-down fragmentation data, are presented in Chapter 10 and Chapter 11, respectively, using ESI Orbitrap and MALDI mass spectrometry technologies. Automation of manual processes of sample extraction, cleaning, and preparation for analysis is described in Chapter 12, which targets the improvement of reliability, consistency, and throughput of analytical workflows. Chapter 13 describes novel approaches for identification and quantitation of HCPs in biotherapeutic products.

The compilation of data and willingness of scientists throughout the biopharmaceutical industry to share their most recent innovations in this volume is a testament to the collaborative nature and interest in furthering a mission to quality therapies. At the time of the first mAb approved for human use, it was unthinkable that one day an image of a single mAb molecule might be attainable. Such astonishing developments have now become a reality, and the excitement only continues to grow. Many of novel and exciting technologies are rapidly advancing and demonstrate that as a village, we will succeed in attaining an even higher level of product characterization.

John E. Schiel
Research Chemist
Biomolecular Measurement Division
National Institute of Standards and Technology
Gaithersburg, Maryland 20899, United States
john.schiel@nist.gov (e-mail)

Darryl L. Davis
Associate Scientific Director
Janssen Research and Development, LLC
Spring House, Pennsyvania 19002, United States
DDavis14@its.jnj.com (e-mail)

Oleg V. Borisov
Associate Director
Novavax, Inc.
Gaithersburg, Maryland 20878, United States
oborisov@novavax.com (e-mail)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.pr001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Editors' Biographies

John E. Schiel

Dr. John E. Schiel received his B.S. (2004) and Ph.D. (2009) in Chemistry from the University of Nebraska-Lincoln, and is currently a research chemist in the NIST Biomolecular Measurement Division. He is leading the LC- and MS-based biomanufacturing research efforts at NIST; developing a suite of fundamental measurement science, standards, and reference data to enable more accurate and confident characterization of product quality attributes. Dr. Schiel is also the technical project coordinator for the recombinant IgG1κ NIST monoclonal antibody Reference Material (NISTmAb) program. He is an author of over 20 publications and recipient of numerous awards, including the ACS Division of Analytical Chemistry Fellowship, *Bioanalysis* Young Investigator Award, and UNL Early Achiever Award.

Darryl L. Davis

Dr. Darryl L. Davis holds a doctorate in Medicinal Chemistry from the Philadelphia College of Pharmacy and Science. His thesis focused on the use of MS in the characterization and quantitation of peptide phosphorylation. He started his career at J&J as a COSAT intern using MS to characterize the glycan linkages found on Remicade. Upon receiving his doctorate he accepted a full-time position within the Bioanalytical Characterization group at Centocor, a J&J company. Since joining J&J he has held a wide variety of responsibilities including starting and leading several sub-groups, analytical CMC lead, member of CDTs, member of technology development teams for alternative production platforms and new technology and innovation lead within analytical development. He has won several innovation awards within J&J for his work on automation and high-throughput analysis which continues to be a current focus. Currently he leads an analytical group within the discovery organization at Janssen R&D.

Oleg V. Borisov

Dr. Oleg V. Borisov earned a B.S. degree (with honors) in Chemistry at Moscow State University (1992), and received his Ph.D. in Chemistry from Wayne State University (1997), after which he completed his post-doctoral studies at Lawrence Berkeley National Laboratories (2000) and Pacific Northwest National Laboratories (2001). His background includes experience with analytical methods for characterization of biotherapeutic proteins and vaccine products, with emphasis on liquid chromatography and mass spectrometry methods.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Dr. Borisov held positions at Genentech and Amgen with responsibilities that included protein characterization, testing improvement, leading innovation and CMC strategy teams. He is currently a Director at Novavax, Inc., developing methods and strategies for analysis and characterization of recombinant vaccines, based on nano- and virus-like particle technologies. His credits include several student awards, a book chapter, and over 25 scientific publications.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Chapter 1

Trends and Drivers for the Development of Next-Generation Biotherapeutic Characterization Tools

Oleg V. Borisov,[*,1] John E. Schiel,[2] and Darryl Davis[3]

[1]Novavax, Inc., 20 Firstfield Rd., Gaithersburg, Maryland 20878, United States
[2]Analytical Chemistry Division, National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, Maryland 20899, United States
[3]Janssen Research and Development, LLC, 1444 McKean Rd., Spring House, Pennsylvania 19477, United States
[*]E-mail: oborisov@novavax.com

Biotherapeutics are recognized as increasingly important modalities for treating human disease. Capitalizing on advances in modern science and clinical experience with biotherapeutics, the field is rapidly expanding in seemingly orthogonal directions targeting new and increasingly sophisticated therapies, such as bispecific and conjugated monoclonal antibody products, as well as making existing therapies more affordable via the establishment biosimilar and follow-on biologics pathways. Collectively, these trends amplify the increasing demand for improvement of existing analytical methodologies as well as the development of new tools to characterize these complex biological products in greater detail. Discussion in this introductory chapter is based on the polled opinions of researchers associated with the development and testing of biotherapeutic proteins. The aim of the survey was to capture a snapshot on current perspectives on the state-of-the-art analytical methods and the need for the development of emerging technologies to address unmet or under-met characterization needs for these products.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Unlike small molecule pharmaceutics, recombinant protein-based therapeutics are large, structurally dynamic, and inherently heterogeneous biological products that are manufactured by living organisms as an ensemble of related species. Over nearly three recent decades, the IgG class of monoclonal antibodies (mAbs) has become the largest modality of therapeutic proteins. The importance of mAbs is evident (Mechanisms of Action chapter/Volume 1, Chapter 2) as are the complexities of these large molecules (Heterogeneity chapter/Volume 1, Chapter 3) and difficulties in the analysis of critical quality attributes (QbD chapter/Volume 1, Chapter 5). Current state-of-the-art technologies have advanced to provide precision characterization and quality control; however, the desire for continuous innovation is fueled by the need for faster-to-market development as well as the increasing complexity of mAb-based therapeutics, including bispecifics, antibody–drug conjugates (ADCs), and combination therapies. We are also amidst a paradigm shift wherein analytical technologies are playing an increasing role in both originator and biosimilar molecule development. Alongside the ever-expanding scientific knowledge of systems biology and continuous improvements in manufacturing and testing capabilities, supported by research undertaken by drug manufacturers, instrument vendors, and academic and government institutions, come the regulatory requirements, driven by the need of world governments to protect their citizens.

Two major factors drive the development of analytical technologies for the characterization of biopharmaceuticals. On one hand, newly gained scientific knowledge or clinical evidence may identify a potential "knowledge gap"—one that challenges the ability and competency of existing analytical methods to answer a critical question. On the other hand, the emergence of new technologies often provides further insight on critical quality attributes. The maturation of a new technology can be a lengthy process, established via a collaborative network of scientists from academia, industry, vendor firms, and regulators, who come together to form a consortium that is established to evaluate and demonstrate the fit-for-purpose capabilities of the new technology. In a sense, industry has the tendency to self-regulate. New technology advances from an academic bench to measuring biopharmaceutical proteins under regulatory constraints upon reaching a tipping point when the benefit of employing the new technology outweighs the associated investment costs and risks. Thus, some technologies may wait for their "prime time" longer than others.

One example is detection methods for residual host cell protein (HCP) impurities in biotherapeutic formulations, for which a number of factors are currently fostering the application of new technologies. Immunological bioassays were developed and used at the dawn of the era of biotechnology for the detection and quantitation of contaminating HCPs (Process Impurities chapter/Volume 2, Chapter 9). At that time, HCP enzyme-linked immunosorbent assay (ELISA) was identified as the only available method to provide good coverage for all the potential contaminants at microgram per gram of product quantities (*1*). To date, ELISA-based methods are providing information on the levels of HCPs in biotherapeutic products for regulatory submissions. There is no defined regulatory limit on levels of HCP in biotherapeutic formulations; however, most biotechnology products reviewed by the Food and Drug Administration

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

(FDA) contain ELISA-based HCP levels of 1–100 μg/g of product (*2*), which over the years became a commonly accepted limit for HCPs in biotherapeutical products ((*3*), Process Impurities chapter/Volume 2, Chapter 9). Despite the unquestionable advantage of reporting the collective sum of immunoreactive proteins, the sensitivity and accuracy of ELISA method depends on the quality of immunoreagents, customized for a particular manufacturing process. Recently, however, the overall sophistication of analytical technologies, increased knowledge and evidence on the significance of HCPs with respect to safety and efficacy of biotherapeutics, and the emergence of biosimilar products have challenged ELISA-based methods. Biosimilar manufacturers have a limited ability to measure HCPs in reference products because immunoreagents used by innovators are not available. State-of-the-art analytical methods (e.g., mass spectrometry) often detect individual HCPs that may be missed by HCP ELISA for a number of reasons (LC-MS HCP chapter/Volume 3, Chapter 13). A movement to incorporate these emerging technologies for use as research tools or in regulatory submissions has accelerated as experience with the use of biotherapeutic proteins in humans has increased, and new evidence has emerged linking HCPs to potential immunogenic reactions to the biotherapeutic product, leading to an increased regulatory concern (*3*). Together, these factors foster the development of new technologies. Liquid chromatography-mass spectrometry (LC-MS) methods, largely adopted from mass spectrometry-based proteomics applications and catalyzed by advances in bioinformatics and the availability of genomic data, are gaining acceptance for the identification and quantitation of individual HCPs (LC-MS HCP chapter/Volume 3, Chapter 13). In our opinion, LC-MS has a strong potential to outperform HCP ELISA because it provides information on levels and identities of HCP in biotherapeutic products at high resolution and without the need for using product-specific immunoreagents. We predict that LC-MS-based methods may eventually become the new standard for reporting HCPs in biotherapeutic products, or at the very least provide increased confidence in the suitability of a given immunoreactive method.

This book series is motivated by the desire that we and others share to provide a public forum by which the vast experience on characterization of mAbs can be critically discussed and continue the scientific dialogue on the state of the analytical technologies that support the development of these products. We believe that wide availability of a common IgG material, characterized by the collective effort of multiple industrial, government, and academic institutions, leading to a well-characterized Reference Material from the National Institute of Standards and Technology (NIST) for this important class of biotherapeutics, can serve as the common ground for this dialogue. In our opinion, this book series is a starting point in this journey. The goal is to promote collaboration and provide a baseline knowledge on the NISTmAb IgG1 molecule to researchers spanning established manufacturers and start-up companies that are currently establishing their characterization toolkit portfolio, as well as fundamental researchers who are working on the development of new technologies that are targeted to address unmet analytical needs.

During the preparation of this book series, we polled researchers associated with the development and testing of biotherapeutic proteins on their current

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

perspectives on the state-of-the-art analytical methods and the need for the development of emerging technologies to address unmet or under-met characterization needs for these products. An anonymous, nonscientific survey was designed asking participants to rank predefined categories and development areas by their role and significance in product characterization and general laboratory operation. The survey was completed by 51 participants, who provided feedback on the following topics. It should be noted that this discussion is based on an indiscriminate collection of opinions and no adjustments were made to compensate for the individual specialties of the participants.

Q1. With respect to the analysis pipeline and laboratory operation, which areas are in need of additional development of emerging technologies, based on your best understanding of the Lab-of-the-Future concept?

Categories related to data collection, processing, handling, collation, and storage were identified as areas requiring the most development. "Laboratory automation and robotics" and "instrumental platform compatibility" categories were regarded as requiring substantial development. In contrast, the "general laboratory layout and ergonomics" category received the lowest ratings (Figure 1).

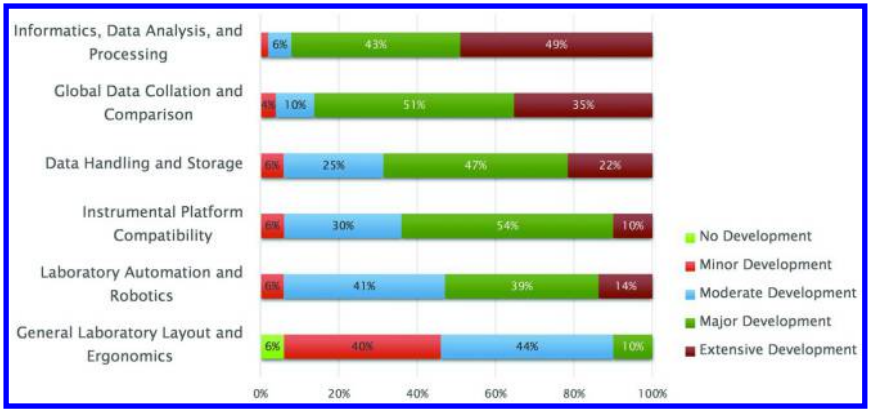

Figure 1. Responses to Q1. (see color insert)

Among other areas requiring further development, respondents named workflow and business intelligence, establishing effective management, and dissemination of gained knowledge. High-throughput technologies and the development of analytical tools that can be directly interfaced with manufacturing process equipment for real-time testing are other areas proposed by the survey participants.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Responses to this question highlight likely trends in the Lab-of-the-Future concept. On one hand, we see a strong need for high-throughput and real-time testing methods that would be targeted to expedite the decision-making process during the development, optimization, and execution of manufacturing runs and increase the breadth of knowledge about the process. On the other hand, modern instruments generate enormous amount of data, which requires storage, proper cataloguing, and processing. Raw data, however, arguably offers low value unless it can be processed (or re-processed) to extract useful information that can be reported in a format convenient for interpretation. The role of informatics tools will undoubtedly increase in the future. Innovative informatics technologies, in our opinion, will not only improve processing speed, availability, and dissemination of large-scale data but will enable the establishment of intelligent databases of knowledge, providing information on the cross-talk between product attributes of a specific molecule or extracting important trends for a particular quality attribute from multiple projects. With enormous amounts of data generated by modern instrumentation and with ever-changing and overlapping timelines, scientists are often limited in their ability to spend enough time on proper analysis of data. A well-catalogued repository of data, combined with the ability to reprocess the data as informatics tools develop, may one day help to inform analysis workflows, yielding the most informative data on the time scale of industrial development.

Participants of the survey also noted that most of the current bioinformatics tools are brought in from adjacent fields and academic research, where they fit slightly different purposes or have limited application for biotechnology tasks. In that regard, further development of bioinformatics tools designed for and targeted to address biotechnology approaches should continue to gain significant attention for future development. Among these software approaches, we see the importance of the development of tools predicting manufacturability properties of mAbs for development as biotherapeutics, such as viscosity, chemical and physical stability, shelf life, clearance, and major degradation pathways, based on in silico analysis of sequences of candidate molecules (*4*). Development of these tools would be supported by systemizing significant amounts of information accumulated over decades of the development of mAb-based biotherapeutics.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Q2. Based on your perspective of current state-of-the-art practices for characterization of biotherapeutics, please rate the following items as to their need for development of emerging technologies.

The rating scale used to analyze this and the following questions is based on a weighted average of the weights assigned to each answer on a 5-point rating scale, as indicated at the bottom of Figure 2.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

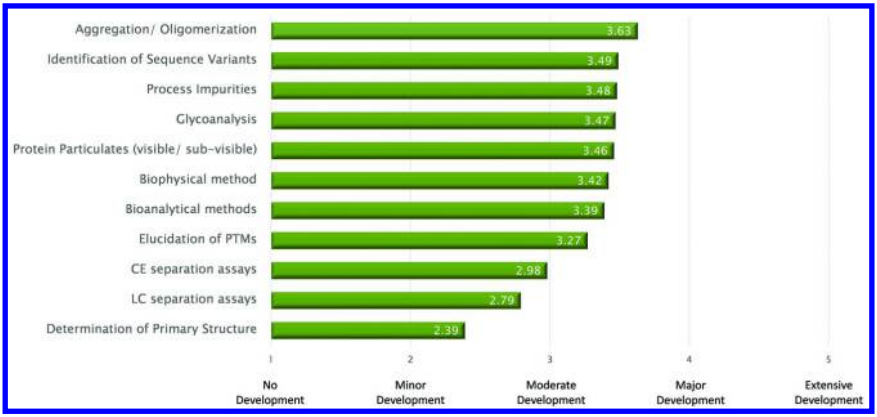

Figure 2. Responses to Q2. In the figure, the following abbreviations are used: CE (capillary electrophoresis), LC (liquid chromatography), and PTMs (post-translational modifications). (see color insert)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Oligomerization and aggregation is a recognized degradation mechanism of biotherapeutic proteins that has potential implications for the safety and efficacy of these products. In fact, aggregation has been identified as one of the areas of regulatory concern (Well Characterized chapter/Volume 1, Chapter 4). The survey highlighted the need for the development of emerging technologies to study protein aggregation. It is not surprising that two chapters in this volume are devoted to the mechanisms and technologies to study aggregation (SMSLS chapter/Volume 3, Chapter 6 and Aggregation chapter/Volume 3, Chapter 5).

Technologies for the identification and analysis of sequence variants, process impurities, glycans, protein visible and sub-visible particulates, post-translational modifications, as well as the improvement of bioanalytical methods, were identified as requiring above moderate development. At the same time, participants agreed that the existing state-of-the art technologies are adequate for the determination and confirmation of the primary structure (amino acid sequence) of proteins. We attribute this largely to the invention of soft ionization (electrospray ionization [ESI] and matrix-assisted laser desorption/ionization [MALDI]) methods for mass spectrometric analyses of biological macromolecules.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Q3. With respect to identification of protein modifications, which attributes require additional technological development for robust identification and quality control (Figure 3)?

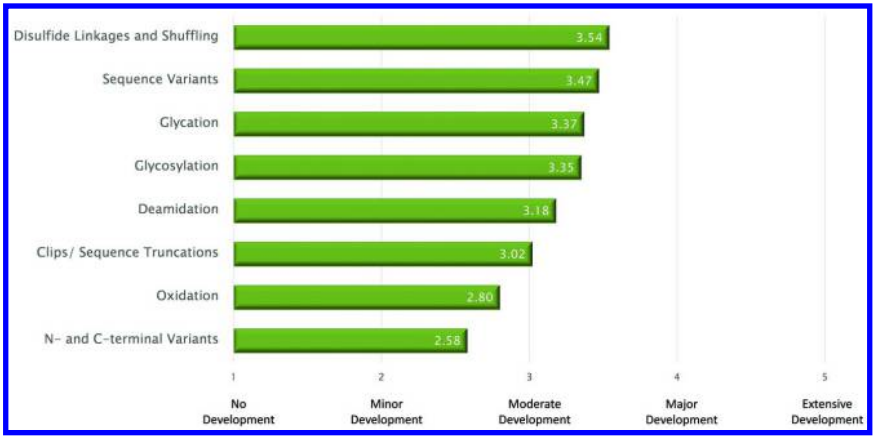

Figure 3. Responses to Q3. (see color insert)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Disulfide linkages (bonds) co-define higher order (tertiary) structure of proteins, which receives significant attention in the scientific community (*5*, *6*) and in recent years has been recognized as a focus area by regulators (Well Characterized chapter/Volume 1, Chapter 4). Peptide mapping with liquid chromatography-UV (LC-UV) and mass spectrometry is a technology frequently used to study disulfides. It often relies on a visual comparison of non-reduced and reduced maps of the same sample to assess changes in peak profiles following reduction with agents such as dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), or β-mercaptoethanol (BME). This process is, however, low throughput, requires two peptide maps, and prone to errors due to manual analysis, which is common. It is not surprising that robust technologies to elucidate disulfide linkages, their reduction–oxidation state, scrambling, and shuffling in biotherapeutic proteins are required to address this need.

The next four highest ranking categories of attribute in need of development of appropriate methods reflect challenges associated with their detailed and independent characterization. One common theme among analysis for sequence variants, glycation, glycosylation, and deamidation (including isomerization products of aspartic acid) is the need for improved workflows and informatics tools to readily identify and quantify these variants. For example, sequence variants may be in very low abundance and/or provide multiple potential isobaric combinations during identification. Glycosylation patterns of mammalian proteins are complex, often containing multiple glycan species with different functional roles and requiring rigorous and methodical structural characterization (Glycosylation chapter/Volume 2, Chapter 4). Deamidation/isomerization

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

analysis also suffer from difficulty in assignment due to the relatively low mass shift or even isobaric overlap in the case of isomerization as well as from the high potential for sample preparation artifacts. Each of these three analysis often require significant manual verification and orthogonal validation through forced degradation protocols and/or orthogonal techniques. It is therefore likely that continued development in targeted analysis of these modifications will continue in the coming years.

Answers to the following two questions are grouped to show trends in the methods for higher order structure determination.

Q4a. With respect to determination of higher order structure, please rate the following approaches for their current use in product characterization (Figure 4).

Q4b. With respect to determination of higher order structure, please rate the following approaches for their prospective impact on product characterization (Figure 4).

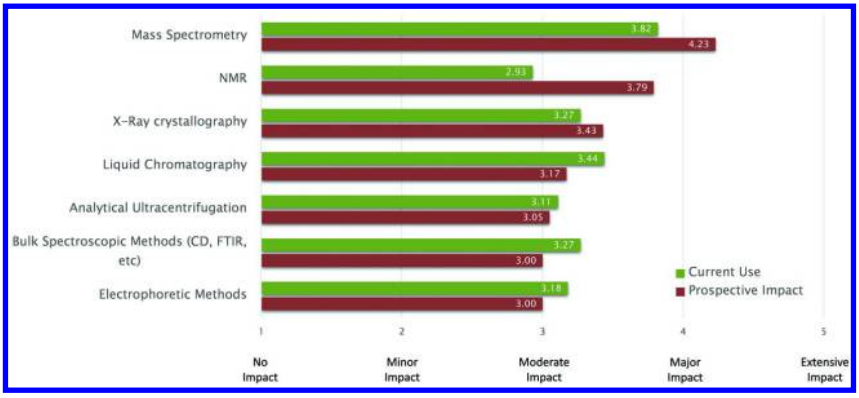

Figure 4. Responses to Q4a and Q4b. In the figure, the following abbreviations are used: CD (circular dichroism), FTIR (Fourier transform infrared), and NMR (nuclear magnetic resonance). (see color insert)

Higher order structure defines function of proteins and is an important quality attribute of biotherapeutics. The ICH Q6B guideline emphasizes that "for complex molecules, the physicochemical information may be extensive but unable to confirm the higher-order structure which, however, can be inferred from the biological activity" (*7*).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Liquid chromatography (size exclusion chromatography [SEC]), electrophoretic (sodium dodecyl sulfate [SDS] gels and capillary electrophoresis [CE]), and analytical ultracentrifugation methods are routinely used to characterize size variants of biotherapeutic proteins, which can be indicative of the higher order structure of proteins. Many biophysical methods, including bulk spectroscopic measurements (such as intrinsic fluorescence, Fourier transform infrared [FTIR], far- and near-UV circular dichroism measurements) and differential scanning calorimetry (DSC), are well established and widely used to characterize and compare higher order structure of proteins. Although results of these methods are often included in regulatory filings to describe higher order structure of biotherapeutic proteins, these methods have relatively low resolution and are often limited to providing domain-specific information at most and have a limited ability to differentiate between different species, which is an intrinsic property of any bulk method.

The survey correctly identifies the increasing demand for technologies that offer improved resolution, such as nuclear magnetic resonance (NMR), X-ray crystallography, and mass spectrometry-based methods. Applications of these methods to the characterization of higher order structure of biotherapeutic proteins are the subject of several chapters of this volume (Higher Order Structure chapter/Volume 3, Chapter 2; Covalent HOS chapter/Volume3, Chapter 3; Ion Mobility chapter/Volume 3, Chapter 4; and Aggregation chapter/Volume 3, Chapter 5). For example, the hydrogen-deuterium exchange method, based on measuring exchange rates of amide hydrogens of the protein backbone, is sensitive to changes in the local environment of these hydrogens, defined by the higher order structure of the protein. This method in combination with mass spectrometry detection is an emerging technology for probing the structure and dynamics of mAbs at a resolution approaching site-specific detail (*8*). The development of this technology in recent years has been a truly collective effort of academic institutions and biotechnology and instrument vendor firms, and it has been highly regarded by regulators as a potential technology to characterize protein conformational attributes.

Interestingly, NMR showed the largest difference in current and prospective utility among those techniques surveyed. NMR is a staple technique for small molecule structure confirmation and routinely is used in small molecule drug development. Its application to biopharmaceutical products has been limited in the past due to the limitations in resolution and sensitivity achievable with natural isotopic abundance of protein drugs. During the most recent decade, however, applications of NMR methods for the structural assessment of biotherapeutic proteins during discovery, production, comparability exercises, and quality assurance efforts have emerged, owing to significant improvements in hardware and methodologies for one-dimensional (1D) and two-dimensional (2D) NMR experiments (Covalent HOS chapter/Volume 3, Chapter 3) (*9*). For example, the Covalent HOS chapter/Volume 3, Chapter 3 demonstrates the feasibility of 2D NMR for spectral mapping of mAb domains to provide high-resolution structural information. The survey indicates a consensus in the field that NMR is at the cusp of the critical tipping point toward widespread implementation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Q5. With respect to mass spectrometry, please rate the following methods and their potential utility for their prospective impact on product characterization (Figure 5).

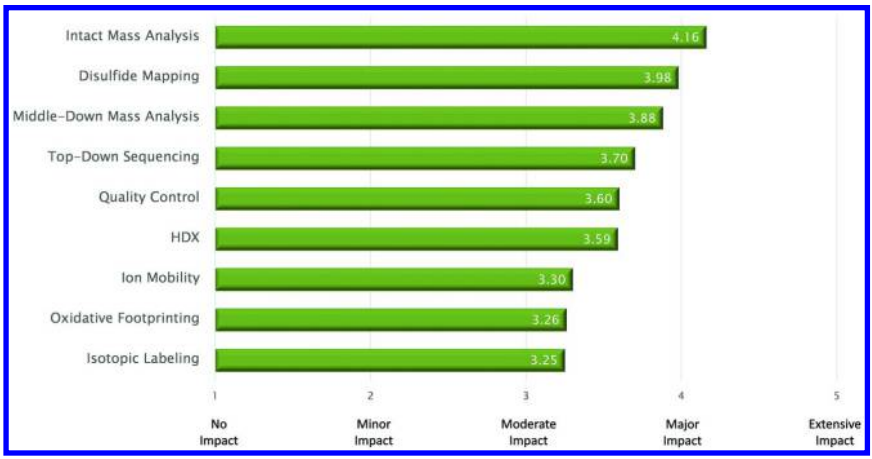

Figure 5. Responses to Q5. In the figure, the following abbreviation is used: HDX (hydrogen–deuterium exchange). (see color insert)

Mass spectrometry has become a key tool for the characterization of proteins. Over the last two decades, mass spectrometry has continued to mature to include numerous applications of this technology for the analysis of biopharmaceutical proteins—from measuring masses of peptides early on to approaches to fragment, detect cross-sections, and probe higher order structure of large intact proteins by modern state-of-the-art instruments. This success has arguably been driven by the successful development and use of biotherapeutics to treat human diseases. In the modern laboratory, mass spectrometry already is providing information on primary, secondary, tertiary, and quaternary structures of proteins. In the survey, we asked for the opinion on the prospective impact of mass spectrometry on the characterization of biotherapeutic proteins. Responses indicated that mass spectrometry methods, dealing with analysis of intact proteins and their fragments, including top- and middle-down methods, as well as methods for disulfide mapping, are expected to contribute to protein characterization the most. The speed and ability to probe the molecule with no sample pretreatment are likely a significant factor to the high rating of intact mass spectrometry. It is interesting to note that applications of mass spectrometry for quality control of biotherapeutics is gaining acceptance and received high ratings in the survey. In our opinion, the truly multi-attribute measurement capability of mass spectrometry will emerge as a Quality Control strategy for biotherapeutic proteins.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Q6. With respect to mass spectrometry instrument performance, please rate the following items as to their need for development of emerging technologies (Figure 6).

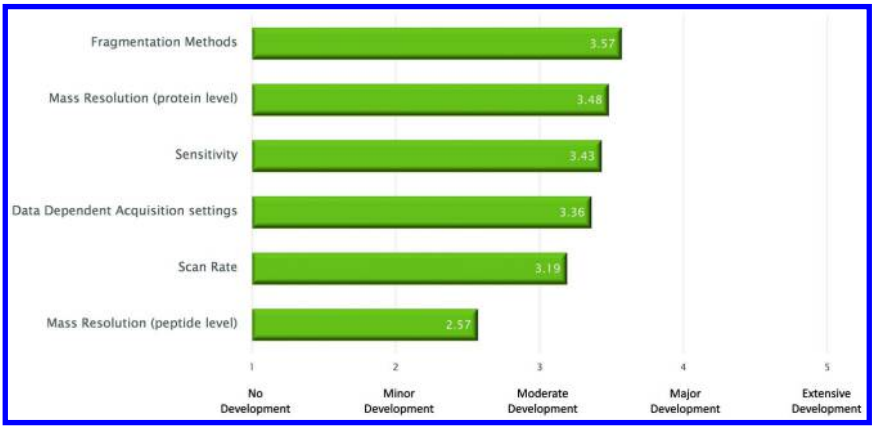

Figure 6. Responses to Q6. (see color insert)

This question polls opinion on selected performance characteristics of modern mass spectrometers requiring further development. Although modern mass spectrometers offer a number of choices to fragment ions of interest, the survey identified the importance of further improvement of these methods. Interrogation of analyte ions in the gas phase by means of fragmentation serves the purpose of obtaining "fingerprint" information on these ions, enabling their structural elucidation. Collision-induced dissociation (CID) methods historically have been used as primary technologies for providing structural data on peptide and protein molecules. In fact, major achievements in proteomics and peptide mapping of biotherapeutic proteins over the last two decades are due to the robust performance of CID methods. Depending on the translational energy supplied to the precursor ion during fragmentation, methods are divided into two regimes—low-energy CID with energies below 1 keV (available on ion traps and triple quadrupole-based instruments and including higher-energy CID [HCD] on Orbitrap instruments), and high energy CID methods energies above 1 keV (available on MALDI-time of flight [TOF]/TOF instruments). Despite the unquestionable advantages of CID methods due to the high speed, efficiency in the overall yield of fragment ions, and robust performance for a wide range of peptides and small proteins, certain factors, such as the incomplete and sequence-dependent fragmentation, overlapping ion series, and poor ability to detect labile modifications, limit application of these methods (*10*). More recently, electron capture dissociation (ECD) and electron transfer dissociation (ETD) methods have emerged as complimentary tools with unique advantages to study larger peptides and proteins and preserving labile modifications intact during the analysis. However, spectra produced by these mechanisms have a lower yield of

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

fragment ions, and spectra can be difficult to interpret due to the overlapping ion series, the presence of fragments in multiple radical and nonradical states, and somewhat less robust performance for a wider range of precursor ions compared to CID methods.

Other fragmentation methods have been developed and are available on different types of mass spectrometers, most notably infrared multiphoton dissociation (IRMPD) and blackbody infrared radiative dissociation (BIRD) on Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers, and surface-induced dissociation (SID) on FT-ICR and TOF instruments. New fragmentation methods, such as charge transfer dissociation (CTD) (*11*), continue to emerge as well.

Mass spectrometry analysis of biological samples, ranging from whole proteomes to a single-component biotherapeutic protein, is based primarily on tandem data that are processed automatically to match with in silico sequences in a protein or genomic database. The drawback of the database searching is that sequences are not always in the database due to a variety of reasons, including but not limited to alternative splice variants, frame shifts, wrong gene predictions, multiple modifications on the same peptide, and other transcription and translation errors (Sequence Variant chapter/Volume 2, Chapter 2). These factors may prevent the correct identification of experimental tandem mass spectrometry data. For example, during a typical analysis of a biotherapeutic IgG by peptide mapping with mass spectrometry, a large number of tandem spectra (~50%) did not match to a known peptide sequence (Bioinformatics chapter/Volume 3, Chapter 7). Thus, the ultimate goal of the fragmentation method, when applied to studies of peptides and proteins, is to provide sufficient sequence information to enable unambiguous identification of amino acid sequences and connectivity without the need for relying on the database for the virtual sequence. In other words, de novo sequencing is at the pinnacle of tandem mass spectrometry data analysis (*12*, *13*). Unfortunately, de novo sequencing has not been widely used for analysis of biotherapeutic proteins due to the relatively low accuracy of identifications, caused in part by the limitations of the tandem data. In fact, fragmentation mechanisms are the basis for de novo sequencing. The use of several existing fragmentation mechanisms, such as concurrent HCD and ETD on the same precursor, shows a promise for increasing sequence coverage by providing complementary fragment information (*14*). However, development of new and further improvement of existing fragmentation mechanisms will be needed to improve the way tandem mass spectrometry data is analyzed.

The resolution of mass spectrometers is expressed as $M/\Delta M$, where ΔM is the full width of the peak at half its maximum height (FWHM) and is an important parameter defining the ability of the instrument to resolve similar masses and affecting its mass measurement accuracy. TOF and Fourier transform, including FT-ICR and Orbitrap systems, are the two major platforms of modern mass spectrometers offering high resolution. Resolution of TOF instruments have increased by over 10-fold since late 1990, when the first TOFs became commercially available, and is now reaching 50,000 and even 80,000. Orbitrap technology, introduced in 2005 in a commercial instrument, now offers mass resolution of over 200,000 and up to 500,000 (at $m/z = 200$). In our opinion,

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the survey reflects such a significant improvement in resolution of modern instruments to accurately measure masses of peptides and small proteins with great isotopic resolution. However, the attainable resolution is still not sufficient to isotopically resolve charge states of larger proteins, such as IgG, and therefore, small mass shift variants may not be confidently identified. In this regard, the desire for higher resolution is reflected in the response to Question 5 in that further development of intact mass measurements would significantly benefit product characterization.

What might be additional goals of the race for high resolution? For example, deamidation is a known degradation pathway of biotherapeutic proteins and is an important quality attribute monitored during e development and stability. Asparagines are the primary amino acid residues affected by deamidation, converting to aspartic acids via an acid- or base-catalyzed processes, resulting in a mass shift of 1 Da. Since deamidation induces relatively small changes to the overall peptide's sequence, chromatographic separation of the amidated parent peptide and its deamidated form(s) can be difficult to achieve during LC-MS analysis of peptide maps. We illustrate the effect of instrument resolution on the example of resolving deamidated and amidated peptide variants from the single spectrum by TOF and Orbitrap-type instruments. First, the fundamental difference in resolution of the two platforms should be considered. Based on the detection principals, the resolution of TOF remains nearly unchanged across the mass range, whereas for Orbitraps, the resolution is inversely proportional to the square root of m/z (*15*). For Orbitraps, resolution is often reported at m/z 200. Thus, with a resolution of 240,000 at m/z 200, resolution at m/z 1200 is around 97,000.

For most peptides, the difference between the first and the second isotopes, is 1.0028(2) Da (dominated, respectively, by the mass difference of carbon-12 and carbon-13 isotopes). Deamidation results in a mass shift of 0.98402 Da, and the mass difference between the second isotope of the amidated peptide and the first isotope of the deamidated form is about 0.0188 Da, which defines the ΔM that the instrument needs to resolve in order to detect deamidation in a single spectrum. Figure 7 defines the requirements for instrument resolution (nominal resolution represents hypothetical instrument resolution at vendor-specified conditions) to detect deamidation as a function of mass of the amidated parent peptide, where red and green areas represent cases, respectively, of not resolved and resolved deamidation. The difference in the shapes of the curves between Orbitraps and TOFs is due to the differences in mass dependence of the resolution for these two instrument types. For example, TOF operating at a resolution of 50,000 resolves the deamidated monoisotopic peak from the second isotope of the parent amidated peptide with mass below 940.2 Da, whereas Orbitrap with resolution of 150,000 (at m/z 200) resolves the two forms of the peptide with mass below 1167.5 Da.

Historically, the analysis of proteins by mass spectrometry, including biotherapeutic products, was conducted using a so-called bottom-up methodology in which structural analysis is based on mass spectrometry fragmentation of proteolytic digests of intact proteins. In combination with LC separation of the peptide mixture, this method is highly sensitive for detection of low-level sequence variants and protein impurities. The method, however, can be labor

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

intensive and lengthy due to the sample preparation and separation requirements. Recently, top-down methods have gained popularity to probe sequences of intact proteins (or fragments in the middle-down version), owing to the improvements in resolution of modern mass spectrometers and the development of ECD and ETD fragmentation methods (Intact chapter/Volume 3, Chapter 10). In-source decay (ISD) technology available on MALDI instruments (MALDI-Sequencing chapter/Volume 3, Chapter 11) is another method to obtain top-down and middle-down information. In the current state of these technologies, top-down methods provide quick and robust information on C- and N-terminus regions of intact proteins, but more work is required to achieve higher coverage of protein sequences with fragment ions.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

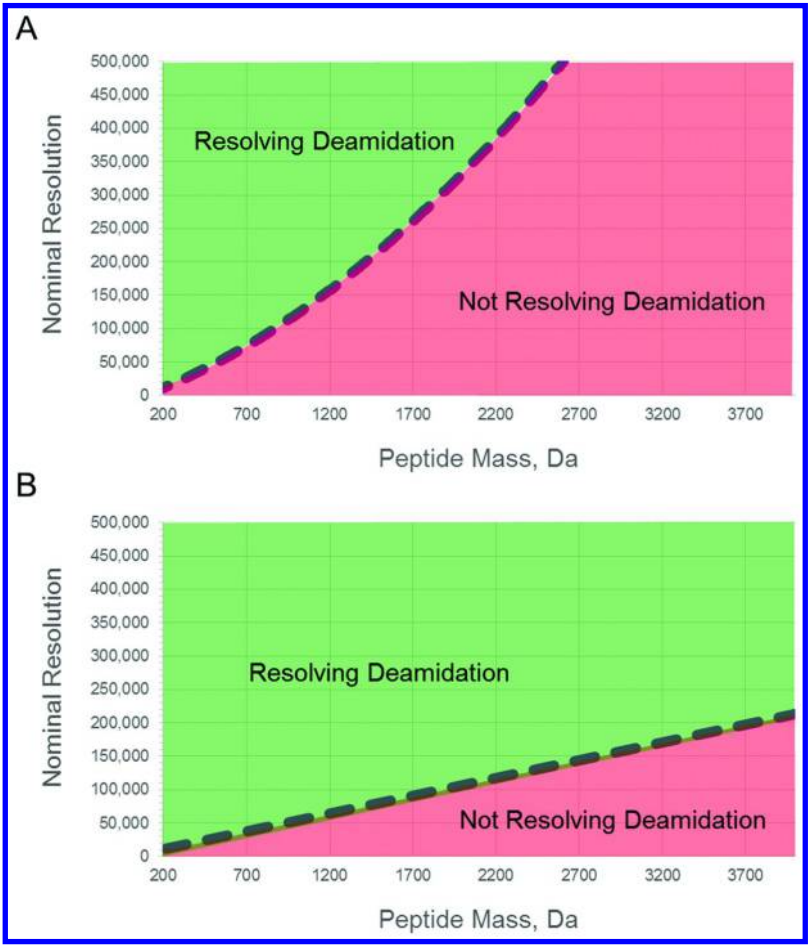

Figure 7. Ability of mass spectrometers based on Orbitrap (A) and time of flight (TOF) (B) technologies to resolve deamidation. (see color insert)

Q7. With respect to the application of mass spectrometry to process-related testing and control, how important do you feel the following areas are for additional development (Figure 8)?

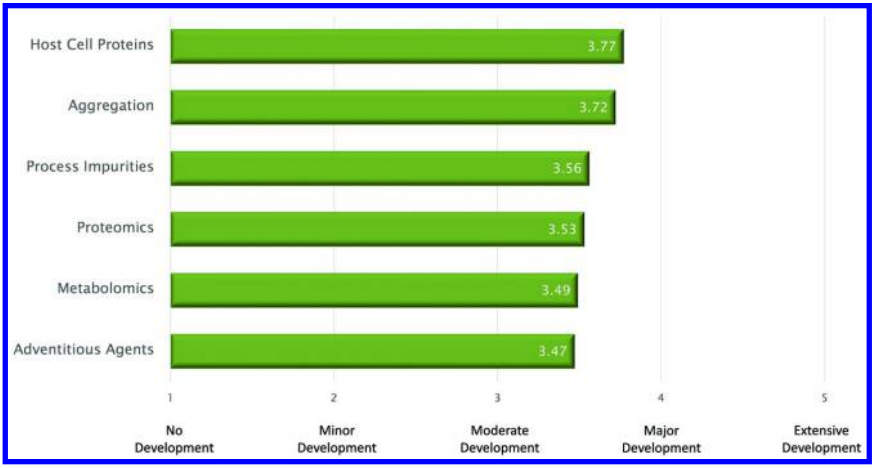

Figure 8. Responses to Q7. (see color insert)

The final question of the survey probed where mass spectrometry may provide the highest impact during process development. The results indicate the need for additional implementation and development of mass spectrometry applications expanding to the early stages of product development. This is not entirely surprising, considering that process analytics are the first-line technologies for obtaining information pertaining to product quality. Earlier and increased implementation of information-rich technologies such as, but absolutely not limited to, mass spectrometry would undoubtedly inform further process decisions relevant to a product during development and manufacturing. Process monitoring technologies are emerging as a predictive means for informing the quality by design of therapeutic proteins.

Summary

Collectively, the survey revealed a need for some level of development in multiple areas and is indicative of the desire of biopharmaceutical researchers to produce products of the highest quality attainable with the technology at hand. Clearly the simultaneous development of innovative solutions in each of these areas would be most beneficial to the community. Moreover, investments in the improvement and development of analytical tools would be capitalized by affording reduced requirements for clinical studies and, thus, faster times to the market.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

References

1. Jones, A. J. S. Sensitive detection and quantitation of protein contaminants in rDNA products. In *The Impact of Chemistry on Biotechnology*; Phillips, M., Ed.; ACS Symposium Series 362; American Chemical Society: Washington, DC, 1988; pp 193–201.
2. Champion, K.; Madden, H.; Dougherty, J.; Shacter, E. *BioProcess Int.* **2005**, *3*, 52–57.
3. Zhu-Shimoni, J.; Yu, C.; Nishihara, J.; Wong, R. M.; Gunawan, F.; Lin, M.; Krawitz, D.; Liu, P.; Sandoval, W.; Vandelaan, M. *Biotechnol. Bioeng.* **2014**, *111*, 2367–2379.
4. Sharma, V. K.; Patapoff, T. W.; Kabakoff, B.; Pai, S.; Hilario, E.; Zhang, B.; Li, C.; Borisov, O.; Kelley, R. F.; Chorny, I.; Zhou, J. Z.; Dill, K. A.; Swartz, T. E. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 18601–18606.
5. Jiang, Y.; Li, C.; Li, J.; Gabrielson, J. P.; Wen, J. *J. Pharm. Sci.* **2015**, *104*, 1533–1538.
6. Gabrielson, J. P.; Weiss, W. F., 4th *J. Pharm. Sci.* **2015**, *104*, 1240–1245.
7. *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products Q6B*; International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Geneva, Switzerland, March 1999. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q6B/Step4/Q6B_Guideline.pdf (accessed June 19, 2015).
8. Majumdar, R.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. *J. Pharm. Sci.* **2015**, *104*, 327–345.
9. Poppe, L.; Jordan, J. B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P. D. *Anal. Chem.* **2013**, *85*, 9623–9629.
10. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 5301–5303.
11. Hoffmann, W. D.; Jackson, G. P. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1939–1943.
12. Allmer, J. *Expert. Rev. Proteomics* **2011**, *8*, 645–657.
13. Medzihradszky, K. F.; Chalkley, R. *J. Mass Spectrom. Rev.* **2015**, *34*, 43–63.
14. Chi, H.; Chen, H.; He, K.; Wu, L.; Yang, B.; Sun, R. X.; Liu, J.; Zeng, W. F.; Song, C. Q.; He, S. M.; Dong, M. Q. *J Proteome Res.* **2013**, *12*, 615–625.
15. Makarov, M.; Denisov, E.; Lange, O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391–1396.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch001

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Chapter 2

Emerging Technologies To Assess the Higher Order Structure of Monoclonal Antibodies

J. P. Marino,[*,1] R. G. Brinson,[1] J. W. Hudgens,[1] J. E. Ladner,[1] D. T. Gallagher,[1] E. S. Gallagher,[1] L. W. Arbogast,[1] and R. Y.-C. Huang[1,2]

[1]Institute for Bioscience and Biotechnology Research, Biomolecular Measurement Division, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, Maryland 20850, United States
[2]Current Address: Bristol-Myers Squibb, Route 206 and Province Line Road, Princeton, New Jersey 08543, United States
[*]E-mail: john.marino@nist.gov

In contrast to small molecule therapeutics whose conformations can be absolutely defined by constitution and stereochemistry, biopharmaceuticals are distinguished by the requirement for folding into higher order structures (secondary, tertiary, and quaternary) for therapeutic function. Whereas proper folding of a protein biologic is critical for drug efficacy, misfolding may impact drug safety by eliciting unwanted immune or other off-target patient responses. In this chapter, we review current and emerging technologies for high-resolution characterization and fingerprinting of the structure and dynamics of monoclonal antibodies (mAbs) with a focus on techniques that can provide data at or near atomic resolution, such as X-ray crystallography, nuclear magnetic resonance (NMR) and hydrogen-deuterium exchange mass spectrometry (HDX-MS). Application of these techniques is illustrated using the NISTmAb.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Introduction

Antibody immunoglobulins (Igs) comprise several classes (IgA, IgD, IgE, IgG, and IgM), with the class defined by the type of heavy chain present (*1*). Further, two types of light chains are found in mammals, kappa (κ) and lambda (λ). In its four isoforms (IgG1, IgG2, IgG3, and IgG4), IgG comprises about 75% of serum Igs in humans and has to date been the dominant protein platform for the development of monoclonal antibody (mAb) drugs. As with all antibody Igs, the higher order structure of IgGs is based on extensive re-use of a single protein folding unit, the so called "immunoglobulin" domain (*2*, *3*). A single Ig domain consists of 7 or 9 beta strands that form a sandwich of two antiparallel sheets, with an intra-chain disulfide bridge linking the sheets for extra stability. An antibody consists of twelve of these Ig domains in four chains (2 domains per light chain and 4 per heavy chain). Interchain disulfides covalently unify the entire molecule. Figure 1 shows a schematic representation of an IgG antibody and a three-dimensional (3D) structural model of the NISTmAb built on the scaffold of Padlan (*4*). To generate this model, the Fab (antigen-binding fragment) from the Protein Data Bank (PDB) file 3IXT and the Fc (the so-called "crystallizable fragment") from PDB file 3AVE were superimposed on the scaffold with the appropriate residue changes made graphically in COOT (*5*). The hinge region was then built in an extended form as in Padlan's model. The second half of the mAb was generated using the models generated for one Fab and half of the Fc by applying twofold symmetry guided by 3AVE and the Padlan model. Lastly, the glycans were modeled from the 3AVE structure. The basic IgG fold confers two features essential for function: it enables linear concatemers of Ig domains with either flexible or tight linkers, and it allows the domain's interstrand loops to project outward, making them available for critical interactions, particularly antigen binding.

For all four polypeptides of an antibody (two heavy and two light chains), the N-terminal Ig domain supplies a set of hypervariable recognition loops at one end of the domain, alternating in sequence with the beta strands of the sandwich structure. This arrangement supports the loops with a framework that is also moderately variable, as the whole domain is produced through the unique gene-recombining and mutational mechanisms of B cells. The main complementarity determining regions (CDRs) are numbered L1, L2, and L3 for light chain and H1, H2, and H3 for heavy chain; a fourth heavy-chain loop, called CDR-H4, sometimes contacts antigen and is increasingly included in structural analyses (*6*). The closely associated N-terminal region of one light and one heavy chain, together called the Fv for variable fragment, form one complete antigen binding site. Together with the Fv, the second (and last) Ig-fold domain in the light chains, called C_L, and the second domain in the heavy chain, called C_H1, complete the Fab (Figure 2). Thus each Fab contains a complete light chain and the first half of a heavy chain. Between the first and second domain is an "elbow", and even though the light and heavy chains are closely coupled within each Fab, angular variations of 30 to 40 degrees are often observed in crystal structures of chemically identical elbows (*7*). Beyond the Fab, the heavy chain continues into the hinge, a flexible region that gives the whole antibody a large range of conformational

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

plasticity. Within the IgG class, the four subclasses (IgG1, IgG2, IgG3 and IgG4) are distinguished primarily by different amino acid compositions and lengths of this hinge region. For subclass IgG1, there are four interchain disulfides in this region, two linking heavy to light chains (close to the light chains' C-termini) and two that join the heavy chains to each other. The hinge also contains a key protease-sensitive site, where papain is routinely used to digest the heavy chains, liberating the two 50 kDa Fab fragments and leaving the C-terminal halves of the two heavy chains, which form a third 50 kDa Fc fragment (Figure 2). The 3rd and 4th domains of the two heavy chains pair up symmetrically to form the Fc fragment, whose sequence is relatively constant within the few defined classes.

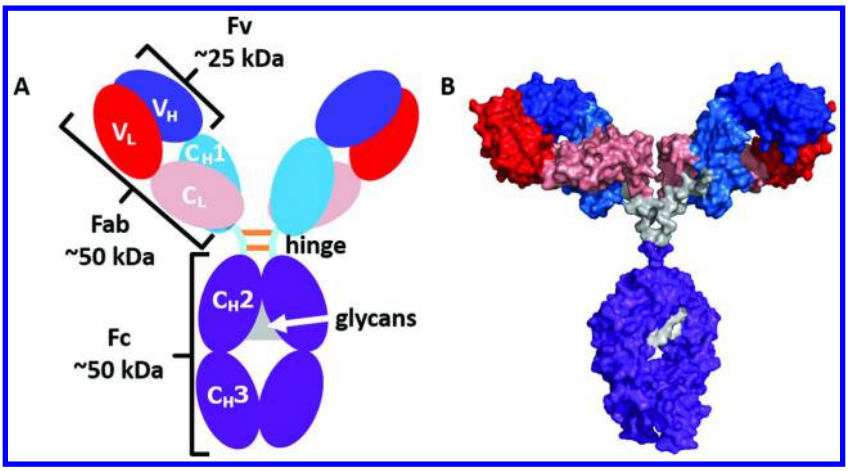

Figure 1. (A) IgG schematic antibody with the immunoglobulin (Ig) domains represented by ovals and the glycans represented by a gray triangle between the two C_H2 domains. The domains are labeled and color coordinated with the space filling model to the right. The light chain variable domain is red, and the constant domain is salmon. The heavy chain variable region is dark blue; the constant domain C_H1 is medium blue; the hinge is light blue; and the crystallizable fragment (Fc) is purple. (B) Space filling model of the NISTmAb. (see color insert)

Post-translational glycosylation greatly increases the challenge of antibody structural analysis. In addition to glycans that modulate function in the antigen-binding regions of some antibodies, all native human IgGs have a glycan attached to Asn297 (according to the Chothia, et al. (*8*) numbering system) in the C_H2 domains of the Fc region. This large adduct, typically about 10 saccharide units, is the result of a complex, multi-enzyme assembly process and is heterogeneous in both saccharide composition and connectivity, even under rigorous conditions of production (*9*). Moreover, different glycoforms are produced based on the host cell expression system, and these differences are known to influence Fc structure and interactions and thus biological function. The need for precise and validated measurements of glycan composition and resulting protein structural heterogeneity is particularly acute because therapeutic antibodies are produced in engineered cell

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

lines, in which the resulting glycoform patterns are not uniform despite rigorous process controls (Glycosylation chapter/Volume 2, Chapter 4).

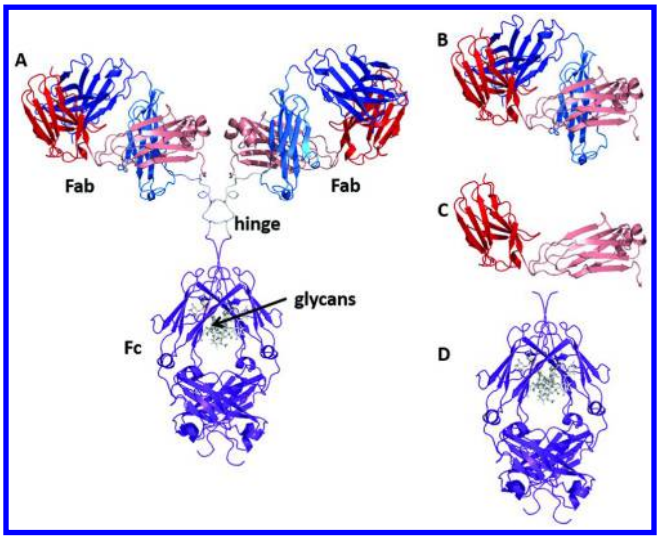

Figure 2. Ribbon diagram structures from the structural model of the NISTmAb. (A) complete IgG antibody; (B) Fab (antigen binding fragment); (C) entire light chain; (D) Fc (crystallizable fragment) with glycans as grey stick models. (see color insert)

In this chapter, we briefly review current standard technologies (X-ray crystallography and, briefly, spectroscopic methods) and then focus on two emerging technologies, nuclear magnetic resonance spectroscopy (NMR) and hydrogen deuterium exchange mass spectrometry (HDX-MS), for assessment of higher order structure of mAbs and provide illustrative examples of applications of these methods using the NISTmAb, an IgG1 kappa antibody and a "drug-like substance" that was donated to National Institute of Standards and Technology (NIST). As the potency and safety of a mAb biopharmaceutical is strongly correlated with its higher order structure, the precision and accuracy with which methods can measure the structural comparability of therapeutic protein drugs is a critical element in ensuring the quality of each therapeutic product (*10*). The strengths, weaknesses, and complementarity of information derived from each of the measurement technologies are discussed.

mAb Crystal Structures

High-resolution structural characterization of antibodies has been achieved primarily using X-ray diffraction techniques. Since the first antibody Fab domain structure was determined in the 1970s (*11*), the technique of protein crystallography has produced numerous structures of antibody fragments, revealing both common fundamental architectures and specific details of

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

molecular interactions. Monoclonal antibodies can be produced in adequate amounts and purity for crystal growth, but their inherent flexibility is problematic because crystals require molecules that can adopt identical conformations in well-ordered lattices. Within a unit cell of a crystal there can exist more than one conformation of the structure; however, if the condition for regular ordering is not met, a crystal will not grow through formation of repeated identical unit cells. In the PDB there are three structures of intact antibodies, 1HZH from human, as well as 1IGT and 1IGY from mouse, all at relatively low resolution (2.7, 2.8, and 3.2 Å respectively). These structures show a wide variety of hinge conformations while clearly showing the familiar overall modular Ab architecture with attached glycans (Figure 3).

Although intact antibodies crystallize poorly and only three structures are available, the separated Fab and Fc fragments are reasonable candidates for crystallization and are represented by about a thousand PDB structures. Most of these are Fabs, and most of the Fabs are in complex with antigen, providing a wealth of data on the specific interactions underlying immunity. Whereas Fab (and Fv fragment) structures explore antigen interactions, the Fc fragment has been crystallized in complex with biological interaction partners to elucidate downstream signaling. Many of the structures are at high resolution, thus providing atomic-level details on the molecular interactions of antibodies and informing engineering and design of antibodies as medicines.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

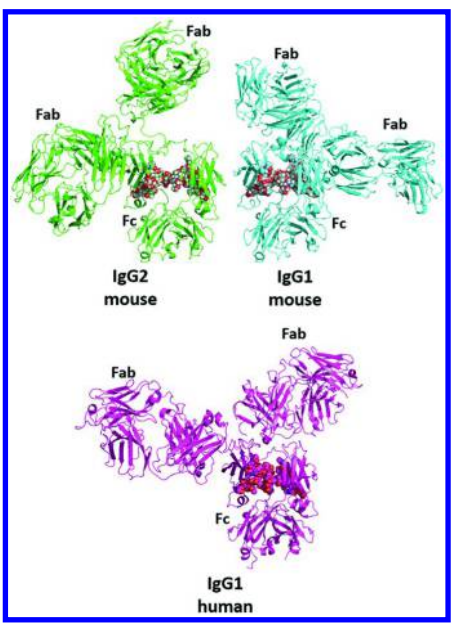

Figure 3. Three IgG antibody crystal structures with their crystallizable fragment (Fc) regions oriented as in Figure 1, showing wide variation of hinge conformations and antibody binding fragment (Fab) orientations. The glycan atoms are shown as space filling balls: IgG2a mouse (1IGT); IgG1 mouse (1IGY); and IgG1 human (1HZH). (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Spectroscopic Methods

Currently, higher order structure is most commonly assessed by low- and moderate-resolution spectroscopic techniques such as intrinsic fluorescence, circular dichroism (CD), vibrational circular dichroism (VCD), Raman, Raman optical activity (ROA), and Fourier transform infrared (FT-IR) spectroscopies. These spectral techniques can provide fingerprints of the structure(s) of protein therapeutics and are used as tools for establishing consistency in drug manufacturing, for detecting drug product variations inherent to or resulting from modifications in the manufacturing process, and for comparing a biosimilar to an innovator reference product. All share the advantage of being relatively high in sensitivity, thus allowing rapid acquisition with small amounts (typically micrograms) of material. Using CD, VCD, Raman, ROA, or FT-IR spectroscopes (*12–16*), the type and aggregate amounts of secondary structural elements (helix, beta sheet, turn) can be identified, monitored as a function of sample conditions (e.g., pH, buffer, temperature), and measured over time to give a readout of possible changes in structure and dynamic behavior. In addition, ROA spectroscopy can distinguish different molecular populations in fast conformational exchange in the nanosecond range (*16*), and VCD has the distinct advantage of allowing measurements at concentrations as high as 50 mg/ml without dilution (*15*). The limitation of these spectroscopic approaches, however, is that they do not provide assignment of signals to specific secondary structural elements within the protein fold, and therefore, the correlation of the observed spectral differences with specific changes in structure is not possible without an orthogonal technique (*14*).

In contrast to CD, VCD, ~~standard~~ Raman, ROA, and FT-IR spectroscopies, intrinsic fluorescence reports more directly on the local molecular environment of a fluorescent amino acid (*17*). Tryptophan fluorescence is most often used for this approach due to its higher quantum yields compared to tyrosine and phenylalanine. For proteins like mAbs that have more than one tryptophan, time-resolved intrinsic fluorescence can be applied to attempt to parse out the contributions from the multiple tryptophans. The relative contributions of the different tryptophans to the total measured fluorescence can, however, be hard to deconvolute and interpret. Although protein mutants can be produced that reduce the number of tryptophans in a protein to one and thus simplify the emission spectrum, such approaches have limited utility in a context of an industrial setting. Further details on spectroscopic and other biophysical techniques for mAb characterization are covered in the Biophysical chapter/Volume 2, Chapter 6 of this series.

NMR Structural Fingerprinting of Protein Biologics

Although X-ray crystallography has generated a wealth of high-resolution structural data for antibodies, particularly for Fc and Fab fragments and complexes between Fab fragments and antigens, the approach is unable to assess the solution structure of a protein therapeutic in formulation. In solution, NMR spectroscopy can in principle provide atomic-level characterization of mAb, Fc, and Fab

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

structure. As each NMR signal represents, for example, a specific proton or a specific proton-nitrogen correlation in a protein molecule, the method can provide an atomic-level readout of structure, which in theory is limited only by the experimental precision of the spectrometer. In the case of modern NMR spectrometers, this precision is very high (in parts per billion). However, the practical resolution of the NMR experiment is governed by the size of the protein and the corresponding complexity of the spectral map. As a protein increases in size, more overlap will naturally occur due to the fact that there are more signals within a given region of the spectra. NMR is also an intrinsically insensitive method owing to the fact that the measured signals arise from a small difference in the energy of the two states of a given nuclear spin, which results in only a small population bias towards the lower energy state and a correspondingly small population inversion upon excitation by a radio frequency pulse. Significant improvements in both NMR hardware and methodology over the past decade, however, have opened up the potential for application of NMR methods for the structural assessment of biologics during discovery, production, and for quality assurance (*18*, *19*). Specifically, console electronics (e.g., digital amplifiers and receivers) and probe technologies, including the development of cryogenically cooled probes (cryoprobes), has enabled the practical application of structural fingerprinting of proteins using NMR-active nuclei at the very low level of natural abundance (e.g., $^{15}N = 0.37\%$, $^{13}C = 1.07\%$) (*20*). Before the advent of cryoprobes, high concentrations (>10 mM) or isotopic labeling of proteins was required for practical application of these methods. Other recent probe design advances have also allowed great reduction in sample volume from the standard 500 to 600 μL down to a few microliters (*21*). Together, these sensitivity gains and sample volume reductions have opened the door to application of NMR methods to mass-limited samples (e.g., samples that are difficult to obtain in the quantities of 1 milligram or greater normally required for standard sample sizes) without the requirement for stable-isotope labeling.

Among NMR methods, one-dimensional (1D) proton ($^{1}H = 99.9\%$ natural abundance) spectroscopy of protein biologics provides the simplest, highest sensitivity approach for structural assessment. Every proton on a molecule will resonate in the ^{1}H NMR spectrum at a given frequency that is dictated by its specific electronic environment, which includes solution conditions, sample temperature, and local chemical structure. It is the unique local electronic environment that will shield a specific proton from the external magnetic field and afford a specific frequency position. Often, a specific NMR-active nucleus will be referred to as a "resonance" because the NMR atom or spin precesses at a given frequency in a manner similar to the way a gyroscope precesses in the earth's gravitational field. Since many different static magnetic field strengths are used, the precise resonance position of a proton nucleus is not normally reported as a frequency. Rather, a chemical shift scale is used that is a normalization of the frequency scale and is given in parts per million (ppm). This allows data collected at different magnetic fields to be easily compared. For ^{1}H nuclei, typical chemical shifts in proteins range from −1.0 ppm to 11.0 ppm. For ^{13}C nuclei, the chemical shift range is 5 ppm to 185 ppm and, for ^{15}N nuclei, 30 ppm to 190 ppm.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Another key component in the application to NMR is spin relaxation that governs the practical implementation of NMR experiments. After a radio frequency pulse, the bulk magnetization of the sample will eventually return to its original equilibrium state. This process, called *longitudinal* or T_1 *relaxation*, defines how quickly an experiment can be signal averaged. T_1 relaxation rates vary with proton type, but practically, 1 to 2 seconds is typically used in between scans for protein measurements. The second type of relaxation, called *transverse* or T_2 *relaxation*, defines the degree to which magnetization coherence is lost after a radio frequency pulse. Each spin within a molecule will experience a slightly different fluctuation in local magnetic field, resulting in the loss of coherence of the individual magnetization vectors and, ultimately, causing the full loss of signal. Resonances from molecules that tumble faster in solution tend to have longer T_2 values, whereas resonances from large molecules, such as mAbs, have short T_2 values and lose magnetization coherence much more quickly. In practice, T_2 rates govern the resolution of a spectrum as these rates influence the observed resonance line-widths.

As a fingerprinting tool, several 1D ^{1}H NMR methods have been developed to assess higher order structure of protein biologics (*18*). In one study, 1D ^{1}H NMR was used to compare two Filgrastim products, the innovator product Neupogen® and follow-on product Zarzio®, and concluded that these spectra could demonstrate "structural similarity" of the two drug products (*22*). Another investigation on intact mAbs under formulated conditions applied 1D ^{1}H NMR to establish comparability using a method termed PROtein FIngerprint by Line shape Enhancement (PROFILE) (*23*). The PROFILE method allows the selective filtering of the mAb proton signals from the water and excipient signals. Any residual signal from excipients that form micelles, such as polysorbate, are removed by collecting a buffer-only spectrum and subtracting it from the sample spectrum. After intensity normalization of the subspectra, a correlation coefficient is calculated. This method therefore allows not only for the evaluation of structure but also the effect of formulation on parameters such as hydration and dynamics, which can be correlated to aggregation behavior. The PROFILE method was shown to allow rapid and precise assessment of the structural comparability of different intact mAbs under formulated conditions and determined that their 1D ^{1}H PROFILE gives highly similar results to two-dimensional (2D) ^{15}N,^{1}H correlation spectra of the isolated $F(ab)_2$ and Fc domains that have been ^{15}N-labeled. It also provides a sensitive measure of overall phenomenological changes like protein unfolding or aggregation, as with the conventional 1D ^{1}H spectrum, even though it does not provide the resolution needed to attribute structural differences to specific sequence elements.

Specific assignment of the 1D ^{1}H spectrum of a protein is typically not practical due to the number of resonance signals and the resulting spectral overlap. Instead, well established methods for detailed structural characterization by NMR, generally limited to proteins approximately 50 kDa in size or less, initially involves isotopically labeling a target protein with ^{13}C and ^{15}N to allow the application of 2D and 3D heteronuclear techniques to assign ^{1}H, ^{13}C, and ^{15}N resonances. The term "heteronuclear" refers to NMR experiments that correlate protons with other heteronuclei (typically carbon and/or nitrogen). With

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

assignments in hand, further NMR experiments can be carried out to generate torsional, distance, and orientational restraints that are used to compute 3D structural models. These methods for structure determination of proteins by NMR have been extensively reviewed elsewhere (*24*).

In the absence of a structural model or even resonance assignments, 2D NMR experiments still provide a high-resolution fingerprint of the structure that can be used for comparability assessment (e.g., to monitor the effect of manufacturing changes or to compare a biosimilar to an innovator product). In particular, NMR methods that correlate one bond-coupled amide protons and nitrogens, such as 2D ^{15}N,^{1}H-heteronuclear multiple quantum coherence spectroscopy (HMQC) and 2D ^{15}N,^{1}H-heteronuclear single quantum coherence spectroscopy (HSQC), offer a unique structural fingerprint of a protein molecule at atomic resolution. Every non-proline amide in a protein sequence is ideally represented by a ^{15}N,^{1}H correlation, whose peak position is determined by its unique chemical environment. The typical chemical shift range of this fingerprint region is from 6.0 ppm to 11.0 ppm for ^{1}H and 100 ppm to 140 ppm for ^{15}N. The chemical environment of each amide is influenced by, among many factors, primary structure, secondary structural elements (i.e., α-helix, β-sheets, etc.), as well as tertiary folding and quaternary interactions. The 2D HSQC therefore serves as structural fingerprint of a protein with any deviations in the folding of the protein resulting in a change of the chemical shift of one or more amide resonances. Solution and temperature conditions also can influence the chemical environment of a nucleus and potentially result in chemical shift perturbations and so need to be matched between samples in any structural comparability exercise. In addition, these 2D NMR measurements can be carried out on the formulated drug without need for sample manipulation. In practice, however, formulations with large aromatic signals can interfere with the amide region of these protein spectra.

The use of 2D ^{15}N,^{1}H HSQC spectroscopy for fingerprinting protein biologic structure(s) was first demonstrated in 2008 using a recombinant protein, human granulocyte macrophage-colony stimulating factor (rhGM-CSF) (*25*). Through comparisons of ^{15}N,^{1}H correlation spectra, the study demonstrated that the spectral fingerprint of ^{15}N-labeled rhGM-CSF produced in *Escherichia coli* (*E. coli*) could be directly overlaid on that of the therapeutic version, Leucotropin™, which was produced in *Streptomyces lividians.* Although no attempt was made to quantitate the degree of similarity in this initial study, the high degree of overlap of the amide resonances in the spectral fingerprint suggested that the drug substances were structurally highly similar. In a subsequent study, the 2D NMR approach for structural mapping and comparability assessment was demonstrated using a second chemokine-class drug, Interferon Alpha-2 (IFN) (*26*). Through application of the 2D NMR fingerprinting methodology in this study, the structure was determined to be unaffected by the process of formulation and deformulation through a wide pH range of 3.5 to 8.0. A slight propensity of IFN to aggregate above pH 5.0 was observed, but this tendency was ameliorated by excipient choice, which destabilized the formation of oligomers. Below pH 3.0, dynamics fluctuations in structure, observed as line broadening and chemical exchange, marked the threshold of protein unfolding. This observation had been previously noted by fluorescence, circular dichroism, and differential scanning calorimetry

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

measurements (*27*). Taken together, these seminal studies illustrated the power of NMR to monitor the folded state of protein biologics in various formulations, and demonstrated its potential as a comprehensive structure comparability tool at atomic resolution in which signals could be sequence-specifically assigned to amino acid residues.

2D ^{1}H homonuclear (e.g., correlations of two or more spins of the same nucleus such as protons) NMR methods, such as nuclear Overhauser enhancement spectroscopy (NOESY) (*28*), also have been used to generate spectral fingerprints for structural comparability that, in principle, provide a greater ability to detect structural changes due to the enhanced resolution provided by the second dimension. A 2D NOESY spectrum represents all pair-wise distance correlations between proton resonances in a protein that are within 5 Å of each other. As such, two resonances can be correlated for amino acid pairs that are distant from each other according to the primary structure but close in space due to the higher order folding of the protein. The closer that two spins are in space, the greater that the cross-peak intensity will be. The cross-peak intensity for two spins 2 Å apart will, in general, be much greater than two spins 5 Å apart as the nuclear Overhauser effect is proportional to r^{-6}, where r is the distance between the spins. The data derived from this type of experiment maps to the higher order structure of a protein and thus can provide a tool for structural fingerprinting of higher order structure. In subsequent work, two statistical methods were proposed to demonstrate structural similarity from the 2D NOESY: (1) a direct comparison of peak position of all cross peaks within a given spectral region; and (2) a utilization of graph theory to link peaks in a graph by their nearest neighborhoods (*29*). Unfortunately, this study found that the two statistical methods did not give an equivalent measure of similarity amongst batches of drug product from various sources and concluded that additional experiments were needed to refine their statistical approach for similarity.

The 2D NMR fingerprinting methodology has been further extended to ^{13}C,^{1}H HSQC-type spectral fingerprints using the well-resolved methyl region (roughly −1.0 ppm to 2.0 ppm for ^{1}H and 5.0 ppm to 40 ppm for ^{13}C) at ^{13}C natural abundance (*30*). In this work, statistical similarity was established by normalizing the intensity of two NMR spectra, binning the resulting normalized spectra in small blocks (e.g., 0.05 ppm), and using linear regression to determine the correlation coefficient (R^2). The method, called ECHOS-NMR (Easy Comparability of Higher Order Structure by NMR), was applied to several proteins ranging from 6.5 kDa to 67 kDa. Using this approach, it was found that batches of the same protein achieved a correlation coefficient of 0.98 or higher, whereas comparison of two different proteins afforded small R^2 values as low as 0.14 and 0.00, which confirmed that they had little to no correlation, as would be expected. It was also noted in this study that the analysis tools developed for ECHOS-NMR could be applied to other types of NMR spectra because this methodology only requires collection of data of the same type for comparability purposes (e.g., two ^{1}H NOESYs, two ^{15}N,^{1}H HSQCs).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

As a rule of thumb, standard NMR measurements are suited for proteins in the approximately 50 kDa or smaller size range. As the molecular weight increases, the slower correlation time (τ_c) or molecular tumbling of the molecule results in shorter transverse relaxation and broader signals that compromise both measurement sensitivity and resolution. To overcome these issues, a 2D ^{15}N,^{1}H correlation method known as Transverse Relaxation-Optimized Spectroscopy (TROSY) was developed that selects for the component of a ^{15}N,^{1}H cross peak where the major contributions to the relaxation are opposite in sign and effectively cancelled for large proteins with slow correlation times. Using a TROSY experiment that selects this component of the ^{15}N,^{1}H cross peak can yield narrow lines and highly resolved spectra for large proteins (*31*). Similar improvements can be made using analogous ^{13}C,^{1}H TROSY methods (*32*, *33*). The proton-rich nature of proteins compromises the performance of this technique, as these resonances provide an efficient secondary pathway for relaxation of the narrow component of an amide resonance; thus, the TROSY spectra become severely line-broadened. To compensate for this relaxation mechanism, proteins are commonly perdeuterated, which effectively removes this secondary relaxation pathway. Indeed, a perdeuterated protein complex 670 kDa in size with selectively protonated methyl groups has been successfully studied using TROSY-type methods and enabled determination of the catalytic residue involved in the first hydrolysis step of the 20S CP proteasome (*34*). Perdeuteration is generally achieved by the expression of a recombinant protein in a bacteria-based culture in a minimal media with greater than 95% D_2O. To date, similar approaches for perdeuteration of protein expressed in mammalian cell culture have not been demonstrated, which limits application of TROSY-type approaches to intact mAbs.

Application of NMR methods to Fc and Fab fragments generated from papain digestion of intact antibodies or produced from *E. coli* expression of the fragment domains has been shown to provide useful, high-resolution measurements for assessing these domain structures (*23*). In a study of the structural consequences of methionine oxidation of *E. coli*-produced unglycosylated perdeuterated Fc fragments, nearly complete resonance assignments were achieved to map structural changes resulting from forced degradation of the protein using hydrogen peroxide (*35*). The results showed that the attenuation in structural stability was due to a weakening of domain-domain interactions between C_H2 and C_H3 that were attributable to changes in specific residues in the C_H2 domain. This correlated loss of stability due to the oxidation was confirmed using differential scanning calorimetry (DSC). A lower melting temperature (T_m drop of 11 °C) was measured for the C_H2 domain, and the T_m of the C_H3 domain was found to be largely unchanged, confirming what was observed by NMR.

In a another study, a glycosylated and selectively stable isotope-labeled mouse IgG2b-Fc fragment was expressed in Chinese hamster ovary (CHO) cells using media supplemented with ^{2}H,^{13}C,^{15}N-labeled amino acids (*36*). The resulting spectral fingerprint of the amide backbone was well-resolved and dispersed. This allowed conventional triple resonance experiments for resonance

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

assignment to be employed to achieve nearly complete assignment of the protein backbone (*24*). Similar approaches were also used to generate high-resolution spectra with a ^{13}C,^{15}N-doubly labeled glycosylated human IgG1-Fc fragment (*37*). However, only 66% of the resonance assignment for the Fc backbone was achieved, presumably due to spectral overlap and dynamics. Using these ^{13}C,^{15}N-labeled constructs, subtle structural changes to the mAb Fc domains upon trimming of the carbohydrate chains could be mapped to specific residues and used to detect the degree to which the glycans maintained the "structural integrity" of the FcγR-binding region of the Fc.

NMR Structural Fingerprinting of the NISTmAb

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

To illustrate the quality of mAb 1D ^{1}H spectra, ^{1}H spectra of the fully formulated NISTmAb were acquired (Figure 4a) using water flip-back Watergate water suppression with 32 scans over 2 minutes at temperatures ranging from 25 to 50 °C. The NISTmAb is a "drug-like" substance that was donated to NIST in its formulation buffer, which consists simply of 25 mM l-histidine {2-(S)-histidine} at pH 6.0. At 25 °C, the protein signature, especially in the amide "fingerprint" region, is weak due to broad lines resulting from the long correlation time of the intact mAb (≈ 150 kDa) and likely additional conformational exchange due to the motions about the linker regions of the mAb that occur on an intermediate (micro- to millisecond) timescale. The sharp lines at 8.03, 7.10, 3.93, 3.20, and 3.11 ppm are due to the formulation buffer, l-histidine {2-(S)-histidine}. At 50 °C, the increased temperature results in a slightly faster correlation time for the mAb and thus slightly narrower lines and detection of more amide resonances. The observation of a dispersed amide/indole region up to 11 ppm and methyl resonances below 0 ppm give confidence that the NISTmAb is properly folded across this temperature range. As described above, however, the sheer number of resonances, coupled with broad lines from the large size, make the 1D spectrum intractable for further detailed analysis despite the small spectral improvements from the increased temperature of 50 °C.

Similarly, the full NISTmAb spectral fingerprint of the amide region acquired with 2D ^{15}N,^{1}H TROSY and HSQC experiments were collected with 4096 scans per transient over 30 hours to demonstrate the quality of these data for an intact, representative IgG. The TROSY spectral fingerprint yielded sharp signals, but less than thirty cross peaks were observed, as might be expected for a non-deuterated sample where an abundance of proton resonances provides efficient pathways for signal relaxation that can significantly compromise the performance of the experiment (data not shown). On the other hand, the HSQC spectral fingerprint (Figure 5) yielded a more complete map of the protein amide correlation despite broad cross peaks and considerable overlap.

In contrast, cleavage of the full NISTmAb with immobilized papain yielded Fab and Fc fragments that could be more readily fingerprinted using the 2D NMR method. Because the cleavage reaction required a phosphate buffer, the samples were exchanged back into the formulation conditions before NMR analysis using deuterated l-histidine. Ideally in a biopharmaceutical analysis, the formulated

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

drug products would be evaluated directly. In practice, sample manipulation has become accepted convention for characterization of mAbs, as in the case for middle-up or middle-down mass spectrometry approaches, which also involve the cleavage of intact mAb into smaller regions (*38*). For NMR experiments, the 50 kDa Fab and Fc domains show both reduced line width and decreased complexity owing to the reduced size of these fragments relative to the intact mAb (Figure 4b). Application of 2D ^{15}N,^{1}H HSQC NMR spectral fingerprinting methodology to these fragments at natural abundance levels of ^{15}N isotope is indeed possible and can be demonstrated using these fragments generated from the NISTmAb (Figure 6). In this example, instead of the standard HSQC experiment, we selected the SOFAST-HMQC experiment, which is a rapid pulsing technique that promotes faster longitudinal relaxation for the spins of interest (e.g., the amide region), less wait time between scans, and hence, collection of more scans in a shorter amount of time (*39*) [Schanda, 2005 #42]. These spectra could be collected in 24 to 30 hours and showed the resolving power of the 2D method for fingerprinting both the Fab and Fc domains. Comparison of the spectra of the Fc and Fab domains derived from the NISTmAb with published spectra of ^{15}N isotope-labeled Fc and Fab fragments expressed in *E.coli* and cleaved using a similar papain protocol (*23*) shows comparable quality in terms of line widths and spectral resolution. For the NISTmAb, a total of approximately 245 and 482 peaks are expected in the ^{15}N,^{1}H correlation spectra for the Fc and Fab fragments, respectively, when accounting for Pro residues as well as Asn, Gln, and Trp side chains. The natural abundance ^{15}N,^{1}H SOFAST-HMQC spectrum of the NIST Fc exhibits a total of 198 peaks with a signal to noise ratio (S/N) greater than approximately 7:1 (chosen to avoid t_1-noise peaks), or 81% of the expected resonances. Of the 198 identified peaks, approximately 110 are located in reasonably well-resolved areas along the periphery of the spectrum, whereas the remaining 88 are located in the heavily overlapped center of the spectrum or side chain (e.g., loosely, the upper right portion of the spectrum) regions of the spectrum. Comparison to the peripheral regions of the ^{15}N,^{13}C,^{2}H-labeled *E.coli*-expressed Fc spectrum (*35*) suggests a total of 120 peaks should be present, giving approximately 92% coverage in the well-resolved region, whereas the remainder of the missing peaks likely result from increased spectral overlap as a result of minimal acquisition times in both dimensions as well a aglycosylation (*E.coli*-expressed Fc) versus glycosyled (NISTmAb) Fc regions. For the natural abundance ^{15}N,^{1}H SOFAST-HMQC spectrum of the NIST Fab, 357 peaks were identified with a S/N above 7:1, or 74% of the expected resonances. As there is no published Fab reference spectrum to which to compare this data, further analysis cannot be made.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

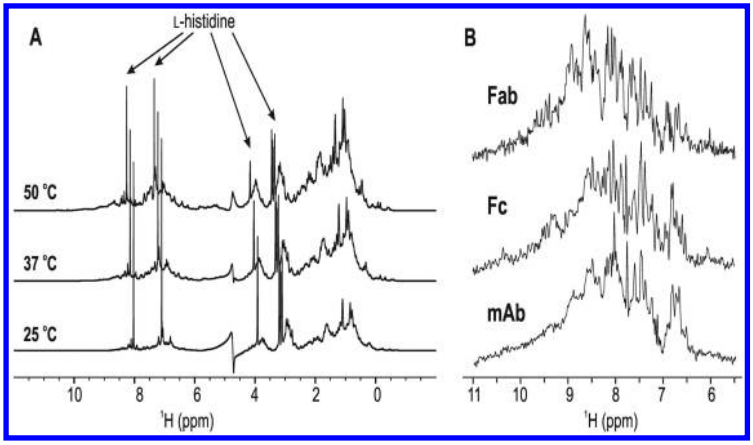

Figure 4. (A) Overlay of one-dimensional (1D) water flip-back Watergate water suppression of intact 0.68 mM NISTmAb in 25 mM L-histidine {2-(S)-histidine}, pH 6.0, at three different temperatures acquired with 32 scans over 2 minutes using a 900 MHz nuclear magnetic resonance (NMR) spectrometer. The peaks in the full ^{1}H spectra corresponding to the L-histidine buffer are labeled.(B) Expansion of the natural abundance, ^{15}N-edited, 1D ^{1}H amide fingerprint region acquired using a $^{15}N,^{1}H$ SOFAST-heteronuclear multiple quantum coherence spectroscopy (HMQC) experiment collected with 2048 scans over 14 minutes at 900 MHz and 50 °C for the intact mAb (bottom), 0.44 mM crystallizable fragment (Fc) (middle) and 0.50 mM antibody binding fragment (Fab) (top). ^{15}N editing of spectra involves the spin labeling of the ^{1}H amide resonances with their corresponding amide nitrogen. Therefore, the resulting 1D ^{1}H spectrum only contains protons attached to nitrogens. For the intact mAb, a Bruker shape tube was used (300 µL). For the Fab and Fc fragments, 5mm Shigemi tubes were used (300 µL).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

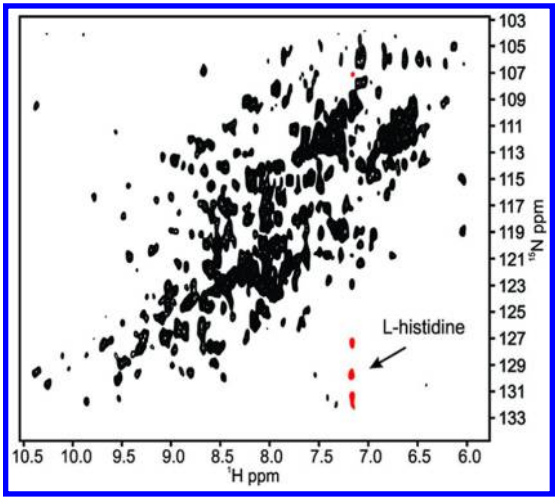

Figure 5. Two-dimensional (2D) spectral fingerprint of 0.30 mM intact NISTmAb in 20 mM bis-tris-d19, pH 6.0, acquired at 900 MHz and 50 °C. ^{15}N,^{1}H SOFAST-heteronuclear multiple quantum coherence spectroscopy (HMQC) was collected with 4096 scans over 30 hours. The artifact at approximately 7.2 ppm is a t_1 ridge arising from the residual L-histidine, the original formulation buffer of the NISTmAb. (see color insert)

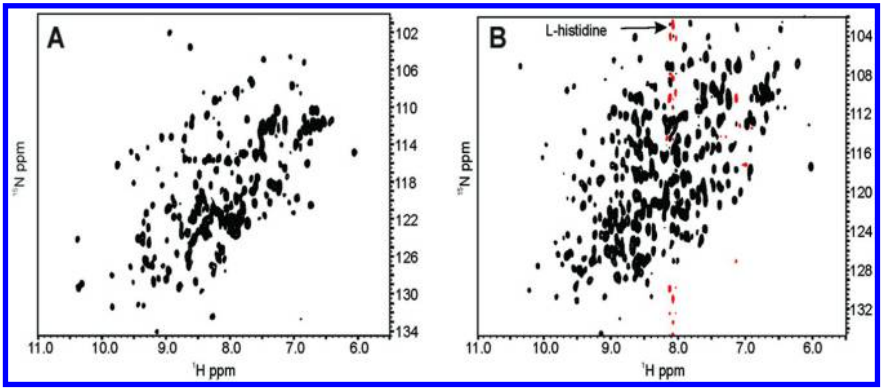

Figure 6. Two-dimensional (2D) spectral fingerprints of unlabeled NISTmAb domains in 25 mM L-deuterohistidine, pH 6.0, acquired using ^{15}N,^{1}H SOFAST-heteronuclear multiple quantum coherence spectroscopy (HMQC) collected with 4096 scans over 24 hours at 900 MHz and 50 °C. (A) 0.44 mM crystallizable fragment (Fc); (B) 0.50 mM antibody binding fragment (Fab). Positive and negative contours are black and red, respectively. In (B), artifacts known as a t_1 ridge from the residual protonated buffer were observed and labeled. (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Beyond simple visual comparisons, the Fc and Fab ^{15}N,^{1}H NMR data collected at natural isotopic abundance can be assessed using more quantitive and statistical methods. For example, previously developed statistical methods, such as ECHOS NMR (*30*), could be employed to analyze these spectra and provide a more statistical measure of the structural similarity between two protein samples. As the data on the NISTmAb represent the first demonstration of 2D NMR spectral fingerprinting of mAb domains, ongoing research will continue to improve the performance of this method and demonstrate its value as a validated tool for high order structural assessment and comparability.

HDX-MS

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

HDX-MS is regarded as one of the mass spectrometry-based footprinting strategies for protein structural assessment (*40–45*). This strategy utilizes subtle chemical modification due to the exchange of deuterium atoms for hydrogen atoms at exchangeable sites in a protein (primarily backbone amides) to generate a mass shift that is readily seen by MS. The modern HDX-MS laboratory, comprising robotic analytical equipment and specialized analysis software, has rendered analyses of 9 to 300 kDa proteins as routine (*46*). HDX-MS can reveal the dynamical behavior of very complex systems. For example, HDX-MS results have revealed the dynamic structure of hepatitis B (≈ 6 MDa) (*47*) and HIV-1 (≈ 275 MDa) (*48*) viral capsids. In short, HDX-MS manifests scant restrictions with respect to protein size.

Other chemical modification strategies can be employed similarly, such as hydroxyl-radical oxidative modification and chemical crosslinking, which label side chains rather than backbone amides, providing complementary structural information (*45*). HDX and hydroxyl radical-mediated protein footprinting offer information on protein conformational changes associated with protein-ligand (ligand: protein, DNA, peptide, drug, etc.) interactions and protein solution dynamics, though each technique has advantages and limitations (*43*, *49–51*).

HDX has unique characteristics in which the protein hydrogens can be classified into three classes based on their different deuterium exchange rates. The first class includes amide hydrogens that exchange with deuterium at medium rates (half-lives of seconds to hours). A second class comprises hydrogens attached to the functional groups of side chain, which exchange with deuterium at very fast rates (half-lives of microseconds). Thus, these hydrogens are easily back-exchanged to hydrogen by returning the protein to a protic solvent. The third class includes hydrogens covalently bonded to carbon. Essentially, these hydrogens do not exchange with deuterium in D_2O. These characteristics enable us to measure the extent of deuterium uptake of the amide residues (amino acids) on a reasonable time frame, thereby revealing a wide range of protein dynamics.

The extent of deuterium uptake of a protein can be measured globally by investigating the average difference between the masses of the deuterated protein and the undeuterated protein. HDX is initiated by diluting the proteins in D_2O buffer at a fixed pH (often physiological pH) and convenient temperature (often 25 °C) and then by allowing the solution to react for a chosen time interval. At

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the conclusion of the time interval, the buffered protein/D_2O solution is added to a cold quenching solution, which establishes a solution environment of pH 2.5 and 0 °C. This cold, acidic environment greatly slows hydrogen exchange processes. The deuterated samples are loaded onto a C_8 or C_{18} trap column, and the chromatographic mobile phase, which contains H_2O, washes away salts. While on the trap column, the labile side-chain sites undergo rapid back-exchange (where deuterium is replaced by hydrogen). After elution from the trap column, the protein may be purified further via a cooled ultrahigh performance liquid chromatography (UPLC) system and analyzed by mass spectrometry. This strategy enables us to monitor global changes of hydrogen bonding and solvent accessibility of the protein of interest originating from perturbations, including ligand binding and the presence of post-translational modifications (PTMs).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

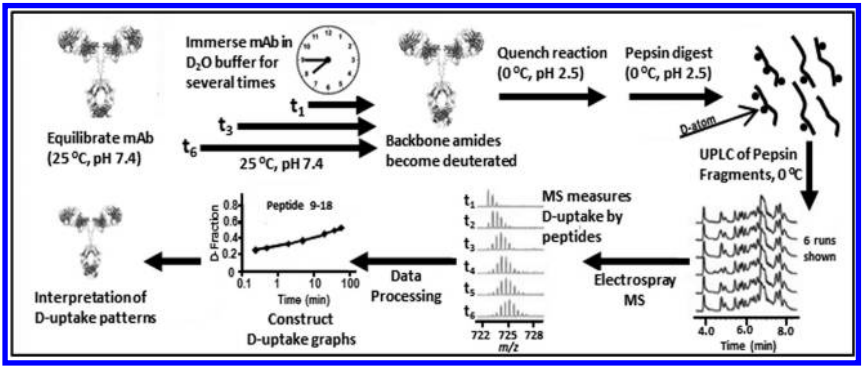

Figure 7. Hydrogen-deuterium exchange (HDX) workflow that provides "peptide-level" information. A protein is diluted into D_2O-initiating exchange. After various durations, the reaction is quenched at pH 2.5 and 0 °C, where the rates of forward and reverse exchange are minimized. The protein is denatured and digested to form peptides that are separated by ultrahigh performance liquid chromatography (UPLC) and analyzed by mass spectrometry (MS), allowing changes in mass to be measured and correlated to the presence of deuterium. The amount of deuterium uptake is measured versus time of exposure to D_2O, permitting rates of exchange in particular regions of the protein to be determined.

In many cases, protein conformational changes occur only at small regions of a large protein, and these subtle changes are difficult to capture by investigating the whole protein. To reveal these minor changes, we use the higher spatial resolution provided by proteolytic fragmentation HDX-MS (Figure 7). In this measurement method, the sample protein undergoes enzymatic digestion and chromatographic separation prior to mass spectrometric analysis. Proteolytic fragmentation HDX-MS typically affords spatial resolution at the "peptide-level" (i.e., regions of the protein that are 6–10 amino acid residues in length (*52*)).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

The extent of deuterium uptake is measured as the mass-to-charge differences between the centroids of the isotopic distribution of the deuterated peptide and the protonated peptide. The proteolytic enzymes used to achieve peptide resolution in HDX-MS are mainly acidic proteases, such as pepsin or protease XIII, which are active under the quenching conditions (i.e., pH ≈ 2.5).

To refine the spatial resolution to the "residue (amide) level", one can incorporate multiple enzymatic digestions and analyze the overlapping peptides (*53*). Alternately, one can apply an appropriate gas-phase fragmentation method to the deuterated peptides released upon pepsin digestion. Fragmenting peptides with electron-transfer dissociation (ETD) or electron-capture dissociation (ECD) has been observed to be more reliable than other gas-phase fragmentation strategies, such as collision-induced dissociation (CID), which can corrupt nascent deuterium content information via randomization of the hydrogen and deuterium occupancies of the amides and through sequence scrambling of the amino acids (*54*, *55*). Furthermore, procedures have been described that can largely prevent the hydrogen and deuterium scrambling in the gas phase (*56–58*). Both the methods of top-down (*54*, *59–62*) and bottom-up HDX (*63*, *64*) allow single-residue resolution to be achieved.

A recent set of experiments demonstrated that HDX-MS is traceable to the International System (SI) of units ; hence, calibration hierarchies to fundamental standards can exist. Englander's group observed proteolytic fragmentation HDX-MS peptide-level data for staphylococcal nuclease, SNase (PDB: 1SNO) at four pHs for several exchange times, t_{ex} (*65*). Each data set comprised approximately 300 proteolytic peptide fragment ions that overlapped extensively in sequence, enabling fits at single amide resolution. After correcting for back-exchange, these datasets were fit to yield the deuterium occupancy at each amide site. Fits of deuterium occupancies versus t_{ex} gave the site-resolved HDX rate coefficients for each amide, which were converted to protection factors (*66*).The measured amide hydrogen exchange rates vary by a factor of 10^7. Previously, Englander's group measured the same HDX rates by 2D NMR to a precision (u_{cA}) of 14 percent (*67*). The HDX-NMR and HDX-MS data have 99 sites in common. Comparison of the HDX-MS protection factors shows that 81 amino acids lie within threefold of the measured NMR protection factors, 16 lie within 10-fold, and 2 are outliers. This comparison is direct—it requires no scaling factors. Thus, because exchange rate coefficients measured by HDX-MS and HDX-NMR correspond exactly and the NMR frequency can be calibrated against a time standard, exchange rate coefficients determined by HDX-MS are also directly traceable to a time standard.

Relevant to HDX-MS studies of antibodies, glycosylation, the N-linked attachment of sugar moieties at asparagine (Asn300 in the present antibody sequence, though conventionally numbered as Asn297), increases mAb microheterogeneity. Glycans can retain deuterium, which can exaggerate the apparent deuterium content of observed glycopeptide ions (*68*). Experiments employing electron transfer dissociation can resolve between the glycan and peptide deuterium content, allowing correct assessment of the deuterium uptake by peptides (*69*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

HDX-MS Footprinting of mAbs

Protein conformational changes affect many biological functions, including enzymatic catalysis and selective interactions of macromolecules. A straightforward strategy for examining the protein conformational changes induced by ligand binding (antigen interactions), by PTMs (e.g., glycosylation, phosphorylation), or by the change of the external environment (e.g., pH, temperature, ionic strength) is to measure the changes in protein dynamics. Proteins consisting of multiple domains joined by flexible linkers can undergo large movements that result in the repacking of adjacent domains. The slow dynamics of protein conformational changes usually involve breaking and making relatively strong hydrogen bonds, and these are associated with low-solvent accessibility at short times of exchange. A HDX kinetics study has the potential to identify domain-packing interactions that correspond to different protein conformational states in large proteins. The recent addition of automation to the standard HDX-MS platform has relieved scientists of the many arduous laboratory hours required to obtain kinetic data. This great reduction of manual labor and the recent development of user-friendly data-analysis software have allowed HDX-MS to become a suitable tool for routine characterization of therapeutic proteins (*70*).

HDX-MS has been successfully applied to study the changes of conformational dynamics of an IgG, a 150 kDa antibody isotype that features glycosylation (*71*, *72*), antigen binding (*73*), site-directed mutation (*74*), and the alteration of ionic strength in the external environment (*75*). Heat maps comprising sets of observed peptides are commonly used to present the protein conformational dynamics of whole proteins. Figures 8 and 9 present the colored "heat maps" of the NISTmAb heavy chain (HC) and light chain (LC), respectively, based on HDX-MS kinetics measurements. Each peptide is depicted by a colored bar and a bracket (black lines), and its location in a map corresponds to the amino acid sequence of the intact mAb. Each bar comprises a stack of seven horizontal stripes, where the color chosen for each stripe (refer to color code bar) reports the percent deuterium content observed in the peptide at exchanged times ranging from 30 s (top) to 120 min (bottom). Warmer colors correspond to greater extents of deuterium uptake. Each bracket spans two amino acids at the N-terminus of the peptide. These amino acids are not colored, as they have undergone rapid deuterium for hydrogen back-exchange during the UPLC analysis, thus losing all information about the nascent deuterium content.

The color variations help us visualize the relative structural stability of each dynamical region. For example, peptide regions HC(95–100) and HC(183–188), located at the Fab region, and peptide regions HC(303–309) and HC(408–413), located in the Fc region, show deep blue color (low deuterium uptake) among seven HDX time points (30 s, 2 min, 5 min, 20 min, 40 min, 60 min, and 120 min) (Figure 8), indicating these regions are structurally stable or solvent protected. In contrast, peptide region HC(238–255), located near the hinge region, displays a gradient of color change from blue to red, indicating the very dynamic nature of this region.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Analyses of the overlapping peptides can improve amide resolution through computation of the uptake differences or through simultaneous fits of all temporal data (e.g., ref. (*65*)). For example, at all times, peptide HC(159–166) exhibits greater deuterium uptake than the overlapping peptides (Figure 8). This is indicative of the very dynamic nature of this short region. The activity along residues HC(159–166) may account completely for the temporal color change observed in the overlapping peptides HC(159–178) and HC(148–167). However, computations of the differences among overlapping peptides must be evaluated carefully for complications, as the results can be significantly perturbed by sequence-specific differences in back-exchange rate coefficients and by peptide-column interactions (*76*). The reduction of such distortions is an active research field (*77*, *78*).

An understanding of the HDX-MS dynamics data is more easily realized when the peptide deuterium uptake values are mapped onto the 3D structure of the protein, as obtained by X-ray crystallography or NMR. Figure 10 illustrates the deuterium uptake data mapped onto the theoretical NISTmAb structure for two time points, 30 s and 60 min. Deuterium uptake is presented with color ranging from blue to red for low to high uptake, respectively. Black sections of the protein indicate regions in which no peptides were detected during the proteomics experiments, such as LC(83–104) (Figure 9). The two structures show that after 60 min of deuterium exchange, the structure is represented by warmer colors, indicating greater deuterium uptake than at the earlier time point. In particular, there are β-sheets within the constant region of the Fab and throughout the Fc region which exhibit significant deuterium uptake at 60 min, as indicated by the yellow and yellow-green colors. Alternatively, there are regions in both the Fab and Fc that show very little deuterium uptake, as evidenced by the blue color in the protein structure. Inspecting the two structures allows the examination of regions of the protein which are more or less dynamic and provides insight into how the 3D structure may affect the measured rates of deuterium uptake. This simple process facilitates an understanding of correlations between dynamics and structure.

Acquisition of the HDX-MS peptide map of a mAb is only the starting point of a study. For example, the present HDX-MS information for the apo-NISTmAb could be compared to NISTmAb engaged in noncovalent bonding, as occurs with antibody-antibody, antibody-antigen, or antibody-excipient interactions (*47*, *79–85*). Difference maps of the apo-NISTmAb and holo-NISTmAb can reveal the changes in dynamics induced by such interactions. Interpretation of difference maps can help identify epitopes and allosteric effects. HDX-MS studies with mutated mAb or antigen can improve the reliability of these results (*79*, *86*, *87*). Alternately, the present NISTmAb could become the control for a comparability study where the changes in dynamics among batches of mAb, environmentally stressed mAb, or mutated mAb forms are examined (*70*, *79*, *88*, *89*). For all studies, computational tools can improve the interpretation of results.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

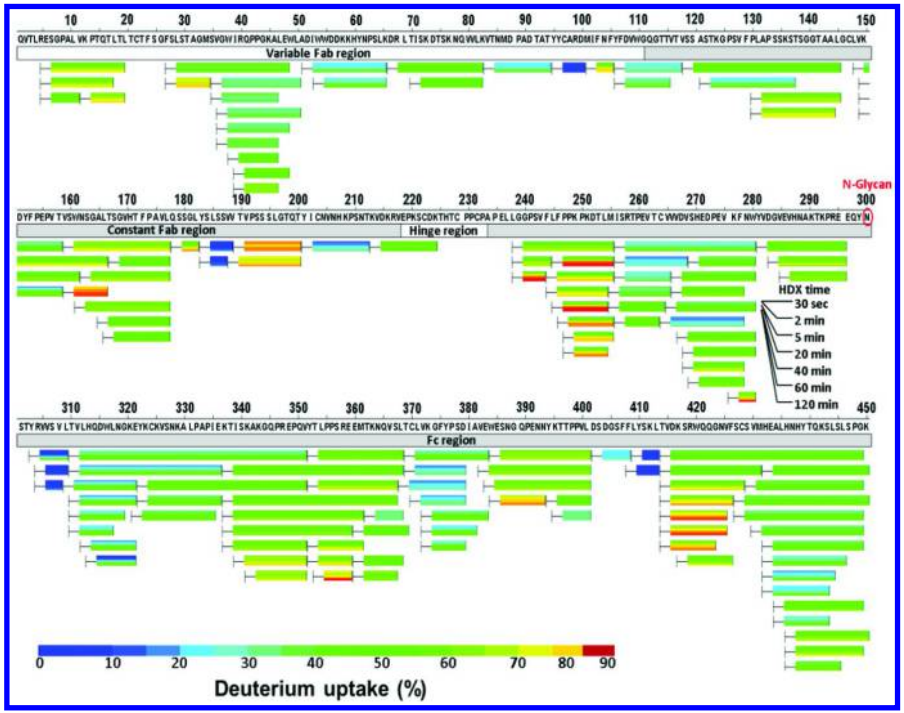

Figure 8. Hydrogen-deuterium exchange (HDX) heat map of the NISTmAb heavy chain. Each bracket and colored bar represent an observed peptide and its placement in the mAb peptide sequence. In each colored bar, the HDX kinetics data from seven HDX time points (30 s, 2 min, 5 min, 20 min, 40 min, 60 min, and 120 min; triplicate) are plotted in stripes from top to bottom. The extent of relative deuterium uptake from low to high is presented with color ranging from blue to red. (see color insert)

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

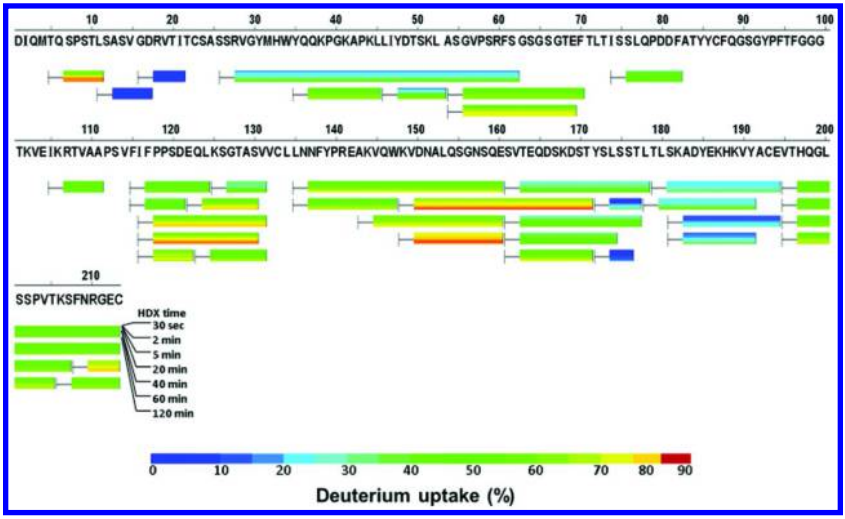

Figure 9. Hydrogen-deuterium exchange (HDX) heat map of NISTmAb light chain. Each bracket and colored bar represent an observed peptide and its placement in the mAb peptide sequence. In each colored bar, the HDX kinetics data from seven HDX time points (30 s, 2 min, 5 min, 20 min, 40 min, 60 min, and 120 min; triplicate) are plotted in stripes from top to bottom. The extent of relative deuterium uptake from low to high is presented with color ranging from blue to red. (see color insert)

The automated feature of the modern HDX-MS platform makes it a high-throughput technique. The extremely small quantities of protein (0.5 to 2 nmol) required for extensive studies and the relatively rapid turnaround of 1 to 2 days for both proteomic and HDX-MS characterization allows this method to fit well into the routine work flow of the discovery and development pipelines of the pharmaceutical industry (*70*, *79*, *90*). Given that HDX-MS is a relatively new area compared to other techniques, much ongoing research and development will continue to improve its performance. In the future, innovations may reduce the deuterium back-exchange (currently 20% to 30% of back-exchange is usually observed); achieve HDX at the single amino acid residue level more precisely; and improve the accuracy of data-analysis software. Thus, due to its aforementioned advantages, in the coming years, HDX-MS will likely be acknowledged as one of the key strategies in therapeutic protein characterization.

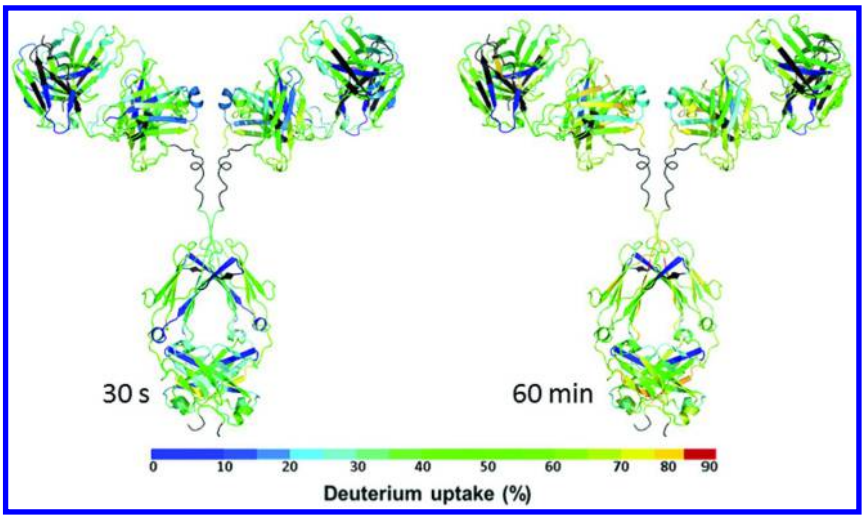

Figure 10. Deuterium uptake mapped onto the theoretical NISTmAb structure for two time points, 30 s and 60 min. Deuterium uptake is presented with color ranging from blue to red for low to high uptake, respectively. Black sections of the protein indicate regions in which no peptides were detected during the experiments. (see color insert)

Summary

NMR and HDX-MS technologies for assessment of higher order structure of mAbs, as described in this chapter, hold great promise as tools for the high-resolution assessment of the structure of protein therapeutics in solution and formulated state at atomic resolution. The NMR and HDX-MS experiments are easily shown to conform to the relevant fundamental units of kilogram, second, ampere, kelvin, and mole. Consequently, these measurements are firmly based in the SI and can be traced through the calibration hierarchy to fundamental standards. This calibration hierarchy will rely on reference materials designed for use in each method. The existence of a traceable calibration hierarchy can promote the use of these techniques for commercial applications (e.g., comparability studies). Furthermore, these approaches and X-ray crystallography also can be combined with other structural methods, such as electron microscopy and small-angle X-ray scattering (*91–94*) to gain a detailed picture of oligomeric and aggregate protein states.

These powerful techniques are still, however, in the relatively early stages of application as structural assessment and comparability tools for protein therapeutics, and many questions still need to be answered. For example, the level at which NMR of HDX-MS will be able to detect subtle differences in mAb conformations and isoforms remains to be determined. The characterization of this property will require studies of a number of mAbs under different conditions. In addition, further developments in instrumentation and protocols can be anticipated, which will continue to enhance both the precision and

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

sensitivity of the methods. Harmonization and standardization of the approaches through interlaboratory studies will also be required to establish confidence in the measurements.

NIST Disclaimer: Certain commercial equipment, instruments, and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

References

1. Vlug, A.; van Remortel, P. *Eur. Clin. Lab* **1989**, *8*, 26–35.
2. Alzari, P. M.; Lascombe, M. B.; Poljak, R. J. *Annu. Rev. Immunol.* **1988**, *6*, 555–580.
3. Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309–320.
4. Padlan, E. A. *Mol. Immunol.* **1994**, *31*, 169–217.
5. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501.
6. Sela-Culang, I.; Kunik, V.; Ofran, Y. *Front. Immunol.* **2013**, *4*, 302.
7. Stanfield, R. L.; Zemla, A.; Wilson, I. A.; Rupp, B. *J. Mol. Biol.* **2006**, *357*, 1566–1574.
8. Al-Lazikani, B.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1997**, *273*, 927–948.
9. Huhn, C.; Selman, M. H.; Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. *Proteomics* **2009**, *9*, 882–913.
10. Beck, A.; Sanglier-Cianférani, S.; Van Dorsselaer, A. *Anal. Chem.* **2012**, *84*, 4637–4646.
11. Padlan, E. A.; Segal, D. M.; Spande, T. F.; Davies, D. R.; Rudikoff, S.; Potter, M. *Nat. New. Biol.* **1973**, *245*, 165–167.
12. Bertucci, C.; Pistolozzi, M.; De Simone, A. *Curr. Pharm. Biotechnol.* **2011**, *12*, 1508–1516.
13. Jiang, Y.; Li, C.; Nguyen, X.; Muzammil, S.; Towers, E.; Gabrielson, J.; Narhi, L. *J. Pharm. Sci.* **2011**, *100*, 4631–4641.
14. Barth, A. *Biochim. Biophys. Acta* **2007**, *1767*, 1073–1101.
15. Nagarkar, R. P.; Murphy, B. M.; Yu, X.; Manning, M. C.; Al-Azzam, W. A. *Curr. Pharm. Biotechnol.* **2013**, *14*, 199–208.
16. Yamamoto, S. *Anal. Bioanal. Chem.* **2012**, *403*, 2203–2212.
17. Royer, C. A. *Chem. Rev.* **2006**, *106*, 1769–1784.
18. Aubin, Y.; Freedberg, D. I.; Jones, C. *Biopharm. Int.* **2010**, S28+.
19. Wishart, D. S. *Trends Anal. Chem.* **2013**, *48*, 96–111.
20. Kovacs, H.; Moskau, D.; Spraul, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131–155.
21. Fratila, R. M.; Velders, A. H. *Annu. Rev. Anal. Chem.* **2011**, *4*, 227–249.
22. Sorgel, F.; Lerch, H.; Lauber, T. *BioDrugs: Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy* **2010**, *24*, 347–357.
23. Poppe, L.; Jordan, J. B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P. D. *Anal. Chem.* **2013**, *85*, 9623–9629.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

24. Marion, D. *Mol. Cell. Proteomics* **2013**, *12*, 3006–3025.
25. Aubin, Y.; Gingras, G.; Sauve, S. *Anal. Chem.* **2008**, *80*, 2623–2627.
26. Panjwani, N.; Hodgson, D. J.; Sauve, S.; Aubin, Y. *J. Pharm. Sci.* **2010**, *99*, 3334–3342.
27. Sharma, V. K.; Kalonia, D. S. *Pharm. Res.* **2003**, *20*, 1721–1729.
28. Freedberg, D. I. *Dev. Biologics* **2005**, *122*, 77–83.
29. Zuperl, S.; Pristovsek, P.; Menart, V.; Gaberc-Porekar, V.; Novic, M. *J. Chem. Inf. Model.* **2007**, *47*, 737–743.
30. Amezcua, C. A.; Szabo, C. M. *J. Pharm. Sci.* **2013**, *102*, 1724–1733.
31. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371.
32. Meissner, A.; Sorensen, O. W. *J. Magn. Reson.* **1999**, *139*, 439–442.
33. Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 10420–10428.
34. Velyvis, A.; Kay, L. E. *J. Am. Chem. Soc.* **2013**, *135*, 9259–9262.
35. Liu, D.; Ren, D.; Huang, H.; Dankberg, J.; Rosenfeld, R.; Cocco, M. J.; Li, L.; Brems, D. N.; Remmele, R. L., Jr. *Biochemistry* **2008**, *47*, 5088–5100.
36. Kato, K.; Yamaguchi, Y.; Arata, Y. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56*, 346–359.
37. Yamaguchi, Y.; Nishimura, M.; Nagano, M.; Yagi, H.; Sasakawa, H.; Uchida, K.; Shitara, K.; Kato, K. *Biochim. Biophys. Acta* **2006**, *1760*, 693–700.
38. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
39. Schanda, P.; Kupce, E.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199–211.
40. Underbakke, E. S.; Iavarone, A. T.; Marletta, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6777–6782.
41. Zhang, H.; Cui, W. D.; Gross, M. L. *FEBS Lett.* **2014**, *588*, 308–317.
42. Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514–3543.
43. Engen, J. R. *Anal. Chem.* **2009**, *81*, 7870–7875.
44. Gau, B.; Garai, K.; Frieden, C.; Gross, M. L. *Biochemistry* **2011**, *50*, 8117–8126.
45. *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics*; John Wiley & Sons, Inc.: Hoboken, NJ, 2008.
46. Iacob, R. E.; Engen, J. R. *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 1003–1010.
47. Bereszczak, J. Z.; Rose, R. J.; van Duijn, E.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. *J. Am. Chem. Soc.* **2013**, *135*, 6504–6512.
48. Monroe, E. B.; Kang, S.; Kyere, S. K.; Li, R.; Prevelige, P. E., Jr. *Structure*, *18*, 1483–1491.
49. Zhang, Z.; Smith, D. L. *Prot. Sci.* **1993**, *2*, 522–531.
50. Konermann, L.; Pan, J.; Liu, Y.-H. *Chem. Soc. Rev.* **2011**, *40*, 1224–1234.
51. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. *Exp. Rev. Proteomics* **2011**, *8*, 43–59.
52. Rand, K. D.; Zehl, M.; Jensen, O. N.; Jorgensen, T. J. D. *Anal. Chem.* **2009**, *81*, 5577–5584.
53. Mayne, L.; Kan, Z.-Y.; Sevugan Chetty, P.; Ricciuti, A.; Walters, B.; Englander, S. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1898–1905.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

54. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc.* **2009**, *131*, 12801–12808.
55. Zekavat, B.; Miladi, M.; Al-Fdeilat, A. H.; Somogyi, A.; Solouki, T. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 226–236.
56. Rand, K. D.; Jorgensen, T. J. D. *Anal. Chem.* **2007**, *79*, 8686–8693.
57. Rand, K. D.; Adams, C. M.; Zubarev, R. A.; Jorgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130*, 1341–1349.
58. Zehl, M.; Rand, K. D.; Jensen, O. N.; Jorgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130*, 17453–17459.
59. Sterling, H. J.; Williams, E. R. *Anal. Chem.* **2010**, *82*, 9050–9057.
60. Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Anal. Chem.* **2009**, *81*, 7892–7899.
61. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *Anal. Chem.* **2010**, *82*, 8591–8597.
62. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *Anal. Chem.* **2011**, *83*, 5386–5393.
63. Rand, K. D.; Zehl, M.; Jensen, O. N.; Jorgensen, T. J. D. *Anal. Chem.* **2009**, *81*, 5577–5584.
64. Landgraf, R.; Chalmers, M.; Griffin, P. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 301–309.
65. Kan, Z. Y.; Walters, B. T.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16438–16443.
66. Houry, W. A.; Sauder, J. M.; Roder, H.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4299–4302.
67. Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. *Prot. Sci.* **2012**, *21*, 987–995.
68. Guttman, M.; Scian, M.; Lee, K. K. *Anal. Chem.* **2011**, *83*, 7492–7499.
69. Huang, R. Y.; Hudgens, J. W. *Biochemistry* **2013**, *52*, 7127–7136.
70. Houde, D.; Berkowitz, S. A.; Engen, J. R. *J. Pharm. Sci.* **2011**, *100*, 2071–2086.
71. Houde, D.; Arndt, J.; Domeier, W.; Berkowitz, S.; Engen, J. R. *Anal. Chem.* **2009**, *81*, 2644–2651.
72. Houde, D.; Peng, Y.; Berkowitz, S. A.; Engen, J. R. *Mol. Cell. Proteomics* **2010**, *9*, 1716–1728.
73. Zhang, Q.; Willison, L. N.; Tripathi, P.; Sathe, S. K.; Roux, K. H.; Emmett, M. R.; Blakney, G. T.; Zhang, H.-M.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 7129–7136.
74. Rose, R. J.; van Berkel, P. H. C.; van den Bremer, E. T. J.; Labrijn, A. F.; Vink, T.; Schuurman, J.; Heck, A. J. R.; Parren, P. W. H. I. *mAbs* **2013**, *5*, 219–228.
75. Majumdar, R.; Manikwar, P.; Hickey, J. M.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. *Biochemistry* **2013**, *52*, 3376–3389.
76. Sheff, J. G.; Rey, M.; Schriemer, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1006–1015.
77. Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 2132–2139.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

78. Venable, J. D.; Okach, L.; Agarwalla, S.; Brock, A. *Anal. Chem.* **2012**, *84*, 9601–9608.
79. Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R. *Drug Discovery Today* **2014**, *19*, 95–102.
80. Ahn, J.; Engen, J. R. *Chim. Oggi* **2013**, *31*, 25–28.
81. Huang, R. Y. C.; Hudgens, J. W. *Biochemistry* **2013**, *52*, 7127–7136.
82. Zhou, B.; Zhang, Z. Y. *Methods* **2007**, *42*, 227–233.
83. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; Southern, M. R.; Griffin, P. R. *J. Biomol. Technol.* **2007**, *18*, 194–204.
84. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; He, Y.; Hendrickson, C. L.; Marshall, A. G.; Griffin, P. R. *Anal. Chem.* **2006**, *78*, 1005–1014.
85. Iacob, R. E.; Bou-Assaf, G. M.; Makowski, L.; Engen, J. R.; Berkowitz, S. A.; Houde, D. *J. Pharm. Sci.* **2013**, *102*, 4315–4329.
86. Brier, S.; Lemaire, D.; DeBonis, S.; Kozielski, F.; Forest, E. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 456–462.
87. Li, Z.; Huang, R. Y.-C.; Yopp, D. C.; Hileman, T. H.; Santangelo, T. J.; Hurwitz, J.; Hudgens, J. W.; Kelman, Z. *Nucl. Acids Res.* **2014**.
88. Federici, M.; Lubiniecki, A.; Manikwar, P.; Volkin, D. B. *Biologicals* **2013**, *41*, 131–147.
89. Manikwar, P.; Majumdar, R.; Hickey, J. M.; Thakkar, S. V.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. *J. Pharm. Sci.* **2013**, *102*, 2136–2151.
90. Wei, H.; Tymiak, A. A.; Chen, G. D. *Bioanalysis* **2013**, *5*, 1299–1313.
91. Iacob, R. E.; Bou-Assaf, G. M.; Makowski, L.; Engen, J. R.; Berkowitz, S. A.; Houde, D. *J. Pharm. Sci.* **2013**, *102*, 4315–4329.
92. Lilyestrom, W. G.; Shire, S. J.; Scherer, T. M. *J. Phys. Chem. B* **2012**, *116*, 9611–9618.
93. Paul, R.; Graff-Meyer, A.; Stahlberg, H.; Lauer, M. E.; Rufer, A. C.; Beck, H.; Briguet, A.; Schnaible, V.; Buckel, T.; Boeckle, S. *Pharm. Res.* **2012**, *29*, 2047–2059.
94. Wu, Y.; West, A. P., Jr.; Kim, H. J.; Thornton, M. E.; Ward, A. B.; Bjorkman, P. J. *Cell Rep.* **2013**, *5*, 1443–1455.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch002

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Chapter 3

Covalent Labeling Techniques for Characterizing Higher Order Structure of Monoclonal Antibodies

Parminder Kaur,[*,1,2] Janna Kiselar,[1] Wuxian Shi,[1] Sichun Yang,[1] and Mark R. Chance[1,2]

[1]Case Western Reserve University, Cleveland, Ohio 44106, United States
[2]NeoProteomics, Inc., Cleveland, Ohio 44106, United States
[*]E-mail: pxk131@case.edu

The higher order structure of the reference IgG1 is investigated using two complementary irreversible covalent labeling (CL) approaches: hydroxyl radical-based footprinting (HRF), and carbodiimide-based carboxyl group labeling by glycine ethyl ester (GEE) tagging. The chapter outlines how experiments are designed and the results interpreted in order to report biologically useful and functionally relevant information on IgG1 structure. Computational models for the IgG1 are generated based on monoclonal antibodies (mAbs) with known crystal structures used as templates. The goals of the study are to understand the sites of mAb labeling and the potential resolution of the structure assessment; determine the reproducibility of the technique in its application to a mAb; and by establishing a model for the mAb structure, compare the CL data to the model. Some of the key applications of the technique such as epitope mapping and conformational characterization due to process variations are discussed. In addition, pitfalls relating to the experimental design of the CL experiments are highlighted. Future developments for enhancing structural resolution and top-down analysis are discussed and will be instrumental in order to maximize the potential of covalent labeling for mAb drug development and assessment.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Introduction

Discovery and development of monoclonal antibodies (mAbs) is important for a wide variety of molecular immunology studies and in providing novel drugs specifically targeted to binding antigens of clinical interest. In contrast to small molecule drugs, whose structures are relatively simple to characterize, mAbs are significantly larger (~150 kDa), have multiple complex structural dimensions, and exhibit sophisticated folding patterns, leading to ensembles of three-dimensional structures (*1*, *2*). However, just as for small molecular drugs, the structural conformations under physiological conditions determine biological functions and efficacy. Consequently, analytical tools that offer detailed information about higher order structure of mAbs are of great importance to the biopharmaceutical industry. Methods like X-ray crystallography and nuclear magnetic resonance have relatively high resolution but are complex and challenging experiments applicable only to a subset of mAbs or their complexes with antigens due to size, complexity, or solution limitations (*3–5*). Other biophysical techniques such as circular dichroism (CD), fluorescence, and infrared are well suited for rapid assessment but are of very low resolution (*6*, *7*). A range of highly sensitive mass spectrometry (MS)-based methods have been developed to examine the primary, secondary, tertiary, and quaternary structure of mAbs and have proven quite powerful in bridging the gap between high-resolution technologies that are challenging and/or expensive and low-resolution approaches that are inexpensive and rapid but lack specificity (*5*, *8–10*). MS has been successfully employed for assessing three-dimensional structure at medium resolution with high sensitivity. This includes identifying conformational changes as a function of drug or metal-ion binding, identifying protein-protein interfaces, and following folding processes (*11–14*).

Of the MS technologies reviewed in this book, covalent labeling (CL)-coupled MS represents a viable approach to assessing the structure of proteins even in the context of a complex mAb. Hydrogen-deuterium exchange (HDX) consists of *reversible* covalent labeling of the selected protein hydrogen atoms with deuterium. The side chains can exchange very quickly, whereas residues buried inside the protein can take days to exchange. The exchange kinetics for backbone hydrogen atoms range from seconds to minutes, and are of primary interest. As these hydrogens are involved in secondary structure, the method provides a quantitative probe of helical, beta-sheet, and higher order structures (*15*, *16*). Minimizing back exchange for sample processing is optimized by lowering the pH close to 2.5; thus, only proteases with activity at acidic pH (e.g., pepsin) can be used to fragment the protein and then read out the data by bottom-up MS. On the other hand, *irreversible* CL, often called protein footprinting, provides a stable probe of the protein structure. A variety of highly reactive reagents are used in these approaches to probe the solvent accessibility of amino acid side chains, with the reactivity governed by tertiary and quaternary structure and the mechanisms of the labeling reaction. As the modifications are stable in protein footprinting, post-experiment solution conditions and proteases can be flexibly applied with regard to digestion times or pH. In addition, strong reducing agents (such as tris[2-carboxyethyl]phosphine, TCEP) can be used to digest heavily

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

disulfide-bonded species with long incubation times. As for HDX, bottom-up MS is used to read out the mass shifts recording the sites and extents of labeling. The investigation of side chains in the case of protein footprinting is complementary to the HDX method, which probes backbone, secondary structure, and structural stability.

As CL indirectly measures side chain accessibility and thus structure for specific sites of macromolecules, it can be used to assess conformational changes upon binding ligands or in the case of environmental perturbations. Ligand binding to a protein, for example, changes the reactivity of sites in contact with the ligand as well as sites that structurally "communicate" with the site of ligand binding. In addition, in the case of mAbs, environmental influences or manufacturing process variations may perturb or denature the protein locally and change solvent accessibility of side chains as well.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Various irreversible CL approaches—including those mediated by OH radicals (generated using X-ray or electron beam radiolysis, Fenton chemistry, or photolysis of peroxide), carbene labeling, carbodiimide, and diethylpyrocarbonate reagents—are routinely employed in CL experiments (*17*). Labeling reactions can be carried out with either highly specific amino acid side chain labels or non-specific labels. Specific labeling approaches target a small subset of the 20 amino acid side chains, whereas non-specific labels, such as hydroxyl radicals, label 15 or more of the 20 side chains (called probes) in a typical experiment.

Figure 1 shows a typical workflow for a PF experiment for higher order protein studies as applied to mapping epitope interfaces. In this method, chemical labels (denoted by circles surrounding the antibody and the antigen) introduced to the solution containing the target antigen impart covalent modifications to its amino acid side chains. This experiment is separately carried out for both the antigen and mAb in isolation, and for the protein complex of interest (e.g., antigen-mAb complex in this case). The labeled samples are proteolytically digested. The solvent-accessible regions of the protein are labeled, as indicated by the circles on the digested peptides, and the relevant peptides are mass shifted. The regions at the binding interface are not exposed to the solution and experience "protection" from the labels, as shown by the unlabeled yellow and green peptides (from the antigen and antibody, respectively) in the case of antigen-antibody complex. The digested sample is subjected to liquid chromatography coupled with mass spectrometry (LC-MS). Tandem MS (MS/MS) is used to identify the peptides and localize the sites of modification. The intensities from the selected ion chromatograms (SICs) are integrated to compute the abundance of each peptide form. The fraction of the unmodified peptide is calculated from the integrated intensity values of the modified and unmodified forms of the peptide. Dose-response (DR) curves are generated to monitor the loss of the unmodified fraction as a function of oxidation exposure time. The amount of labeling of a given region is a function of solvent accessibility, the inherent reactivity of the constituent residues, and the solution conditions (*17*, *18*).

MS has been increasingly gaining attention as a viable method to interrogate the biochemical and biophysical properties of biomolecules (*19–23*). Various MS approaches have been employed in multiple phases of antibody and biologics development to characterize primary structure, glycan structure, and various

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

post-translational modifications, including those generated by in-process and in-storage conditions (*8*, *24*). Applications of hydroxyl radical-based protein footprinting for biophysical characterization of mAbs and biologics is quite recent (*25*), although the technique of protein footprinting-based structural characterization of proteins using MS is now over 10 years old (*26*). The technique has been successfully applied to determine the interface regions of a mAb dimer, study conformation variations of expired forms of granulocyte colony-stimulating factor (GCSF), observe conformational differences of mAb IgG2 isomers, and for epitope mapping of serine protease thrombin (*27–30*). Alternative protein labeling strategies also are suitable for characterizing mAbs. One of the popular techniques employs carbodiimide-based carboxyl group labeling with glycine ethyl ester (GEE) tagging that targets the solvent-accessible C-terminal Glu, and Asp residues (*31*, *32*). This labeling approach has been applied to probe multiple proteins, including the mammalian polyamine transport system; the membrane-attached antenna protein, including mapping a protein-protein interface; and the phosphorylation-induced structural changes of a membrane-associated kinase; and very recently has been used to characterize a mAb (*33–39*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

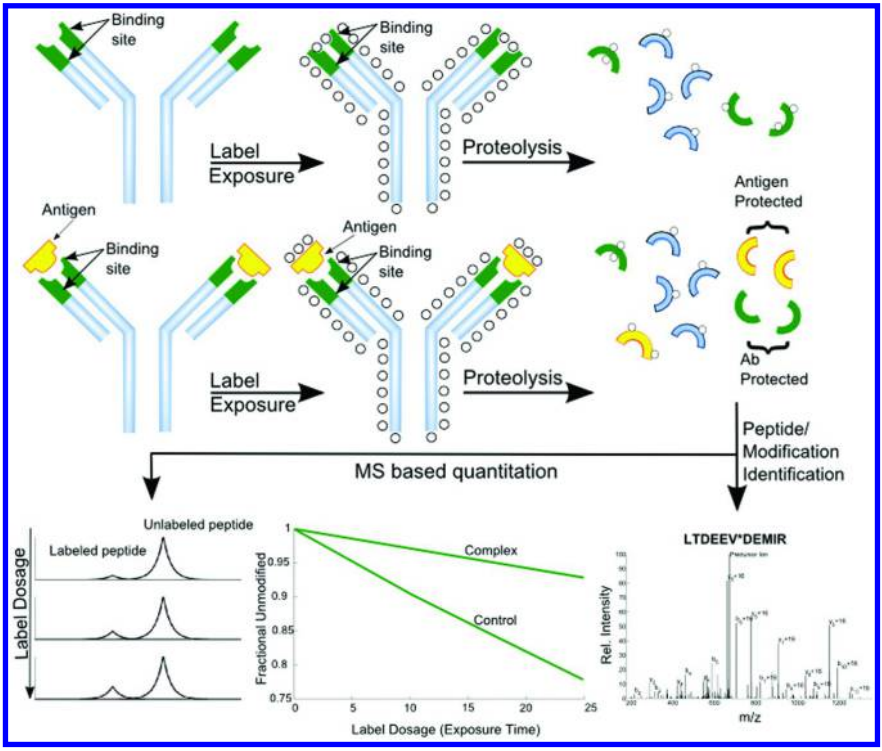

Figure 1. Workflow for a typical covalent labeling experiment as applied to mapping epitope interfaces. (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

In this chapter, the structure of the reference IgG1, in terms of the reactivity of its side chains to irreversible CL, is characterized using two complementary protein footprinting approaches: (a) hydroxyl radical-based protein footprinting (also called hydroxyl radical footprinting [HRF]) and (b) carbodiimide-based carboxyl group labeling (with GEE tagging). The goals are to define the sites of labeling and thus the potential resolution of the structure assessment; determine the reproducibility of the technique in its application to a mAb through the comparison of replicates; and compare the protein footprinting data to a theoretical model of the mAb structure. These goals will provide a baseline for potential future studies where the structure and function of the mAb is perturbed to assess precisely the sources of structure variation at the side chain level.

In the first approach, synchrotron X-ray radiation was used to irradiate aqueous solutions containing the IgG1, producing a high flux of radicals on millisecond timescales. These reactive hydroxyl radical species attack the solvent-accessible amino acid side chains of the protein, hence imparting stable oxidative modifications on the 15 or more residues that are typically targeted. In the second approach, carboxyl group labeling used 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) in the presence of GEE to label solvent-accessible C-terminal carboxyl groups, plus the side chain carboxyls of Glu and Asp residues of the protein on timescale of minutes. Because the Glu and Asp residues are inefficiently labeled in the case of hydroxyl radical-based approaches, the two methods offer complementary information. The chapter outlines how experiments are optimized to preserve the structural integrity of the protein and report biologically useful information on structure. Three replicates of the experiments in the case of hydroxyl radical-based CL, and two replicates for carboxyl group labeling, are performed in order to determine the reproducibility of the results. The protein sequence is used to generate computational models based on homology with mAbs with known crystal structure. Finally, the labeling results from the two approaches are mapped to the homology-based models. The results from modeling and the experimental data are compared against each other in order to understand the similarities and differences observed in the two cases. These labeling patterns provide a distinctive assessment of the IgG1, highlighting a number of useful pharmaceutical applications of protein footprinting technologies, including epitope/paratope mapping and overall structure assessment for determining the quality and reproducibility of specific manufactured drug lots.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Methods

Protein Labeling

The NISTmAb sample was buffer exchanged into 10mM phosphate buffer, 150 mM NaCl, pH 7.4 using a Thermo Scientific Pierce concentrator with a molecular-weight cutoff (MWCO) of 10,000. For HRF experiments, a 5 µl sample of the mAb (5 µM) was exposed to the X-ray beamline X-28C of the National Synchrotron Light Source at Brookhaven National Laboratory (Long Island, NY) for 0 to 15 milliseconds at ambient temperature. All experiments were performed

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

at the ring energy of 2.8 GeV and at the beam current between 212 and 200 mA. To quench secondary oxidation of methionine, 10 mM Met-NH_2×HCl buffer (pH 7.0) was added immediately after irradiation to all samples.

For carboxyl group labeling experiments, 10 µg of the NISTmAb at the concentration of 1 µg/µl was taken for each labeling reaction. Stock solutions of GEE and EDC (Thermo Fisher Scientific, Rockford, IL) were made in 10mM phosphate buffer, 150 mM NaCl, pH 7.4, at a concentration of 1M and 0.025 M, respectively. GEE and EDC were added to the reaction vial so their final concentration was 210 mM and 7 mM, respectively. The labeling reaction was carried out at room temperature for 0 and 10 min. The reaction was quenched by the addition of 1% formic acid (Thermo Scientific, Rockford, IL) to a final concentration of 0.1% (v/v).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Deglycosylation and Digestion

Radiolyzed mAb samples at 3.7 µg of total protein per sample were deglycosylated using 1 µL (500,000 U/mL) of Peptide-N-Glycosidase F (PNGase F) enzyme (New England Biolabs, Co.) for 30 min at 37 °C, and precipitated with 10% trichloroacetic acid/acetone overnight. These protein samples then were washed with cold acetone 3 times and air dried. Protein samples were then reconstituted in 50 mM Tris, 8 M urea buffer (pH 7.8), reduced with 10 mM dithiothreitol (DTT) for 45 min at 37 °C and alkylated with 25 mM iodoacetic acid (IAA) for 45 min at room temperature in the dark. Protein samples were digested with Lys-C for 3 h at 37 °C at 1:20 w/w enzyme to protein ratio followed by trypsin digestion at 1:20 w/w enzyme to protein ratio at 37 °C overnight.

The GEE-labeled mAb samples were buffer exchanged two times with 20 volumes excess of 8 M urea and 40 mM DTT using a 0.5 mL 3000 MWCO filter (Millipore, Billerica, MA), concentrated to approximately 50 µL, and then deglycosylated at 37°C for 1 h on the filter using 2 µL (500,000 U/mL) of PNGase F (New England Biolabs, Co.). Next, protein samples were buffer exchanged two times with 20 volumes excess of 50 mM Tris, 2 M urea buffer (pH 7.8), concentrated to approximately 50 µL, and then reduced and alkylated with DTT and IAA, respectively, on the filter. Finally, mAb samples were transferred to Eppendorf Tubes and digested with Lys-C and trypsin at 37 °C overnight.

Mass Spectrometry

Peptide mixtures derived from the hydroxyl-labeled and GEE-labeled proteins were separated by reversed phase high pressure liquid chromatography (HPLC) using a Waters nanoAcquity LC system (Waters, Taunton, MA) and a gradient formed with mobile phase A (100% water with 0.1 % formic acid) and mobile phase B (100% acetonitrile [ACN] with 0.1% formic acid). The gradient program consisted of ramping mobile phase B from 2 to 40% over a period of 60 min, 40% to 43% over a period of 5 min and 43% to 83% over a period of 10 min at ambient temperature and a flow rate of 300 nl/min. Peptides eluted from the column packed with C18 BEH130, 1.7µm, 130Å (Waters, Taunton, MA) were detected by an Orbitrap Elite mass spectrometer (Thermo Fischer Scientific, San

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Jose, CA) equipped with a nanospray ion source using a needle voltage of 2.4 kV. A full MS1 scan was obtained for eluted peptides in the range of 380 to 1800 *m/z* followed by twenty data-dependent MS/MS scans. MS/MS spectra were generated for peptides with a minimum signal of 2000 by collision-induced dissociation of the peptide ions at normalized collision energy of 35%, an isolation width of 2.5, and an activation time of 30 ms to generate a series of b- and y-ions as major fragments.

Data Analysis

Data analysis for both hydroxyl radical labeling and the carboxyl group labeling experiments was performed using ProtMapMS, a commercial software package developed by NeoProteomics Inc. (Cleveland, OH) specifically for the automated analysis of covalent labeling experiments (*40*). ProtMapMS uses the tandem MS data for identification and localization of labeled residues. It extracts single ion chromatograms from the MS data for quantifying the extent of modification. Areas under the SIC plots are used to construct DR curves for each labeled peptide as a function of label exposure time. The resulting DR curves are used to calculate the rate constants for the oxidation reaction.

Structural Model for NISTmAb

A theoretical model of the NISTmAb was developed based on combining representative IgG templates from the protein data bank (PDB). An Fab/antigen complex solved to a high resolution of 1.9 Å (***PDB ID: 3QWO***) was used to model the intact NISTmAb Fab region (*41*, *42*). A caveat of the template for modeling is that there are potential structural differences between the "naked" Fab and the Fab/antigen complex. The Fc region was modeled using the HIV glycoprotein-specific IgG antibody b12, which has previously been solved to 2.7 Å by X-ray crystallography (***PDB ID: 1HZH***) (*43*)**.** The overall arrangement of the antibody was modeled based on 1IGY. There are only a few full IgG antibody crystal structures in the PDB bank, and they showed very different orientations of Fabs versus Fc. All of these structures represent structural snapshots of flexible IgG antibodies. 1IGY was chosen in this case as it demonstrates a moderate, roughly symmetrical conformation. It provides a structural model for one of the conformations of the antibody in solution. Homology modeling was performed using the Swiss-Model, an automated protein homology-modeling server (*44*).

To generate a full model of the NISTmAb, two copies of Fabs and one copy of Fc were overlaid on the Fabs and Fc in a 3.2 Å resolution crystal structure of an intact IgG1 structural template (***PDB ID: 1IGY***) (*45*). The hinge regions (residues 220–240) were manually built and regularized/energy minimized using the program named COOT (*46*). The sequence identity of the NIST-mAb hinge and the 1IGY hinge is 21%. In summary, the final mAb structure model contains two Fab domains based on 3QWO, one Fc domain based on 1HZH, and an overall IgG Fab/Fc arrangement based on 1IGY. The model can be obtained at http://proteomics.case.edu/pkaur/models/nistref.pdb.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Results

Labeling and Deglycosylation

The introduction of stable, irreversible modifications during protein footprinting allows the labels to be maintained during post-labeling sample handling. This robustness is particularly important for characterizing mAbs that are often heavily glycosylated. However, the presence of diverse glycan modifications can interfere with trypsin digestion due to the steric hindrance by the glycan units (*47*). In addition, it often adds to the complexity of the resulting mass spectra since the high intensity of glycan fragments interferes with the detection of relatively weaker peptide fragment ions. Therefore, a deglycosylation step is added before trypsin digestion in order to increase the sequence coverage and facilitate peptide identification. It is important to note that since the labeling is performed on the originally glycosylated form of the protein prior to the deglycosylation step, the derived structural information pertains to the glycosylated form.

Hydroxyl Radical Footprinting

The trypsin-digested light and heavy chains of the protein were identified with a total sequence coverage of 85.0% and 86.7%, respectively. This included 11 and 25 peptides along the light chain and heavy chain, respectively. The missing regions consisted entirely of short peptides with lengths smaller than 5 residues. Such peptides are typically difficult to detect due to their weaker retention on the chromatography column. Of the 36 (11+25) detected peptides, 34 were also identified as containing oxidative modifications resulting from hydroxyl radical exposure. Nearly 22.6% (150/663) of the total residues across the protein were confirmed to be oxidized. The 150 oxidatively modified probes were comprised of 48 and 102 residues on the light and heavy chains, respectively. These oxidative modifications were detected on 18 unique amino acids across the protein, including Ala, Cys, Asp, Glu, Phe, His, Ile, Lys, Leu, Met, Pro, Gln, Arg, Ser, Thr, Val, Trp, and Tyr. This included a range of common oxidative mass shifts such as +14, +16, +32, and +48, which arise from common hydrogen abstraction and radical attack chemistries when oxygen is present (*18*). We also observed the usual range of less abundant modifications arising from the more complex set oxidation reactions (e.g., for Arg, Asp, Glu, His, Val) such as −43, −30, −22, and +5. The relative reactivity of the side chains under aerobic conditions using MS methods for the detection of the oxidation has been previously established to be the following: Cys > Met > Trp > Tyr > Phe > Cystine > His > Leu ~ Ile > Arg ~ Lys ~ Val > Ser ~ Thr ~ Pro > Gln ~ Glu > Asp ~ Asn > Ala > Gly (*18*).

ProtMapMS-extracted SICs of the labeled and unlabeled peptide forms were prepared from the MS data using a window of 15 ppm around the precursor *m/z* value of the corresponding monoisotopic peaks. For peptides with molecular weight >4 kDa, the SICs are extracted for the most abundant isotope because the intensity of monoisotopic peak drops significantly. This was followed by the calculation of areas under the SICs. The area under the SIC curves for all oxidative forms are summed together to calculate the amount of total oxidation. The fraction

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

of the unlabeled peptide, compared to the sum total area of all of its oxidative and non-oxidative forms, was calculated at each exposure time for the most abundant charge state. These values were used to plot the fraction of the unlabeled peptide versus exposure time in the form of a DR curve. This curve helps to confirm the pseudo-first order reaction kinetics that the labeling reaction is expected to follow; this reaction is described by equation 1:

$$y(t) = e^{-kt} \quad (1)$$

where y(t) is the fraction of the unlabeled peptide, k is the rate constant (RC) in s^{-1}, and t is the exposure time in seconds. Deviation from first order kinetics may be observed at increased exposure times and indicates overoxidation of the sample, compromising the native conformation. The longest time points that evidence such phenomenon should be removed from the curve to provide reliable data.

In order to examine the kinetics of labeling and to assess the reproducibility of the experiments, the DR plots from three replicate experiments are compared. Figure 2 shows four examples of DR plots for peptides from different regions: L(1-18), H(125-136), H(348-358), and H(343-350), along with the best fit to the data (solid line). The X-axis shows that the mAb sample was exposed to hydroxyl radicals for 0, 10, and 15 milliseconds. The percentage of the peptide that remained unoxidized is plotted on the Y-axis. Each of the data points is normalized against the unexposed sample (X = 0 ms) in order to correct for any background oxidation. The circles at each time point indicate results from three different experiments with identical labeling conditions. The resulting RCs range from 1.3 to 4.4 s^{-1} as calculated from a best fit according to Equation 1. Each of the values on the DR plots fall within 1% of the absolute value of the fit (10% relative error in RC values), showing the level of reproducibility of the experiment.

Table 1 shows a summary of tryptic peptides of IgG1 detected to have oxidation upon exposure to hydroxyl radicals. The first and second columns indicate the peptide location and the representative sequence, respectively. The third column indicates the peptide-level rate constant of oxidation along with deviations calculated from three replicates. The fourth column shows the protection factor values (more details in the following section), and the last column denotes the oxidized residues detected by tandem MS and verified manually. The parentheses in the last column represent a theoretically predicted value of the fraction of the solvent-accessible side chain surface area (fSASA) of the labeled residues. This value was calculated from the homology-based model using VADAR (*48*). Theoretical fSASA values range between 0 and 1, with the respective values indicating that 0 and 100% of the surface area of a given side chain is expected to be exposed to the solvent. For example, the side chain of Leu is known to have accessible surface area of 137 $Å^2$ (*49*). An fSASA value of 0.20 indicates that the homology model shows that only 27.4 $Å^2$ (137 × 0.20) of the accessible surface area is exposed to the solvent. This calculation normalizes the surface area calculation for side chains that vary in total accessible area. Amino acid residues with larger fSASA values are expected to correspond to larger experimentally determined RC values (more rapid labeling kinetics). Such

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

a comparison is useful for evaluating the correlation between a theoretical model and experimental results.

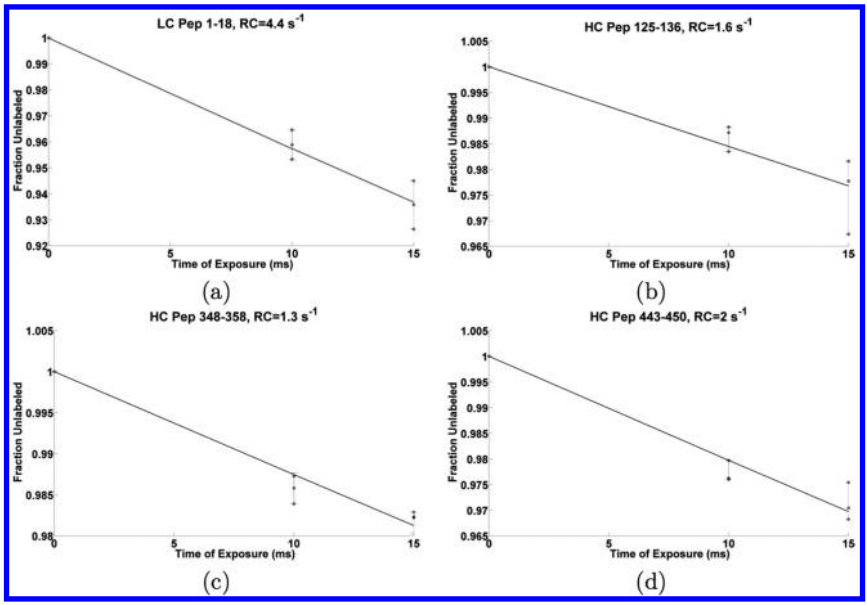

Figure 2. Dose-response plots from the HRF experiment: (a) light chain peptide 1-18; (b) heavy chain peptide 125-136; (c) heavy chain peptide 348-358; and (d) heavy chain peptide 443-450.

Table 1. Results from Hydroxyl Radical Footprinting Experiments

Peptide Affected	*Peptide Sequence*	*RC (HRF) (s-1)*	*Protection Factor*	*HRF Labeled Residues (Fractional SASA)*
H(6-13)	ESG**P**A**L**V**K**	1.7±0.1	9.4±0.7	P9 (0.49), L11 (0.12), K13 (0.59)
H(6-40)	ESG**P**A**L**V**K**PTQ TLTLTCTFSG**F**S LS**T**AG**M**SVGW IR	10.4±2.3	13.6±3.0	P9 (0.49), L11 (0.12), K13 (0.59), F27 (0.12), T31 (0.66), M34 (0.05)
H(46-59)	AL**E**WLAD**I**W**W** DDK**K**	13.6±0.7	5.3±0.3	E48(0.32), I53 (0.00), W55 (0.25), K59 (0.31)
H(60-66)	**HY**N**P**SL**K**	2.6±0.5	12.0±2.4	H60 (0.21), Y61 (0.14), P63 (0.46), K66 (0.66)
H(67-73)	D**R**L**T**ISK	1.3±0.2	13.6±2.1	R68 (0.19), T70 (0.65)
H(69-77)	LTISKDTSK	0		None
H(78-83)	NQVVLK	0		None

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Table 1. (Continued). Results from Hydroxyl Radical Footprinting Experiments

Peptide Affected	*Peptide Sequence*	*RC (HRF) (s-1)*	*Protection Factor*	*HRF Labeled Residues (Fractional SASA)*
H(84-99)	VTN**M**D**P**A DTATYYCAR	0.9±0.0	95. 8±4.3	M87 (0.00), P89 (0.55)
H(100-124)	D**MIF**N**F**YFDV WGQGTTVTVS SASTK	42.5±0.2	2. 6±0.0	M101 (0.27), I102 (0.44), F103 (0.69), F105 (0.32),
H(125-136)	GPSV**F**PL**A**P**S**SK	1.6±0.4	17.6±4.6	F129 (0.12), A132 (0.46), S134 (0.47)
H(137-150)	S**T**SGG**T**AALG **C**LVK	0.7±0.03	71.5±3.2	T138 (0.27), T142 (0.34), C147 (0.00)
H(151-208)	DYFPEPVTVSW NSGALTSGVHT FPAVLQSSGLY SLSSVVTVPSSS LGTQTYICNVN HK	8.7±0.1	23.1±0.3	Modification sites could not be localized due to insufficient fragmentation
H(226-249)	**THTCPPCP**A**P**E **LL**GGPSV**F**L**F**P P**K**	11.0±0.2	10.9±0.2	T226 (0.39), H227 (0.8), T228 (0.72), C229 (0.27), P230 (0.72), P231 (0.54), C232 (0.63), P233 (0.76), P235 (0.59), L237 (0.68), L238 (0.65), F244 (0.55), F246 (0.55), K249 (0.61)
H(252-258)	DTL**MI**S**R**	29.9±0.7	1.2±0.0	M255 (0.27), I256 (1.00), R258 (0.36)
H(259-277)	TPEVT**C**VVV**D** VS**H**E**DP**EV**K**	6.9±0.6	9.0±0.8	C264 (0.00), D268 (0.52), H271 (0.38), D273 (0.32), P274 (0.17), K277 (0.76)
H(278-291)	**F**N**WYV**DG**V** E**VH**NA**K**	6.2±0.2	9.6±0.3	F278 (0.06), W280 (0.01), Y281 (0.19), V282 (0.18), V285 (0.88), V287 (0.23), H288 (0.78), K291 (0.87)
H(292-304)	**TKPR**EEQ**Y**N STY**R**	10.4±1.1	3.8±0.4	T292 (0.39), P294 (0.95), R295 (0.52), Y299 (1.00), R304 (0.33)
H(305-320)	VVS**V**LTV**LHQ** D**WL**NG**K**	1.6±0.2	34.6±4.4	V308 (0.11), L312 (0.77), H313 (0.29), Q314 (0.87), W316 (0.04), L317 (0.21), K320 (0.44)
H(330-337)	AL**PA**P**I**E**K**	2.2±0.2	6.4±0.6	P332 (0.45), P334 (0.36), K337 (0.49)

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Results from Hydroxyl Radical Footprinting Experiments

Peptide Affected	*Peptide Sequence*	*RC (HRF) (s-1)*	*Protection Factor*	*HRF Labeled Residues (Fractional SASA)*
H(348-358)	EPQV**Y**TL**P**PS**R**	1.3±0.0	22. 9±0.9	Y352 (0.06), P355 (0.32), R358 (0.7)
H(364-373)	**NQ**VSLT**C**LV**K**	1.0±0.2	45.0±8.4	N364 (0.89)/Q365 (0.56), C370 (0.00), K373 (0.22)
H(374-395)	GFYPSDIAV**E**W ES**N**G**QP**ENN**YK**	2.2±0.1	32.5±1.5	E383 (0.28), N387 (0.88), Q389 (0.76), P390 (0.65), Y394 (0.18), K395 (0.52)
H(396-412)	TT**PP**V**L**D**S**DGS FFLYSK	1.9±0.3	29.7±4.6	P398 (0.65)/P399 (0.29), L401 (0.58), S403 (0.57)
H(420-442)	W**Q**QG**N**VFS**C**S V**M**HEALHN**HY** TQK	23.3±0.5	5.9±0.1	Q421 (0.47), N424 (0.36), C428 (0.00), M431 (0.04), H438 (0.50), Y439 (0.40)
H(443-450)	SLSLS**P**G**K**	2.0±0.2	8.0±0.8	P448 (NA), K450 (NA)
L(1-18)	**DI**QM**T**QS**P**ST**L** SASVG**D**R	4.4±0.4	10.5±1.0	D1 (0.51), I2 (0.22), T5 (0.87), P8 (0.46), L11 (0.18), D17 (0.54)
L(19-28)	**VT**IT**C**SASS**R**	1.9±0.1	24.2±1.3	V19 (0.09), T20 (0.42), T22(0.38), S24 (0.38), R28 (0.79)
L(29-38)	VG**Y**MHWYQQ**K**	0.4±0.1	191. 8±33.6	Y31 (0.57), K38 (0.21)
L(45-52)	**L**LI**Y**DTS**K**	1.2±0.1	25.1±2.0	L45 (0.07), Y48 (0.25), K52 (0.56)
L(53-60)	**L**ASGV**PS**R	2.5±0.7	5.3±1.5	L53 (0.30), P58 (0.50), R60 (0.24)
L(61-102)	FSGSGSGTEFT LTISSLQPDDFA TYYCFQGSGYP FTFGGGTK	5.3±0.4	32.4±2.4	Modification sites could not be localized due to limited fragmentation
L(108-125)	**TV**AAPSVFIF**P**P S**D**EQ**LK**	3.0±0.1	15.6±0.3	T108 (0.96), V109 (0.39), P118 (0.42), D121 (0.73), L124 (0.16), K125 (0.75)
L(126-141)	SGTASVV**C**LLN N**FY**P**R**	0.9±0.0	82.7±0.9	C133 (0.00), F138 (0.00), P140 (0.21), R141 (0.53)
L(149-168)	VDNA**L**QSGNS **QESVTEQDSK**	5.0±0.4	4.6±0.9	L153 (0.46), E160 (0.39), V162 (0.25), E164 (0.43), D166 (0.34), K168 (0.91)

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Table 1. (Continued). Results from Hydroxyl Radical Footprinting Experiments

Peptide Affected	*Peptide Sequence*	*RC (HRF) (s-1)*	*Protection Factor*	*HRF Labeled Residues (Fractional SASA)*
L(169-182)	DSTYS**L**SSTL **TLSK**	1.2±0.1	33.2±2.8	L174 (0.06), T179 (0.39), L180(0.13), K182 (0.40)
L(190-206)	VYA**CE**V**T**H**Q**G **L**SS**PVTK**	3.2±0.1	22.3±0.4	C193 (0.00), E194 (0.20), T196 (0.50), Q198 (0.72), L200 (0.14), P203(0.44), V204 (0.35), T205 (0.47), K206 (0.38)

Note: labeled residues marked in bold letters. * Site of glycosylation. Abbreviations: HRF = hydroxyl radical footprinting, RC = rate constant, SASA = solvent-accessible surface area.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

It is apparent from Table 1 that there is a significant variation in the rates of oxidation (ranging between 0 to 42.5 s^{-1}) across different protein segments. Rate constant of 0 indicates none to minimal oxidation of the peptide, where any signal from oxidized peptide is below the detection limit of the instrument. The individual rate constant values fall within 10% of their average values from the triplicate experiments, demonstrating the technical variation expected for the technique. We compared the observed labeled residues against the homology model to understand the thresholds of solvent accessibility where labeling is observed to occur under these conditions. For example, nearly 88% of the labeled residues in Table 1 have fSASA values greater than 0.10; this means that the nearly all of the labeled residues have at least 10% of their side chain surface areas exposed to the bulk solvent, allowing for attack by the hydroxyl radicals. In addition, the labeled residues were found to have a median of 40% of their side chain surface area accessible to the solvent. The observed rate of oxidation of a given region is a function of its solvent accessibility, the reactivity of the constituent residues, and the relative ionization efficiencies of the various oxidative and non-oxidative isoforms. We emphasize that different amino acid residues are expected to have different thresholds for a minimum fSASA value in order to be observed to be oxidized. Thus, it is important to be cautious in comparing the absolute values of rate constants for two different peptides within the same protein in an experiment, as they may have widely varying intrinsic reactivity. On the other hand, comparisons can be reliably made for the reaction kinetics of a given peptide under different conformational states of the same protein as could be performed in future studies with stressed material or during epitope mapping studies, as described in Figure 1. The current data is a "baseline" measure of the native molecule, and with appropriate considerations, some general features can be highlighted as discussed below.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Low Oxidation Rates (Rate Constant < 1.0)

Peptide segments H(69-77) and H(78-83) were not modified. The model revealed that both of these regions are represented by inner beta strands of the Fab fragment. The H(69-83) region consists primarily of low reactive residues, with the exception of three moderately reactive residues (L69, I71, and L82), each of which exhibits an fSASA value 0.0. The low solvent accessibility of reactive residues, in conjunction with the low reactivity of the relatively accessible residues, are consistent with the lack of observed labeling. The peptides H(137-150) and L(29-38) were also found to form inner beta strands. Peptide L(29-38) primarily contains probes shielded from the solvent (fSASA < 0.03), with the exception of Y31 (fSASA = 0.57) and K38 (fSASA = 0.21) residues, leading to small extent of oxidation. Peptide H(137-150) is comprised of low solvent-accessible probes (fSASA < 0.08), with the exception of moderately solvent-accessible (fSASA > 0.25) but low-reactive T138, S139, and T142 residues. The overall low solvent accessibility of reactive residues, in conjunction with the low reactivity of the relatively accessible residues, leads to minimal oxidation in these regions, as evidenced by their rate constants representing the smallest values in Table 1.

These peptides serve as a representative example that fSASA values must be considered in combination with residue reactivities in understanding the observed oxidized peptides. Despite different theoretical fSASA values, each of these peptides showed low RCs. For this reason, the protection factor (PF) value discussed below has been introduced.

High Oxidation Rates (Rate Constant > 15)

The highest rate constant values in Table 1 are represented by the peptides H(100-124), H(252-258), and H(420-442). Peptide H(100-124) contains highly reactive and solvent-accessible Met and Phe residues, leading to high oxidation. Similarly, both peptides H(252-258) and H(420-442) contain multiple reactive residues, including minimally solvent-accessible but very reactive Met (fSASA > 0.03) in addition to other highly solvent-accessible probes such as I256 (fSASA = 1.00), H438 (fSASA = 0.50), and Y439 (fSASA = 0.40). In addition, the structural model showed that a significant portion of each of these three peptides spans flexible regions such as loops, allowing for greater possibility of being exposed to the solvent in solution. Although the solvent accessibility of Met is moderate in each case, the high reactivity of Met oxidation contributes to the high rates of oxidation observed for the three peptides (*50*, *51*).

Outliers

Some of the notable exceptions between the experimental data and the homology model include peptides containing highly reactive heavy (H) chain residues M87, C147, C264, C370, C428; and light (L) chain residues C133 and C193. Although the fSASA values for each of these residues is less than 0.01, indicating low solvent accessibility, peptides containing these residues were

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

labeled, and the site of modification was localized to the above residues. It would be desirable to investigate the oxidation reaction kinetics of the individual residues in order to resolve their respective extent of oxidation. Because the quantitative characterization is performed at the level of the peptide, it is difficult to estimate the contribution from individual residues when there are multiple oxidized residues per peptide. However, low rate constants for peptides with fewer probes (number of probes ≤ 4, such as for H[137-150], H[364-373], L[126-141]) suggest that each of these sites experienced minimal oxidation. Another possible explanation for their observed labeling may lie in the alternative oxidation mechanism for sulfur-containing residues that is mediated by hydrated electrons (*50*). Other outliers with low solvent accessibility that were labeled include I53, W280 from the H chain and F138 from the L chain. This could suggest some differences between the theoretical model and solution state experimental structure (*52*, *53*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Results from Carboxyl Group Labeling

The trypsin-digested fragments of the L and H chains of the NISTmAb were identified with total sequence coverage of 85.0% and 85.8% respectively, including 11 and 24 peptides from the L and H chains respectively. Table 2 shows a summary of results from carboxyl group labeling experiments. Columns 1 and 2 show the locations of the tryptic peptide and the corresponding sequence respectively, with the heavy and light chains indicated by the letters H and L in the first column. The third column indicates the GEE-labeled residues, with the parentheses indicating the corresponding fSASA values as determined from the homology-based model. The last column shows the target probes (D/E/C-terminus) that were not observed to be labeled during the experiment along with the corresponding fSASA values. The detected peptides in Table 2 represent a total of 47 D/E residues, of which 28 were labeled by the GEE tag. We compared the trend of labeling against the solvent accessibility profile generated by the model. For example, none of the heavy chain residues (E6, D52, D81, D91, D100) with fSASA values less than 0.1 were observed to be labeled. In addition, 76% (13/17) residues (H chain: D88, E236, D268, E272, E275, E297, E348; L chain: D1, D17, E69, D80, D121, D169) with fSASA ≥ 0.50 observed to be labeled. Only 2 H chain (D67, D404) residues with fSASA > 0.75 were not labeled. Closer examination of the data revealed that H chain D67 is represented by a peptide that includes a missed cleavage (DRLTISK), and the MS1 signal intensity for this peptide was 10 times smaller than the average signal from other peptides. Since the labeled forms are typically present at sub-stoichiometric concentrations, it is possible that the signal from the labeled form of the peptide was below the detection limit of the instrument, and hence, it could not be detected. Thus, only one of the highly solvent-accessible probes (H chain D404) did not get labeled. On the contrary, some of the less solvent-accessible (fSASA < 0.20) residues, such as H chain D315 and E433 and L chain D49, were observed to be labeled. This could suggest that there are some differences between the solution form of the protein and the homology model that is derived from the existing crystal structures (*52*, *53*). Although the labeling status of a given residue

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

is helpful at determining whether or not it is accessible to the solvent, numeric values of the rate constants provide insights into determining the extent of solvent exposure. For example, as a future experiment, the reaction kinetics can be characterized by varying the time of the labeling reaction, as in the case of HRF experiment.

Table 2. Results from Carboxyl Group Footprinting Experiments

Peptide Affected	*Peptide sequence*	*GEE-Labeled Residues (Fractional SASA)*	*Unlabeled Residues (Fractional SASA)*
H(6-13)	ESGPALVK	None	E6 (0.00)
H(6-40)	ESGPALVKPTQTLTLTCTFS GFSLSTAGMSVGWIR	None	E6 (0.00)
H(46-59)	ALEWLADIWWDDKK	None	E48 (0.32), D52 (0.00), D56 (0.51), D57 (0.53)
H(60-66)	HYNPSLK	Detected, but no potential probes	
H(67-73)	DRLTISK	None	D67 (0.77)
H(78-83)	NQVVLK	Detected, but no potential probes	
H(84-99)	VTNM**D**PADTATYYCAR	D88 (0.57)	D91 (0.02)
H(100-124)	DMIFNFYFDVWGQG TTVTVSSASTK		D100D100 (0.00), D108 (0.20)
H(125-136)	GPSVFPLAPSSK	Detected, but no potential probes	
H(137-150)	STSGGTAALGCLVK	Detected, but no potential probes	
H(151-208)	DYFPEPVTVSWNSGALT SGVHTFPAVLQSSGLYSL SSVVTVPSSSLGTQTYICN VNHK		
H(226-249)	THTCPPCPAP**E**LLG GPSVFLFPPK	E236 (0.56)	None
H(252-258)	**D**TLMISR	D252 (0.16)	None
H(259-277)	TP**E**VTCVVV**D**VSH**E**DP**E**VK	E261 (0.38), D268 (0.52), E272 (0.78), E275 (0.80)	D273 (0.32)

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Table 2. (Continued). Results from Carboxyl Group Footprinting Experiments

Peptide Affected	*Peptide sequence*	*GEE-Labeled Residues (Fractional SASA)*	*Unlabeled Residues (Fractional SASA)*
H(278-291)	FNWYV**D**GV**E**VHNAK	D283 (0.49), E286 (0.37)	None
H(292-304)	TKPR**EE**QYNSTYR	E296 (0.48), E297 (0.50)	None
H(305-320)	VVSVLTVLHQ**D**WLNGK	D315 (0.19)	None
H(330-337)	ALPAPI**E**K	E336 (0.42)	None
H(348-358)	**E**PQVYTLPPSR	E348 (0.67)	None
H(364-373)	NQVSLTCLVK	Detected, but no potential probes	
H(374-395)	GFYPS**D**IAV**E**W**E**SNGQ PENNYK	D379 (0.34), E383 (0.28), E385 (0.28)	E391 (0.12)
H(396-412)	TTPPVLDSDGSFFLYSK		D402 (0.13), D404 (0.75)
H(420-442)	WQQGNVF-SCSVMH**E**ALHNHYTQK	E433 (0.13)	
H(443-450)	SLSLSPGK	Detected, but no potential probes	
L(1-18)	**D**IQMTQSPSTLSASVG**D**R	D1 (0.51), D17 (0.54)	None
L(19-28)	VTITCSASSR	Detected, but no potential probes	
L(29-38)	VGYMHWYQQK	Detected, but no potential probes	
L(45-52)	LLIY**D**TSK	D49 (0.16)	None
L(53-60)	LASGVPSR	Detected, but no potential probes	
L(61-102)	FSGSGSGT**E**FTLTI SSLQP**DD**FATYYCFQG SGYPFTFGGGTK	E69 (0.54), D80 (0.57)	D81 (0.02)
L(108-125)	TVAAPSVFIFPPS**D**EQLK	D121 (0.73)	E122 (0.35)
L(126-141)	SGTASVVCLLNNFYPR	Detected, but no potential probes	

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 2. (Continued). Results from Carboxyl Group Footprinting Experiments

Peptide Affected	*Peptide sequence*	*GEE-Labeled Residues (Fractional SASA)*	*Unlabeled Residues (Fractional SASA)*
L(149-168	VDNALQSGNSQ**E**SVT**E**Q**D**SK	E160 (0.39), E164 (0.43), D166 (0.34)	D150 (0.35)
L(169-182)	**D**STYSLSSTLTLSK	D169 (0.53)	
L(190-206)	VYACEVTHQGLSSPVTK		E194E194 (0.20)

Note: labeled residues marked in bold letters. * Site of glycosylation. Abbreviations: GEE = glycine ethyl ester; SASA = solvent-accessible surface area.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Comparison of HRF and Carboxyl Group Labeling Results

As HRF and carboxyl group labeling techniques offer two fundamentally very different approaches to probe the protein structure, it is interesting to compare the results gained from the two techniques. Table 1 and Table 2 represent 36 tryptic peptides where the modification sites could be confirmed.

These peptides were comprised of a total of 571 residues, representing 514 target probes (Ala, Cys, Asp, Glu, Phe, His, Ile, Lys, Leu, Met, Pro, Gln, Arg, Ser, Tyr, Val, Trp, and Tyr residues) for HRF, and 48 (Asp/Glu residues and C-terminus) target probes for the case of carboxyl group labeling. On an absolute scale, HRF offers about 11 times more target probes than carboxyl group labeling, thus offering potentially higher structural resolution. Table 1 and Table 3 indicate that of the 514 target probes for HRF, 150 were labeled and quantitated, or 29.2% of the targets. Table 2 and Table 3 show that of the 48 target probes for carboxyl group labeling, 28 residues were labeled, allowing for the potential quantitative mapping of 58.3% of the target probes. The lower percentage of labeling in the case of HRF can be explained by the hydrophobic nature of some of the probe residues, such as Cys, Phe, Ile, Leu, Met, Val, and Trp, which statistically speaking, are more often than not buried in the hydrophobic core of the protein and hence may not be accessible to hydroxyl radicals. However, these residues can sometimes be present on the surface of a protein, for example, when they function as part of a protein-protein binding interface or in the case of protein unfolding. Due to the wide variation in reactivity across different probes in the case of HRF, some accessible residues may not get labeled due to their low reactivity, and it can be difficult to confidently prove they are modified. Therefore, although HRF offers nearly 11 (514/48) times more probes than carboxyl group labeling, hence allowing for greater structural resolution and information, the number of probes observed to be labeled across the whole sequence was only about 5-fold higher (150/28).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 3. Comparison of Hydroxyl Radical-Based Footprinting (HRF and Glycine Ethyl Ester (GEE) Labeling Results

	HRF	*GEE Labeling*
Probes Available	514	48
Probes labeled and identified	150	28
Detection efficiency	29.2%	58.3%
Sequence coverage across detected peptides	90.0%	8.4%

The carboxyl group labeling method with GEE tagging offers a relatively easy and accessible alternative to radiolysis- or photolysis-based HRF as it is completed on a bench top. The limitations of the carboxyl group labeling approach include the limited number of probes and inability to probe biomolecular dynamics because the reactions tend to be slow and range in time scale in the order of minutes. The sub-millisecond time scales of labeling achievable by HRF experiments make it particularly suited to study the conformational dynamics of proteins by obtaining information that may not be accessible by carboxyl group labeling and HDX studies (*54*, *55*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Homology Model

A complete structural model of the IgG1 molecule is required for full corroboration with the footprinting studies as the CL studies provide information across the entire structure. Several crystal structures of intact IgG antibodies have been reported previously (1IGY (45) , 1HZH (43) ; 1MCO) (*56*). All these structures demonstrate unique conformations and provide snapshots of otherwise flexible IgG antibodies. The structure selected as a template to model the mAb in this study is from an anti-phenobarbital subclass IgG1 antibody (*45*). It displays a distorted Y-shaped molecule. The two Fab domains, both with elbow angles of 155°, are roughly related by a 2-fold rotation with an additional 9 Å translation. The two peptide chains in the Fc domain are also related by a 2-fold axis that is independent of the 2-fold axis for Fabs. As this crystal structure demonstrates a moderate, roughly symmetrical conformation, it has been selected to serve as a template to construct a structural model for the NISTmAb in this work.

Figures 3a and 3b respectively show the mapping of the oxidatively and GEE-labeled residues onto the generated IgG1 structural model, allowing exploration of the relationship between the experimentally observed data and the theoretical solvent accessibility calculations of side-chain surface area obtained from the model. Heavy chains are shown in green and blue, while the light chains are shown by yellow and cyan colors. Side chains of the modified residues are depicted in red, and they are approximately consistent with their apparent solvent accessibility predicted by the model. Note that there are two main “hotspots” with a high density of HRF-labeled residues in Figure 3a: the blue H chain area just beneath the hinge region and the hinge area. The first “hotspot” also

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

gets significantly labeled in case of GEE labeling in Figure 3b. However, the hinge region did not get labeled in case of GEE labeling in Figure 3b. Further examination revealed that the hinge region consists of H(220-240). There are only 2 GEE probes (D224, E236) present within that region. Peptide containing D224 could not be detected since it is only 4 residues long (SCDK). However, E236 did get labeled as shown in Table 2. These examples highlight the similarity and differences of information gained from the two approaches.

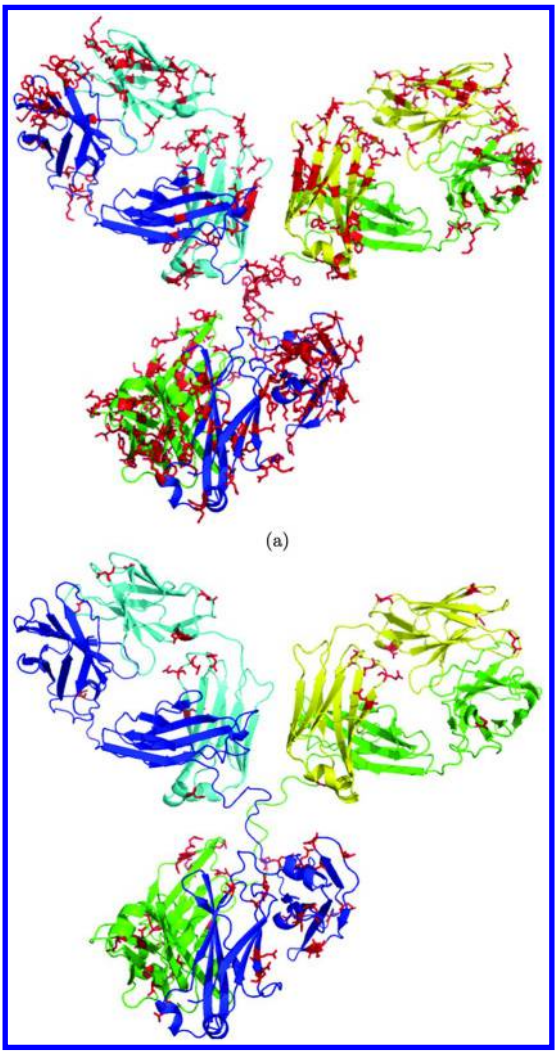

Figure 3. Homology model of the NIST monoclonal antibody (mAb). Heavy chains are shown in green and blue. Light chains are shown in yellow and cyan. (a) Side chains of the oxidized residues from hydroxyl radical-based footprinting (HRF) experiments are displayed in red. (b) Side chains of the glycine ethyl ester (GEE)-labeled residues are displayed in red. (see color insert)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Structural Prediction

Traditionally, footprinting experiments are used to draw structural comparisons of a particular region under different conditions to infer the underlying conformational change. However, such interpretation does not provide any information for cross-site (i.e., different regions of a protein) comparisons, which becomes complicated due to the widely varying reactivity of individual residues. Recently, we introduced the concept of the PF, where rate constant values of peptides were normalized based on known reactivity values of the constituent residues towards hydroxyl radicals (*57–59*). This facilitates measurement of the absolute solvent accessibility across different peptides in the case of HRF. The PFs are a measure of the degree of protection from covalent labeling that incorporate amino acid residue reactivity, and are used to map the solvent accessibility profile of the mAb peptides. The fourth column in Table 1 shows the calculated PF values for NISTmAb using the peptide rate constant and reactivity values of corresponding residues. In order to visualize the relationship of the PFs with the homology model, these PF values are represented onto the homology model in Figure 4 with the left panel showing the same molecular orientation as in Figure 3. The two views in Figure 4 show a 180° rotation of the mAb around the y-axis. The PF is color coded, with the red colored residues (PF $\geq$ 23) clearly oriented in the interior of the protein and the blue colored residues (PF $\leq$ 9) preferentially outside. The preferential orientation of these residues correlates well with the theoretical model, indicating PF values correspond to the expected solvent accessibility. The purple residues represent intermediate protection from the solvent, with $9 \leq PF \leq 23$, while the black color indicates unmapped regions due to undetected peptides. A comparison of the log of PF values against peptide fSASA values calculated from the homology model show a Pearson's correlation coefficient of −0.36 (p-value = 0.05).

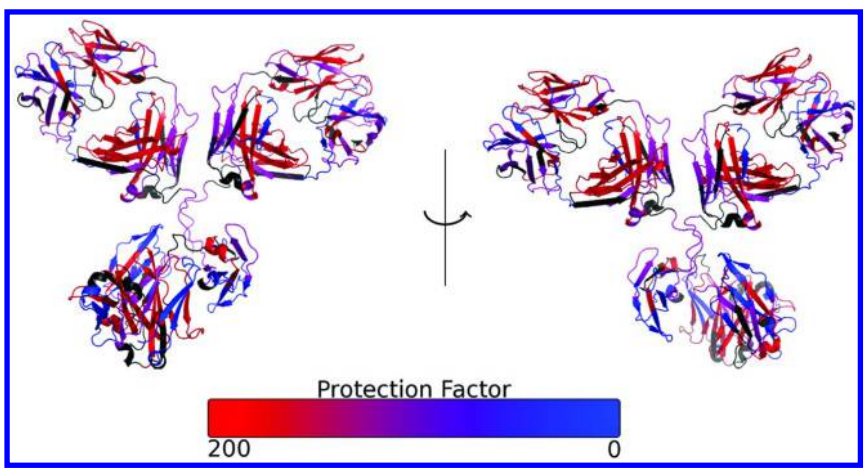

Figure 4. Protection factor (PF) mapping: blue: 0 < PF < 9, purple: 9 < PF < 23, red: PF > 23. Black indicates unmapped regions (no corresponding peptides detected). (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Although the correlations are not high, they are significant, and thus the structural model is consistent with the experimentally derived data, as seen in Figure 4. Caveats in the analysis at this level include uncertainties in the homology model described above and the peptide-level resolution of the data. Nevertheless, the PF analysis shows the potential power of the footprinting data to assess structure *de novo.*

Future Applications

The results presented in this chapter refer to the folded native state of the mAb. They can serve as a baseline for the native conformation form of the mAb, and can be utilized to perform comparative assessments against an altered conformation under similar experimental conditions. The altered conformation can arise due to processes such as aggregation, antigen binding, manufacturing process variation, and so forth. The rate constant values from the two forms can be compared. The peptides showing change in the values will highlight the regions undergoing changes in solvent accessibility as a function of structural reorganization, as illustrated in the following example studies.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Epitope/Paratope Mapping

Understanding antibody-antigen interactions and determining the epitope/paratope are important components of therapeutic antibody design and eventual commercialization. Structural MS has been increasingly popular to define both linear and conformational epitopes on a target antigen (*16*, *30*, *60*). This is done by comparing the solvent accessibility profiles of a free antigen as a reference experiment against the mAb-antigen complex, as shown in the workflow in Figure 1. The "protected" regions of the antigen/antibody experience significant reduction in labeling at the sites of epitope/paratope, hence helping locate the binding region.

In a recent study, HRF has been used to characterize the conformational epitopes of serine protease thrombin (*30*). Thrombin was subjected to HRF, both in its apo- and its antibody-bound form. Nine of the 34 identified peptides showed significant differences in the extent of modification between the two states. A decrease in the modification extent upon antibody binding was apparent in two regions (i.e., amino acid regions 114–119 and 130–171) of the protein, thus showing solvent protection in the bound state. The definition of the binding site within the latter region of 41 (130–171) residues was further refined to 18 (133–150) residues by examining tandem MS data in order to localize sites of modification. When both of these regions with decreased modification (hence, decreased solvent accessibility) were mapped onto the thrombin structure, they came together in the fold of the protein even though they were far apart in the primary structure. The proposed epitope agrees well with a previous study using HDX (*61*). In addition, two loop regions of thrombin experienced increased modification (hence increased solvent accessibility) in the antibody-bound state of the protein. This was attributed to allosteric changes likely serving to modulate

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the catalytic activity and substrate specificity by means of flexible loops. The increased modification extents for the antibody-bound state of thrombin in the two loops were not observed in the previous HDX study. This could be explained if the changes do not affect the hydrogen-bonding pattern of the protein, and only alter the side chain orientations. This highlights the complementary information gained from the two approaches. As shown in the study, footprinting data from the NISTmAb presented in this chapter can be used as a reference for the native form, whereas comparing it against the antigen-bound form would provide a map of the binding interface in the complex.

Characterization of Conformational Changes Due to Process Variations: Protein Quality Measures and Biosimilars

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Analysis of multiple variables related to identity, concentration, purity, and quality are defined characteristics of clinical-grade biological material that are regularly required to be assessed consistently with regulatory filings around the drugs. Quality measures are assessed by a variety of techniques in order to ensure the consistency of higher order structure, among other characteristics. This is necessary because biologics are subject to degradation, oxidation, and also higher order conformational changes that can occur without a change in protein primary structure. Improperly folded proteins can go through undesirable degradation pathways, and can be prone to aggregation, potentially triggering an unfavorable immune response (*62–64*). Such structural changes can be induced by a variety of processes such as chemical modification, denaturation caused by differences in manufacturing processes, or storage conditions (*65*). To ensure equivalency with the originally approved form and avoid extensive additional clinical trials, follow-on (or biosimilar) formulations of the therapeutic proteins need to establish the equivalency of both the primary and secondary structure (*5*, *66*). More importantly, neither X-ray crystallography nor nuclear magnetic resonance (NMR) spectroscopy are particularly suited to rapidly analyze protein aggregation or highly heterogeneous mixtures of protein conformations, which are common problems with therapeutic protein formulations.

MS-based methods for structural analysis provide a sensitive approach to characterizing both primary and higher order structure using a single analytical platform. The most widely used method is HDX analysis coupled with MS (*9*, *20*). In recent studies, CL has been applied to study conformational variations of Neupogen® (a patented GCSF), comparing it against several expired samples of recombinant GCSF, as well as heat-treated Neupogen® (*28*). Conformations of different samples of the therapeutic proteins interferon α-2A and erythropoietin also were compared (*28*). The study showed conformational differences between Neupogen® and expired recombinant GCSF samples. In addition, conformational equivalence between two different samples of GCSF produced 9 months apart was confirmed. Differences observed in the HRF between various Neupogen® samples were consistent with results from CD spectroscopy. Samples with identical HRF results were also found to be indistinguishable by CD spectra. In another study, HRF was applied to characterize the structural orientation of an IgG1 mAb dimer by using the mAb monomer as a control (*27*). The footprinting data showed that

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

peptides displaying decreased rates of oxidation in the dimer form were present in the light and heavy chains of the Fab domain. The interface region for the monomers comprising the dimer was thus inferred to be between the two Fab arms, allowing a modeling of two possible theoretical dimer orientations.

The footprinting results presented in this chapter could be used in a future study to characterize NISTmAb aggregates (e.g., as formed under stressed conditions) by using the present data as a control for monomeric form. In order to draw meaningful comparisons, the aggregate form will need to be subjected to footprinting analysis under similar experimental conditions as the present study. Any significant variation from the present results will serve to highlight the regions of conformational variation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Challenges

While there are many advantages to CL for structural characterization, there are also challenges, as discussed below.

Preserving Structural Integrity

One of the caveats is that covalent modification itself can alter the conformation of the protein if not carefully controlled. Fortunately, this has been addressed as a fundamental concept during the historical development of both nucleic acid and protein footprinting (*67*, *68*). This limitation is claimed to be offset in the case of HRF by completing the oxidation process faster than the timescale of a potential protein conformational change (e.g., microseconds for photolysis of peroxide), limiting the extent of overall modification. Other useful approaches are quantitatively measuring the oxidation rates to conclusively detect any untoward conformational changes (*55*). In addition, the linear (actually semi-log) reaction kinetics of the modification reaction ensures that structural integrity is preserved (*67*, *69*).

Sometimes, protein-ligand binding events induce changes in regions distant from the binding interface, called allosteric effects (*4*, *70*). Binding events also can provide kinetic and thermodynamic stability to the protein complex, possibly causing it to be more strongly protected from labels than the free form, if dynamic fluctuations are suppressed. In such cases, the interpretation of data becomes more complex. In order to confirm the cause of protection, a practical approach is to integrate the results with computational analysis and molecular dynamics. In addition, site-directed mutagenesis can be used for validating the results. In this process, the amino acid sequence encoded in the putative binding region is altered systematically. Mutations of the "true" binding regions are expected to diminish the ability of protein-ligand formation.

It is important to ensure the integrity and composition of the sample under investigation. Results may be ambiguous if size variants are present due to formation of aggregates, polypeptide chain cleavage, or disulfide bond alterations. Thus, the purity and homogeneity of the sample should be established by techniques such as size exclusion chromatography or native gels analysis prior

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

to the experiment. In particular, artifactual changes such as Trp/Met oxidation or Asn deamidation arising from sample preparation need to be minimized and controlled within specified ranges.

Quality Control and Software Solutions

Although the success of protein footprinting has been evidenced by a variety of studies, wider adoption by industry and regulatory authorities poses some barriers. Such barriers include the need for standardization in terms of internal controls, system suitability, and software solutions for data analysis. The high volumes of data generated for large molecules such as mAbs are not amenable to manual analysis. Although progress has been made for the automation of data processing using ProtMapMS (NeoProteomics, Inc., Cleveland, OH) and ByOnic (Protein Metrics, San Carlos, CA), elevating the technique and the software to the standards of good laboratory practices (GLPs) requires significant work (*40*, *71*). Systematic frameworks need to be established allowing for the studies to be designed with system suitability in mind, planned with internal controls, performed, analyzed, and reported. Executing projects within the GLP framework will ensure that the results can be efficiently obtained, and are a true reflection of the underlying variables being studied. This can extend the utility of the technique to larger scales in industry and could drive acceptance by regulatory authorities.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Future Directions

Higher Resolution and Use of Alternative Enzymes

Present approaches to CL merge the solvent accessibility measures at the level of an entire peptide, making it difficult to estimate the contributions from the constituent amino acid residues. In case of HRF, the dominating oxidation products from a few highly reactive residues interfere with the relatively weak signals originating from less reactive residues. These two factors currently limit the overall structural resolution of the technology. Protein digestion with alternative enzymes such as pepsin in combination with ultra high pressure chromatography can often isolate modifications of interest and thus reveal the relative contribution of individual residues. Specifically, pepsin enzymatic digestion generates a set of small overlapping peptide fragments ranging from 5 to 15 amino acid residues, making possible quantification of the extent of modification for each specific site. However, this approach relies largely on the chromatographic separation of each modified species and thus lacks efficiency, speed, and generality. In addition, the effect of oxidation on the efficiency of pepsin digestion at a given site needs to be carefully characterized in order to obtain reliable results. This necessitates the development of novel experimental and computational methods in order to fully realize the potential of spatial resolution that can be gained from such experiments. Such approaches have recently been attempted with some success and rely on tandem ion quantification to achieve residue-level resolution (*59*).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Intact and Top-Down MS

Mass shifts in the intact analysis of mAbs and their large domains can be particularly useful for detecting and monitoring irreversible modifications introduced in a footprinting experiment. This can be especially useful for estimating the total extent of modification on the protein. The fast turnaround of native MS is suited for it to be used to distinguish batch-to-batch heterogeneities by measuring the intact mAb masses and comparing them against the corresponding reference standard values.

The traditional bottom-up approach employed in footprinting provides reliable identification of the protein and its modifications but suffers the risk of peptide losses during digestion and analysis, thus limiting important information. An alternative strategy called "top-down" investigation of intact and reduced forms of mAb by LC-MS provides fast and accurate profiles of entire mAbs and the constituent domains. Top-down analysis involves characterization of the intact protein (as opposed to its proteolytically digested peptides), in which fragmentation methods such as collision-induced dissociation, electron transfer dissociation, and electron capture dissociation are used to fragment the intact protein within the mass spectrometer into smaller pieces for more detailed analysis.

Bottom-up and top-down strategies offer complementary information to analyze differences in protein structures. Bottom-up experiments offer high dynamic range, determining the fractional modification for individual amino acids ranging less than 0.2%. Top-down sequencing requires fewer sample preparation steps and with improvements in the technology may offer adequate sequence coverage even for large proteins. However, this approach lacks the sensitivity and dynamic range of the bottom-up approach.

Conclusions

The solvent accessibility measures of the reference IgG1 from NIST are examined using two irreversible covalent labeling approaches: hydroxyl radical-based footprinting, and carboxyl group labeling by GEE tagging. The overall experimental design is discussed. Trypsin-digested fragments of the mAb are identified with total sequence coverage of 85%. These peptides consist of a total of 571 residues, representing 514 target probes for HRF, and 48 target probes for the case of GEE labeling. These probes are sensitive to the solvent exposure of the corresponding residues, which is influenced by the conformation of a protein. The results highlight that the solvent accessibility values must be considered in combination with residue reactivities in understanding the observed oxidized peptides. PF measures are introduced in order to normalize the oxidation rate of peptides based on known reactivity values of the constituent residues. Homology-based models for the IgG1 are generated. The labeling results are mapped onto the model, highlighting the similarity and differences of the information gained using two approaches. The presented data can serve as a baseline measure for the unmodified (native) molecule, and can be utilized in future comparative studies with stressed material or for epitope mapping studies.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Acknowledgments

The authors would like to thank Dr. Benlian Wang for MS experiments, Dr. Jen Bohon for hydroxyl radical exposure through synchrotron, Irene Christian for helping with creating figures, and Sean Maxwell for supporting with ProtMapMS. This work was supported in part by the National Institute for Biomedical Imaging and Bioengineering under grants NIH R01-EB-09998 and P30-EB-09866 (to MRC).

References

1. Alzari, P. M.; Lascombe, M. B.; Poljak, R. J. *Annu. Rev. Immunol.* **1988**, *6*, 555–580.
2. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2013**, *85*, 715–736.
3. Schotte, F.; Lim, M.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Phillips, G. N., Jr.; Wulff, M.; Anfinrud, P. A. *Science* **2003**, *300*, 1944–1947.
4. Mayne, L.; Paterson, Y.; Cerasoli, D.; Englander, S. W. *Biochemistry* **1992**, *31*, 10678–10685.
5. Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. *Nat Rev Drug Discov* **2012**, *11*, 527–540.
6. Lundblad, R. *Approaches to the Conformational Analysis of Biopharmaceuticals*; Chapman Hall/CRC Press: Boca Raton, FL, 2009.
7. Jiskoot, W., Crommelin D. *Methods for Structural Analysis of Protein Pharmaceuticals*; AAPS Press: Arlington, VA, 2005.
8. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
9. Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R. *Drug Discovery Today* **2014**, *19*, 95–102.
10. Zhang, H.; Cui, W.; Gross, M. L. *FEBS Lett.* **2014**, *588*, 308–317.
11. Bohn, S.; Beck, F.; Sakata, E.; Walzthoeni, T.; Beck, M.; Aebersold, R.; Forster, F.; Baumeister, W.; Nickell, S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20992–20997.
12. Maiolica, A.; Cittaro, D.; Borsotti, D.; Sennels, L.; Ciferri, C.; Tarricone, C.; Musacchio, A.; Rappsilber, J. *Mol. Cell. Proteomics* **2007**, *6*, 2200–2211.
13. Lasker, K.; Forster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1380–1387.
14. Kalisman, N.; Adams, C. M.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2884–2889.
15. Houde, D.; Berkowitz, S. A.; Engen, J. R. *J. Pharm. Sci.* **2011**, *100*, 2071–2086.
16. Pandit, D.; Tuske, S. J.; Coales, S. J.; E, S. Y.; Liu, A.; Lee, J. E.; Morrow, J. A.; Nemeth, J. F.; Hamuro, Y. *J. Mol. Recognit.* **2012**, *25*, 114–124.
17. Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785–815.
18. Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514–3543.
19. Bobst, C. E.; Abzalimov, R. R.; Houde, D.; Kloczewiak, M.; Mhatre, R.; Berkowitz, S. A.; Kaltashov, I. A. *Anal. Chem.* **2008**, *80*, 7473–7481.

20. Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Wang, G.; Baykal, B.; Wang, S. *Biotechnol. Adv.* **2012**, *30*, 210–222.
21. Sinz, A. *Mass Spectrom. Rev.* **2006**, *25*, 663–682.
22. Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158–170.
23. Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.
24. Wei, H.; Tymiak, A. A.; Chen, G. *Bioanalysis* **2013**, *5*, 1299–1313.
25. Beck, A.; Sanglier-Cianferani, S.; Van Dorsselaer, A. *Anal. Chem.* **2012**, *84*, 4637–4646.
26. Kiselar, J. G.; Maleknia, S. D.; Sullivan, M.; Downard, K. M.; Chance, M. R. *Int. J. Radiat. Biol.* **2002**, *78*, 101–114.
27. Deperalta, G.; Alvarez, M.; Bechtel, C.; Dong, K.; McDonald, R.; Ling, V. *mAbs* **2013**, *5*, 86–101.
28. Watson, C.; Sharp, J. S. *AAPS J.* **2012**, *14*, 206–217.
29. Jones, L. M.; Zhang, H.; Cui, W.; Kumar, S.; Sperry, J. B.; Carroll, J. A.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 835–845.
30. Jones, L. M.; Sperry, J.; Carroll, J.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 7657–7661.
31. Hoare, D. G.; Olson, A.; Koshland, D. E., Jr. *J. Am. Chem. Soc.* **1968**, *90*, 1638–1643.
32. Hoare, D. G.; Koshland, D. E., Jr. *J. Biol. Chem.* **1967**, *242*, 2447–2453.
33. Zhang, H.; Shen, W.; Rempel, D.; Monsey, J.; Vidavsky, I.; Gross, M. L.; Bose, R. *Mol. Cell. Proteomics* **2011**, *10*, M110 005678.
34. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6134–6139.
35. Ferreira, J. P.; Sasisekharan, R.; Louie, O.; Langer, R. *Eur. J. Biochem.* **1994**, *223*, 611–616.
36. Chen, D. H.; Daron, H. H.; Aull, J. L. *J. Enzyme Inhib.* **1992**, *5*, 259–268.
37. Torossian, K.; Audette, M.; Poulin, R. *Biochem. J.* **1996**, *319* (Pt 1), 21–26.
38. Taha, T. S.; Ferguson-Miller, S. *Biochemistry* **1992**, *31*, 9090–9097.
39. Kaur, P.; Tomechko, S.; Kiselar, J; Shi, W.; Deperalta, G.; Wecksler, A.; Gokulrangan, G.; Ling, V.; Chance, M. R. *mAbs* **2014**, *6*, 1486–1499.
40. Kaur, P.; Kiselar, J. G.; Chance, M. R. *Anal. Chem.* **2009**, *81*, 8141–8149.
41. McLellan, J. S.; Correia, B. E.; Chen, M.; Yang, Y.; Graham, B. S.; Schief, W. R.; Kwong, P. D. *J. Mol. Biol.* **2011**, *409*, 853–866.
42. McLellan, J. S.; Chen, M.; Kim, A.; Yang, Y.; Graham, B. S.; Kwong, P. D. *Nat. Struct. Mol. Biol.* **2010**, *17*, 248–250.
43. Saphire, E. O.; Parren, P. W.; Pantophlet, R.; Zwick, M. B.; Morris, G. M.; Rudd, P. M.; Dwek, R. A.; Stanfield, R. L.; Burton, D. R.; Wilson, I. A. *Science* **2001**, *293*, 1155–1159.
44. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195–201.
45. Harris, L. J.; Skaletsky, E.; McPherson, A. *J. Mol. Biol.* **1998**, *275*, 861–872.
46. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501.
47. Lee, J. Y.; Kim, J. Y.; Park, G. W.; Cheon, M. H.; Kwon, K. H.; Ahn, Y. H.; Moon, M. H.; Lee, H. J.; Paik, Y. K.; Yoo, J. S. *Mol. Cell. Proteomics* **2011**, *10*, M111 009290.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

48. Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. *Nucleic Acids Res.* **2003**, *31*, 3316–3319.
49. Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1987**, *196*, 641–656.
50. Maleknia, S. D.; Kiselar, J. G.; Downard, K. M. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 53–61.
51. Li, X.; Li, Z.; Xie, B.; Sharp, J. S. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1767–1776.
52. Heidorn, D. B.; Trewhella, J. *Biochemistry* **1988**, *27*, 909–915.
53. Baldwin, E. T.; Weber, I. T.; St Charles, R.; Xuan, J. C.; Appella, E.; Yamada, M.; Matsushima, K.; Edwards, B. F.; Clore, G. M.; Gronenborn, A. M.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 502–506.
54. Hambly, D. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057–2063.
55. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 6563–6571.
56. Guddat, L. W.; Herron, J. N.; Edmundson, A. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4271–4275.
57. Huang, W. R.; Ravikumar, K. M.; Chance, M. R.; Yang, S. *Biophys. J.* **2015**, *108*, 107–115.
58. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75–86.
59. Kaur, P.; Kiselar, J.; Yang, S.; Chance, M. R. *Mol. Cell. Proteomics* **2015**, *14*, 1159–1168.
60. Padayatti, P. S.; Wang, L.; Gupta, S.; Orban, T.; Sun, W.; Salom, D.; Jordan, S. R.; Palczewski, K.; Chance, M. R. *Mol. Cell. Proteomics* **2013**, *12*, 1259–1271.
61. Baerga-Ortiz, A.; Hughes, C. A.; Mandell, J. G.; Komives, E. A. *Protein Sci.* **2002**, *11*, 1300–1308.
62. van Beers, M. M.; Bardor, M. *Biotechnol. J.* **2012**, *7*, 1473–1484.
63. Kessler, M.; Goldsmith, D.; Schellekens, H. *Nephrol. Dial. Transplant.* **2006**, *21* (Suppl 5), v9–12.
64. Maas, C.; Hermeling, S.; Bouma, B.; Jiskoot, W.; Gebbink, M. F. *J. Biol. Chem.* **2007**, *282*, 2229–2236.
65. Kueltzo, L. A.; Wang, W.; Randolph, T. W.; Carpenter, J. F. *J. Pharm. Sci.* **2008**, *97*, 1801–1812.
66. Kozlowski, S.; Woodcock, J.; Midthun, K.; Sherman, R. B. *N. Engl. J. Med.* **2011**, *365*, 385–388.
67. Takamoto, K.; Chance, M. R. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 251–276.
68. Brenowitz, M.; Chance, M. R.; Dhavan, G.; Takamoto, K. *Curr. Opin. Struct. Biol.* **2002**, *12*, 648–653.
69. Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895–2904.
70. Zhang, J.; Adrian, F. J.; Jahnke, W.; Cowan-Jacob, S. W.; Li, A. G.; Iacob, R. E.; Sim, T.; Powers, J.; Dierks, C.; Sun, F.; Guo, G. R.; Ding, Q.; Okram, B.; Choi, Y.; Wojciechowski, A.; Deng, X.; Liu, G.; Fendrich, G.; Strauss, A.; Vajpai, N.; Grzesiek, S.; Tuntland, T.; Liu, Y.; Bursulaya, B.; Azam, M.; Manley, P. W.; Engen, J. R.; Daley, G. Q.; Warmuth, M.; Gray, N. S. *Nature* **2010**, *463*, 501–506.
71. Bern, M.; Cai, Y.; Goldberg, D. *Anal. Chem.* **2007**, *79*, 1393–1400.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch003

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Chapter 4

Ion Mobility and Mass Spectrometry Measurements of the Humanized IgGk NIST Monoclonal Antibody

Iain D. G. Campuzano,*[,1] Carlos Larriba,[2] Dhanashri Bagal,[3] and Paul D. Schnier[3]

[1]Molecular Structure and Characterization, Amgen Inc., Thousand Oaks, California 91320, United States
[2]Mechanical Engineering Department, Yale University, New Haven, Connecticut 06520, United States
[3]Molecular Structure and Characterization, Amgen Inc., South San Francisco, California 94080, United States
***E-mail: iainc@amgen.com**

Monoclonal antibodies are an important class of therapeutic agent that are in widespread use for the effective treatment of many human diseases. We have used a number of mass spectrometers (Synapt G2 and OrbiTrap Exactive Plus extended mass range [EMR]) and ion mobility instruments (Synapt G1 modified to an radio frequency [RF]-confining drift cell and a Synapt G2 travelling wave system) to characterise the humanised IgG1k NIST monoclonal antibody molecule under native-MS and buffer conditions. On both the Synapt G2 and the OrbiTrap instrumentation, charge state distributions are very similar, and all major glycoforms were resolved on all observable charge states; however, only on the OrbiTrap instrument are the glycoforms fully resolved to baseline. Ion mobility measurements were made in both helium and nitrogen drift gases on an RF-confining drift cell device with derived collision cross-section values for charge states +21 to +26 range from 6696 Å^2 to 6892 Å^2 in helium and 7223 Å^2 to 7403 Å^2 in nitrogen, respectively. This small but gradual

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

increase in collision cross section with increasing charge can be attributed to a combination of increasing ion-induced dipole interaction between the charged protein and the neutral drift gas and the enlargement of the protein due to the repulsion of the surface charges. The higher charge states (+27 to +29) show a significant amount of gas-phase unfolding, evident by increased collision cross-section values and bimodal ion mobility arrival times. Collision cross sections measured on a travelling wave device (in nitrogen) were consistent with the mobility measurements made in nitrogen on the RF-confining drift cell device. Molecular dynamic simulations on the theoretically generated NIST monoclonal antibody (NISTmAb) coordinate structure indicate that the gas-phase structure undergoes a significant (up to 40%) amount of compaction. Theoretical collision cross-section calculations on the optimised molecular dynamic-derived NISTmAb structure are consistent with both helium and nitrogen instrument-derived collision cross-section values. Finally, collision-induced unfolding experiments also were performed on the +26 charge of the NISTmAb and could potentially be used as a means of providing structural information in addition to a single, and potentially limiting, gas phase-derived collision cross-section measurement.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Introduction

Ion mobility (IM) has the ability to rapidly separate, typically in the millisecond time frame, isomeric species based on differences in their collision cross sections (Ω; physical size and shape) in the gas phase as the ion moves through a neutral drift gas, typically helium (He) or nitrogen (N_2), under the influence of a weak electric field. When combined with molecular and quantum mechanical modelling, IM can provide specific information on ionic configuration and potential structural conformation of isobaric species. IM coupled to mass spectrometry (MS) can be traced back to the original experiments of E. W. McDaniel in the 1950s and 1960s, where he used a low-field drift tube instrument to study ion molecule reactions as a function of their mobilities (*1*). Some 60 years later, IM coupled to orthogonal acceleration time-of-flight mass (oaToF) MS is becoming a routine "workhorse" instrument of choice in many industrial and academic laboratories. For an excellent historical review of IM-MS development, we refer the reader to the review articles of Hill (*2*) and Eiceman (*3*). IM currently can be performed using a number of instrument platforms, ranging from drift tube instrumentation operated at reduced pressure (*4*, *5*) and atmospheric pressure (*6*), field asymmetric ion mobility analysers (*7*), differential mobility analysis (*8*), and the travelling wave ion mobility (TWIM) device (*9*). Currently, IM is used in many research areas, ranging from chemical warfare agent detection

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

(*10*); fundamental gas-phase peptide and protein structure determination (*11–14*); pharmaceutical compound analysis (both small and large molecules) (*15–18*); gas-phase radical cation structure determination (*19*, *20*); polymer analysis (*17*); phospholipid (*21*) and carbohydrate (*22*) structural analysis; and large native protein analysis (*23–25*), including mega-Dalton viral capsid analysis (*26*). Over the past 15 years, IM has experienced somewhat of a renaissance due to various groundbreaking gas-phase structural analyses performed on drift tube instruments. Examples include the work by groups such as Clemmer when analysing naked protein conformations of cytochrome-c (*27*); Bowers analysing gas-phase conformations of bradykinin (*28*); and Jarrold (*29*) analysing carbon and metal containing clusters, sodium chloride nanocrystals, and cytochrome-c as a function of activation energy and temperature. Most recently, the Robinson group have pioneered gas-phase structural biology analyses of both aqueous- soluble and membrane protein complexes (*30*, *31*) using the TWIM instrumentation. Following the early native protein TWIM research, predominantly performed by the Robinson group, many subsequent manuscripts have demonstrated the utility of IM (mainly in the form of TWIM) and its application into structural biology (*32–39*).

Monoclonal antibodies (mAbs) are a very important class of therapeutic agent that have found widespread use in the treatment of many human diseases. The use and application of IM in biopharmaceuticals is still lagging behind the rapidly developing area being coined "gas-phase structural biology" (*40–44*). However over recent years, a number of research articles have been published demonstrating native-MS and importantly IM and their application for native mAb analysis. In 2010, Bagal et al. (*45*) demonstrated how TWIM can potentially characterize disulfide variants in IgG2 mAbs. It was observed that the +26 charge state of an IgG1 molecule under native-MS and TWIM conditions showed a single but tailing TWIM arrival time distribution (ATD), in contrast to a clear bimodal distribution for an IgG2 molecule for the same charge state. In 2012, Beck et al. (*46*) using trastumab (Herceptin; Roche) and cetuximab (Erbitux; Lilly/Merck-Serono) as examples, used denaturing-MS, native-MS and IM analysis to compare these biological therapeutics to potential biosimilars. In 2012, Valliere-Douglass et al. (*47*) developed a native-MS method for the rapid analysis and determination of intact, noncovalently associated heavy and light chain antibody drug conjugate (ADC) molecules. The native MS conditions are critical, because traditional denaturing liquid chromatography (LC)-MS methods utilized to determine interchain cysteinyl-linked ADCs would result in denaturation of the ADC complex, therefore making identification impossible. Yamaguchi et al. (*48*) demonstrated that coupling IM-MS with ultrahigh performance liquid chromatography (UPLC)-hydrophilic interaction liquid chromatography (HILIC) chromatography has the potential to separate glycans based on their gas-phase mobility differences. They demonstrated that 10 pyridylamino glycans can be separated in both the LC and IM dimension and used molecular dynamics (MD) to probe the structures of two monogalactosylated glycan isomers. Rosati et al. (*49*) showed how the OrbiTrap-extended mass range mass spectrometer can be used to elegantly and unambiguously identify and quantify 8 out of 10 distinct deglycosylated mAbs. In 2013, Jones et al. (*50*) use a combination of rapid and

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

sensitive analytical methodologies, such as TWIM, top-down MS and protein footprinting in the form of fast photochemical oxidation to monitor changes in higher order structure. They demonstrated the detection of subtle changes in conformation of cysteine to serine mutants of an IgG2 molecule, each representing a single disulfide isoform. Finally, and very importantly, in 2014, Thompson et al. (*51*) summarize all aspects of mAb sample preparation, native nanoelectrospray ionisation (nESI), oaToF-MS, OrbiTrap-MS, and data interpretation and analysis. The above example are not intended to be an exhaustive literature review but instead give the reader an idea of the recent and pertinent examples of MS and IM applications within current mAb research and development.

The aim of this chapter is to discuss current IM technology (radio frequency [RF]-confining drift cell and TWIM) and native-MS for the analysis of mAbs in both industry and academia. We also discuss the theory and application of each technique and we demonstrate, with the use of the humanised IgG1k NISTmAb, how IM analysis and separation can be achieved on both an RF-confining drift cell device and TWIM instrumentation, and collision cross-section (Ω) values, both Ω_{He} and Ω_{N2} values, can be easily generated. These Ω values then can be used as protein standard reference values for comparison to other mAb samples or theoretically generated Ω values, or become additions to the already growing denatured and native protein Ω value database (*52*). We also report a detailed MD study of the NISTmAb, using the NISTmAb coordinate file (Higher Order Structure chapter/Volume 3, Chapter 2) and report how this structure appears to undergo a significant amount of gas-phase collapse. Finally, we also will discuss the potential limitations of generating a single Ω value for a molecule such as a mAb, and how the emerging area of collision-induced unfolding (CIU) of proteins may potentially be a valuable technique of "fingerprinting" native protein complexes. We have chosen to only report native gas-phase MS and IM data, rather than analysis under denaturing solvent conditions, of the full intact NISTmAb within this manuscript because ionisation under native-MS solvent conditions will yield gas-phase Ω values that are more representative of a native-like structure.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Ion Mobility

IM spectrometry is a gas-phase chromatographic technique that separates ionic species based on their mobility differences as a result of their short- and long-range interactions with the neutral buffer gas, typically He (or N_2), under the influence of a weak electric field. In principle, packets of ions are driven through a background gas using an electric field gradient across a drift cell device, and a temporal separation of differing mobility species (ions) ensues. If we consider a protein with two distinct conformations, one extended and one compact, both with equal amounts of distributed surface charge, the more extended conformation would traverse the drift cell more slowly than the compact conformation. Instrument-derived Ω values directly relate to the structure of the ion species in the gas phase and can be compared with calculated values from MD and or quantum mechanical calculations (depending on the complexity

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

of the ionic species being studied), therefore allowing the researcher to infer structural information about the ion species under investigation. Additionally, IM separation also provides additional peak capacity to an MS-system, which can be particularly beneficial for complex samples, such as classic LC-MS/MS proteomics experiments (*53*–*55*).

The primary components of both IM devices described herein are:

- An **ion source**, nano-electrospray ionisation (nESI) is commonly used because many of the biological IM experiments are concerned with retaining native peptide/protein structure, and the sample is introduced into the IM device from aqueous buffered solvents. Matrix-assisted laser desorption/ionisation (MALDI) ion sources have been coupled to IM devices (*21*, *56*) (the RF-confining drift cell instrument is fitted with a dual electrospray ionisation (ESI)/MALDI source) but are less commonplace for native biological IM analyses.
- An **ion selection device**, typically a quadrupole ion guide (*57*–*59*) that allows for ion selection for possible collision-induced dissociation (*60*) and ion heating experiments (*61*).
- A **collision cell**, typically filled with a neutral gas such as argon or the larger, more massive xenon (*60*) or sulphur hexafluoride (*58*), which are more commonly used for more effective collisional cooling (*62*) of large multimeric protein complexes.
- The **IM device**, whether that be the low-field drift tube (*4*); an RF-confining, stacked-ring ion guide (*52*); or a TWIM device (*9*).
- The **analyser**, which can be in the form of a quadrupole (*63*), an ion trap (*64*), or an oaToF analyser (*57*) as described herein.

With respect to this chapter and also structural biology, coupling oaToF and IM is advantageous since IM separation typically occurs on the millisecond time scale and oaToF analysis is on the microsecond time scale. For example, if one considers the RF-confining drift cell instrument described herein, for every IM separation (on the order of 10 to 20 msec) there are 200 oaToF pusher pulses sampling an individual IM separation. OaToF is very well-suited to biological analysis because it has a very high mass range (up to *m/z* 100,000 on the instruments described herein), and large biological complexes can routinely display *m/z* values in excess of 8,000 Th (*58*, *65*), and in extreme cases, viral particles ranging from 3 to 18 MDa (*66*, *67*) displaying *m/z* values in excess of 20,000 Th. However, although not yet coupled with IM, the OrbiTrap detector (*68*, *69*) also is rapidly becoming the instrument of choice for native multimeric protein complex analysis due to its apparent efficient desolvation and ion transporting properties. For this reason, we have also included data acquisitions from the OrbiTrap Exactive Plus extended mass range (EMR) instrument as a direct comparison to traditional oaToF instrumentation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Drift Tube Instrumentation

Theory is well-developed for drift tube instrumentation (including the RF-confining drift cell; *vide infra*) that relates the measured mobility (K) of an ionic species to the rotationally averaged gas-phase collision cross section (Ω) for the interaction with a neutral background gas, typically He (or N_2) under low-field conditions (*70*). The velocity (v) of the ion under investigation is proportional to the electric field gradient (E) and the mobility (K) of the ion:

$$v = KE \quad (1)$$

The mobility (K) of the ion is directly proportional to the drift tube gas pressure and typically is converted to a reduced mobility (K_o) to enable direct comparison with other K values determined over a range of different experimental conditions. The mobility (K) is therefore corrected to standard temperature (T_o = 273.15 K) and pressure (p_o = 760 Torr):

$$K_0 = \frac{pT_o}{p_o T} K \quad (2)$$

where the neutral gas temperature (T) is ideally the measured temperature of the drift gas within the drift cell. For drift tube systems, in which uniform and time-invariant electric fields are used, K can be determined using equation 3 under low-field conditions. The Ω value of an ion is related to its mobility (K) through the Mason-Schamp equation (*70*):

$$\Omega = \frac{3ze}{16N}\left[\frac{2\pi}{\mu kT}\right]^{\frac{1}{2}}\frac{1}{K} \quad (3)$$

where e is the unit electronic charge, z is the number of charges on the ion, N is the neutral gas number density, k is Boltzmann's constant and μ is the reduced mass of the ion and the neutral, given by $[m_i m_n/(m_i+m_n)]$, where m_i and m_n are the ion and neutral masses, respectively.

The determination of K and therefore Ω values from drift tube (*4*) and RF-confining drift cell instruments (*18*, *52*) have been described previously; however, it is worth reiterating this information again, because an RF-confining device has been utilized heavily within this manuscript. Typically, on a drift tube-based instrument, multiple drift time measurements are made at a single pressure (usually 2–3 Torr; He or N_2) using different electric field gradients (3–11 V/cm) that are well below the low-field limit (*4*). For the RF-confining drift cell instrument described herein, the drift gas pressure was accurately measured using an MKS capacitance manometer (Type 626; range 10 Torr; accurate to 0.25%). This device directly measured the neutral gas pressure in the RF-confining drift cell device. Ion ATDs are measured for the ion of interest at each drift cell voltage gradient and plotted as a function of ATD versus the reciprocal of the drift

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

tube voltage gradient (1/V). Typically, the relationship between ATD and 1/V is linear, with R^2 correlation coefficients of greater than $R^2 = 0.9992$. The mobility (K) of the ion can be calculated from the slope of the linear drift time versus 1/V relationship; and therefore, the Ω value of the ion also can be calculated. Additional useful information can be obtained from this linear relationship of ATD versus 1/V; the intercept on the Y-axis representing t_0, which corresponds to the time the ions have spent outside the drift cell. By subtracting t_0 from the ATD, the actual ion drift time (t_d) within the cell can be calculated. The linearity of the relationship also demonstrates that there are minimal higher order effects resulting from ion activation that can affect the measured drift times. Our drift time versus 1/V relationship plots are reported in the Appendix (Figure A1, A3 and A4).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Travelling Wave Instrumentation

In 2004, the travelling wave was introduced as a means of propelling an ion through a stacked ring ion guide (*9*). TWIM offers analytical advantages, including improvements in cross-talk during precursor ion scans and multiple reaction monitoring acquisitions and oaToF duty cycle enhancements; and importantly, in the context of this chapter, IM separation was achieved with travelling wave technology on compounds such as leucine-enkephalin, gramicidin-S, and bradykinin. In 2006, the first commercial TWIM device was released (*57*), and in 2009, the second-generation TWIM device was introduced (*71*, *72*). The first generation TWIM device was based around the following electrode geometry: 122 gold-coated electrodes (plates), each 0.5 mm in thickness, with a centre-to-centre spacing of 2.5 mm, each with a single orifice of 0.5 mm. The device was 18.5 cm in length. An equal but opposite RF voltage was applied to adjacent electrodes (2.7 Mhz, 300V peak-peak), which acts to radially confine the transported ions. Superimposed on top of this RF is a direct current (DC) travelling wave. This DC travelling wave is sequentially applied to pairs of plates and travels axially along the device, propelling packets of ions along the TWIM device toward the analyser. In the presence of a neutral gas, typically N_2 at a pressure of 0.4 Torr (0.5 mbar), the propelled ions experience "roll-over" events. The number of "roll-over" events experienced by the ions depends on their mobility; lower mobility ions take longer to traverse the TWIM device and therefore experience more roll-over events. The second-generation TWIM device (Figure 1) is made up of more electrodes (168), therefore it is longer (25.4 cm); it operates at higher pressures 2.2 to 3.0 Torr (3 to 4 mbar N_2); and the travelling wave is applied to four adjacent electrodes, as opposed to two adjacent electrodes, as in the first-generation device (*9*). Along with other modifications, the second-generation device affords an approximate three- to fourfold increase in TWIM resolution, based on $\Omega/\Delta\Omega$ (*71*). It is this second-generation device (Synapt G2 High Defnition Mass Spectrometer [HDMS]) that is used to generate the TWIM data discussed within this manuscript (Table 1).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

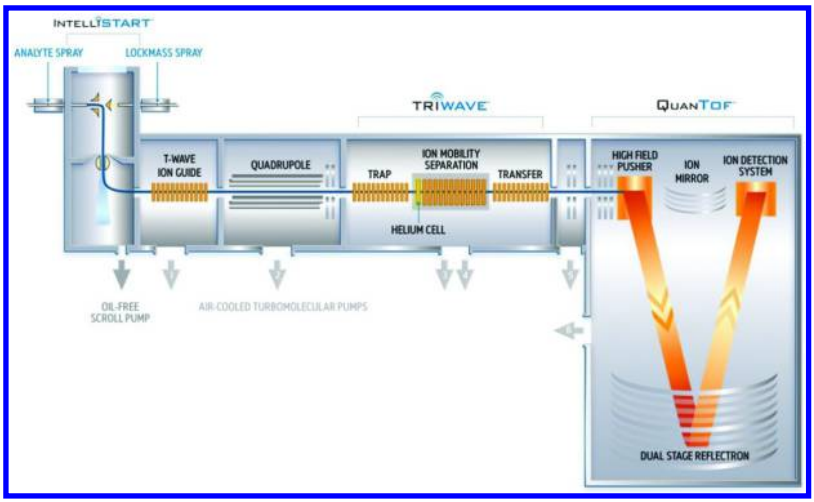

Figure 1. A schematic of the second-generation travelling wave ion mobility mass spectrometer, the Synapt G2 HDMS instrument. (courtesy of Waters Corporation, MS-Technologies Centre, Manchester, UK).

Currently, Ω values cannot be determined by TWIM without prior calibration of the drift time scale (*52*, *65*) due to the complex nature of the ion trajectory through the TWIM device (*73*). Therefore, a complete analytical relationship between the ion's mobility and the experimental separation parameters has not yet been developed, although initial progress has been made to overcome this limitation (*73*, *74*). In the context of native protein analysis, practitioners would typically use Ω_{He} values determined for denatured proteins such as ubiquitin, cytochrome-C, and myoglobin (*75*) that were acquired using a home-built, research-grade drift tube instrument. These values were then used to calibrate the native protein ATD data acquired in the drift gas N_2 on the TWIM device. Experimentally this is not ideal; however, the TWIM-derived Ω values correlated well with theoretical Ω values for many Protein Data Bank (PDB) structures, both X-ray crystallographic- and nuclear magnetic resonance-derived (*30*, *76*). Significant progress was made in 2010 (*52*) and 2012 (*18*, *77*) when protein, small molecule, and peptide Ω_{N2} values, respectively, were published, which allowed for the first time accurate TWIM-derived Ω_{N2} values to be generated using a Ω_{N2} calibration data set.

The TWIM calibration procedure has been well documented (*18*, *52*, *65*, *77*, *78*); however, we will briefly cover the fundamentals. The literature drift tube Ω_{N2} values (in this case native protein complexes) were adjusted by multiplying by the square root of the reduced mass and dividing by the charge state (z) of the calibrant species to provide a term Ω' that is proportional to the reciprocal mobility (K) of the ion:

$$\Omega' = \frac{\Omega \sqrt{\mu}}{z} \quad (4)$$

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

The measured TWIM calibrant ATDs were corrected for their mass-dependent flight times between the TWIM device and the oaToF analyser entrance:

$$t' = ATD - c\sqrt{\frac{m}{z}} \quad (5)$$

where c is an instrument constant that can be obtained from the MassLynx software. The corrected Ω_{N2} values (Ω') are plotted against corrected ATD values (t') determined using the TWIM device and the data fitted using an empirically determined power form $y = ax^n$. The derived coefficient, n, is obtained from the calibration curve and used to calculate the Ω_{N2} of the ion of interest from its measured TWIM ATD, obtained under identical operating conditions to those of the calibrant ions. Typically, three TWIM measurements are made, either varying the amplitude or the speed of the travelling wave, so an average TWIM Ω_{N2} value can be derived, including standard deviation values. Our TWIM calibration plots are reported in the Appendix (Figure A2) for the TWIM amplitudes 19, 20, and 21 V (all acquired at a speed of 250 m/sec).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

RF-Confining Drift Cell Instrumentation

The quadrupole-time-of-flight (QToF) (*79*) and travelling wave technologies (*9*, *71*) have proven themselves to be workhorse platforms for high-mass native protein complex analysis due to their easily achievable data acquisitions over broad *m/z* ranges. For example, operating the oaToF instruments described herein with a pusher frequency of 450 μsec allows for routine data acquisitions up to *m/z* 100,000. Because most protein complexes possess a low number of charges, the observable charge states often appear very high in the *m/z* scale, often above 8000 Th, well out the range of standard quadrupole and ion trap instrumentation, making the oaToF the *de facto* instrument for large native protein complex analysis.

Despite the mass range of the oaToF analyser, an area clearly lacking at the time in TWIM development was its ability to generate Ω values of ionic species which appear high in *m/z* scale (typically, native protein species) due to lack of sufficient native and high-mass protein IM calibrants. In 2010, the Robinson Research Group (Oxford, UK), with the help of Waters Corporation (MS-Technologies Centre, Manchester, UK) converted a production-grade travelling wave Synapt G1 HDMS instrument into an RF-confining drift cell (*52*) system (Figure 2; no applied travelling wave), therefore allowing Ω values to be derived for any ionic species in any gas (in theory) without the need for prior IM device calibration. In brief, the TWIM device was replaced with an RF-confining (2.1 MHz and typically 150 V peak-to-peak) drift cell device of 18.5 cm in length, containing 122 plates. A drift gradient typically ranging from 60 to 200 V can be applied across the device, resulting in efficient IM separation of ionic species ranging from small molecules (*18*), peptides (*80*), and large protein complexes such as GroEL (*52*), and importantly, generating a large Ω_{He} and Ω_{N2} database used for TWIM calibration.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

This RF-confining device is operated at typical Synapt HDMS operating pressures of 1 to 2 Torr (2–3 mbar) with a variety of neutral drift gases (He and N_2). The main benefit of this instrument is that Ω values for any ionic species (in the case of this study, native protein complexes) can be derived (*52*) not only in He, but importantly in N_2 (and any other gas for that matter), the *de facto* IM gas which the TWIM device is routinely operated in (*9*). Additionally, the highly debatable issue of ion heating as a function of the travelling wave speed and amplitude within the TWIM device (*73*, *81*) also can be avoided in the RF-confining drift cell device. The conversion of the standard Synapt G1 HDMS instrument to an RF-confining drift cell instrument described here was validated using previously measured molecules of known Ω value. The protein complex alcohol dehydrogenase (ADH, molecular weight [Mw] 147.5 kDa; Sigma-Aldrich, St. Louis, MO, USA; A3265) and the small drug-like molecules alprenolol (Mw 251; Sigma-Aldrich, St. Louis, MO, USA; A8676), and reserpine (Mw 609; Sigma-Aldrich, St. Louis, MO, USA; R0875) were used. All measured Ω values (both He and N_2) were within 1 percent of previously published data (*18*, *52*) (Appendix, Figures A3, A4 and Table A1).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

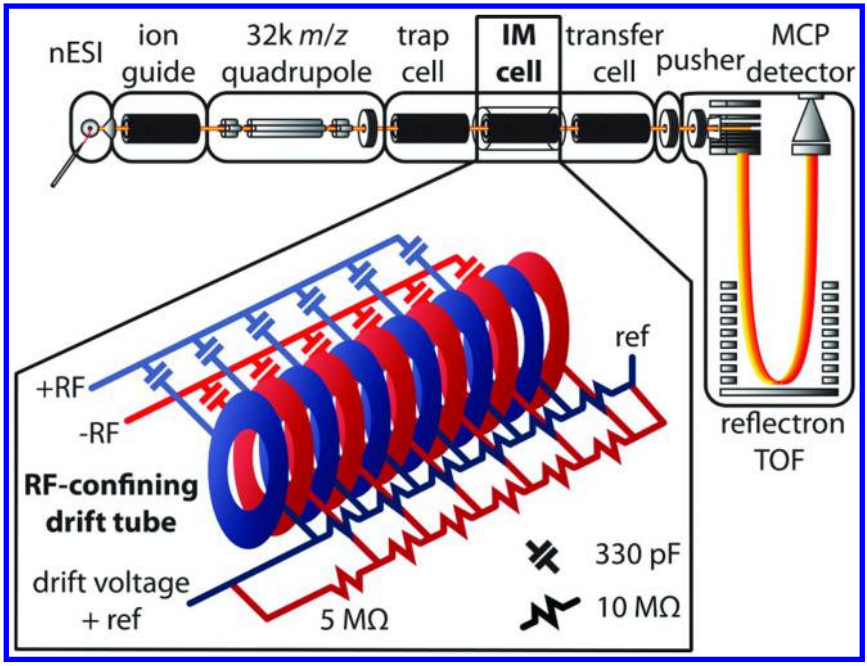

Figure 2. A schematic of the Waters Synapt G1 HDMS instrument modified with an RF-confining drift cell. IM = ion mobility, MCP = micro-channel plate, nESI = nanoelectrospray ionisation, RF = radio frequency, TOF = time of flight. Reproduced with permission from reference (52). Copyright 2010 American Chemical Society.

Sample Preparation

The NISTmAb humanised IgG1k sample (Candidate RM 8670, lot 31fb) was received at a concentration of 10 μg/μL (~67 μM). For MS analyses it was buffer exchanged two times using a BioRad P6 spin desalting column (6 kDa molecular weight cutoff [MWCO]; BioRad, Hercule, CA, USA) into 50 mM ammonium acetate pH 7.0, which is generally considered to be the *de facto* buffer for native protein MS analysis. The buffer exchanged NISTmAb was then further diluted in 50 mM ammonium acetate to an MS working solution of 5 μM and introduced into the Synapt G2 HDMS instrument and the RF-confining drift cell instrument using a nESI source and gold-coated nanoflow needles. External instrument calibration was performed using a 50 μg/μL 50% (v/v) aqueous acetonitrile solution of caesium iodide over the *m/z* range 500 to 32,000. The TWIM ATD data (Synapt G2 HDMS instrument only) was calibrated using the protein standards (*52*) bovine serum albumin (BSA, 69 kDa, monomer; Calbiochem, La Jolla, CA, USA; 12657), serum amyloid protein (SAP, 125 kDa; Calbiochem, La Jolla, CA, USA; 565190), ADH (147.5 kDa; Sigma-Aldrich, St. Louis, MO, USA; A3263), pyruvate kinase (PK, 230 kDa; Sigma-Aldrich, St. Louis, MO, USA; P9136), and glutamate dehydrogenase (GDH, 330 kDa; Sigma-Aldrich, St. Louis, MO, USA; G7882). The protein standards were prepared identically to the NISTmAb, as described above, using 100 mM ammonium acetate. For representative native MS spectra of the TWIM calibrants, please refer to Figure A5 (Appendix). For more details describing native protein preparation for native MS analyses, please refer to the work of Ruotolo (*65*) and Campuzano (*58*).

OrbiTrap Exactive Plus EMR Sample Preparation

The NISTmAb was buffer exchanged into 50mM ammonium acetate using a 10 kDa MWCO spin filter column (Vivaspin 500, Sartorius Stedim Biotech GmbH, Goettingen, Germany). The sample was further diluted to a concentration of 5 μM for analysis. External instrument calibration was performed using a 25 μg/μL 50% (v/v) aqueous isopropanol solution of caesium iodide over the *m/z* range 1000 to 20,000.

Synapt G2 HDMS Instrument Voltages and Pressures

The Synapt G2 HDMS instrument was operated in positive nanoflow ESI mode. All critical instrument voltages and pressures are as follows: capillary voltage 0.8 to 1.0 kV; sample cone 25 V; extraction cone 1 V; source block temperature 30 °C; trap collision energy 4.0 V to 100 V, depending on the experiment; transfer collision energy 3 V; trap entrance 3.0 V; trap bias 45 V; trap DC −2.0 V; trap exit 0.0 V; ion mobility spectrometer (IMS) entrance 10 V; IMS helium cell DC 25 V; IMS helium exit −5.0 V; IMS Bias 3.0 V; IMS exit 0.0 V; transfer entrance 4.0 V; transfer exit 5.0 V; IMS wave velocity 250 m/sec; IMS wave amplitude 19.0 V to 21.0 V; transfer velocity 47 m/sec; transfer wave amplitude 4.0 V; mobility trapping release time 200 μsec; trap height 20.0 V; extract height 0.0 V; source RF-amplitude (peak-to-peak) 450 V; triwave

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

RF-amplitudes (peak-to-peak) trap 380 V, IMS 250 V, transfer 380 V; source backing pressure 6.0 mbar; trap/transfer pressure SF_6, $3.25e^{-2}$ mbar (Pirani gauge indicated; flow rate 4.5 mL/min); IMS pressure N_2 3.54 mbar (Pirani gauge indicated; flow rate 90 mL/min); He cell flow rate was set to 180 mL/min. Instrument control and data acquisition was carried out through MassLynx 4.1 SCN 781.

RF-Confining Drift Cell Instrument Voltages and Pressures

The modified Synapt G1 HDMS instrument was operated in positive nanoflow ESI mode. All critical instrument voltages and pressures were as follows: capillary voltage 0.8 to 1.0 kV; sample cone 40 V, extraction cone 1 V; source block temperature 30 °C; trap collision energy 4.0 V; transfer collision energy 60 to 200 V, corresponding to 3.3 V to 11.1 V/cm applied across the RF-confining drift cell; trap entrance 3.0 V; trap bias 10 V (He) and 16 V (N_2; an increased potential is required to inject the ions in the drift cell when operated with N_2); IMS DC entrance 5.0 V; IMS DC exit 0.0 V; transfer DC entrance 0.0 V; transfer DC exit 2.0 V; transfer wave velocity 70 m/sec; transfer wave amplitude 4.0 V; mobility trapping release time 250 μsec; trap height 20.0 V; extract height 0.0 V; source RF amplitude (peak-to-peak) 450 V; triwave RF amplitudes (peak-to-peak), trap 380 V, IMS 150 V, transfer 380 V; source backing pressure 6.0 mbar; trap/transfer pressure SF_6, $3.3e^{-2}$ mbar (Pirani gauge indicated; flow rate 3.0 mL/min); IMS pressure N_2 2.05 mbar (1.54 Torr; flow rate 38 mL/min); IMS pressure He 2.70 mbar (2.03 Torr; flow rate 70 mL/min). The pressure within the RF-confining drift cell was accurately measured using an MKS Baratron capacitance manometer, type 626 (range 10 Torr; accurate to 0.25%) and an MKS PDR2000 power supply. Ambient temperature was measured using an Oakton Temp10T thermocouple. The temperature was measured at the point where the capacitance manometer is connected to the instrument ion optics lid. Potentially, accurate drift gas temperature can be measured by incorporating a thermocouple inside the RF-confining drift cell optics housing. Instrument control and data acquisition was carried out through MassLynx 4.1 SCN 639, SCN744. As mentioned earlier, the mobilities and, therefore, the Ω values were made by accurately measuring the pressure within the drift cell device and the ambient temperature of drift cell and making up to 10 mobility measurements at different drift cell voltages (60 V to 200 V). This procedure was repeated twice and an overall average Ω value between the two experiments was derived. Each set of 10 mobility measurements took approximately 10 minutes (1 minute per acquisition); therefore, very little temperature or pressure variation was observed. However, individual temperature and pressure measurements were taken for each acquisition, and for the final Ω value calculation, an average temperature and pressure value was used. Typical temperature and pressure variations over 10 minutes were ≤ 0.2 °C and ≤ 0.002 Torr.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

OrbiTrap Exactive Plus-EMR Voltages and Pressures

Experiments were performed on an Exactive Plus EMR instrument (Figure 3; ThermoFisher Scientific, Bremen, Germany) (*49*, *68*) equipped with a nanoflow ESI source. All critical instrument voltages and pressures are as follows: Capillary voltage 0.8 to 1.0 kV. Ions formed by nESI were passed through a stainless steel capillary (ion transfer tube) maintained at 250 °C into an S-Lens stacked ring ion guide with an applied S-Lens RF-Level of 200. Ions then travelled through a transport multipole and entered the higher energy collisional dissociation (HCD) cell, where they were stored at a high pressure before they were returned to the C-trap. This feature allows efficient trapping and desolvation of large protein ions and dramatically improves sensitivity. Nitrogen gas was used in the C-Trap as well as the HCD cell. Utilising a trapping gas pressure setting of 7.0 (software-determined) the C-Trap pressure is approximately $2.0e^{-4}$ mbar and the ultra high vacuum pressure (OrbiTrap analyser) is $7.5e^{-10}$ mbar. The voltage offsets on the transport multipoles were manually tuned to increase the transmission of large complexes (C-Trap entrance lens, 0 V; bent flatapole DC 4 V; inter-flatapole DC 4 V; injection flatapole DC 4 V. An in-source collision-induced dissociation (CID) voltage of 125 V and an HCD voltage of 25 V were required to achieve efficient NIST sample desolvation. Transients detected in the Orbitrap were processed using enhanced Fourier transformation (eFT™) for converting the transients into frequency before *m/z* conversion (*82*, *83*). The instrument was set at a nominal resolving power of 70,000 at *m/z* 200, and mass spectra were acquired for 2 minutes by averaging 10 microscans per analytical scan. Data was analysed using Xcalibur™ 2.2. No additional data processing (smoothing) was performed.

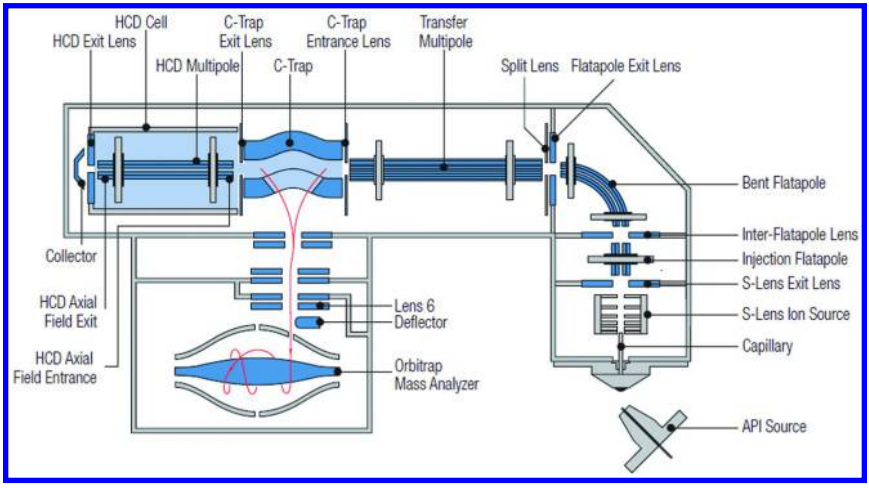

Figure 3. A schematic of the OrbiTrap Exactive Plus extended mass range (EMR) mass spectrometer. (courtesy of Thermo Fisher Scientific, Bremen, Germany). HCD = Higher-energy collisional dissociation.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

MD Simulations

In order to provide further insight into the experimental RF-confining drift cell and TWIM results, the NISTmAb model (Higher Order Structure chapter/Volume 3, Chapter 2) was used to theoretically characterize the gas-phase structure. IMoS mobility software (*84*, *85*) was utilized to calculate Ω_{He} and Ω_{N2} values, using four different theoretical methods: the projected area approximation (PA); the elastic/exact hard sphere scattering (EHSS) and diffuse hard sphere scattering (DHSS) methods, both with ion-induced dipole potentials (4–∞); and the trajectory method (TM) using Lennard-Jones potentials, including polarization (4-6-12 potential). Prior to MM2 MD, an energy minimisation step was performed on the initial NISTmAb coordinate set using Chem 3D Pro. MM2 MD then was performed on the energy-minimised NISTmAb coordinate set as the initial starting point. MM2 force fields were used in Chem 3D Pro to run energy refinement optimisation simulations, using cutoff distances of 80 Å, step values of 2.0 psec, a heating/cooling rate of 1.0 Kcal/atom/psec, and a target temperature of 300 K. For faster conversion, the structure was assumed to be globally neutral. This potentially could result in theoretically inferred gas-phase structures and Ω values being different from instrument-derived Ω values due to charge repulsion volume enlargement for heavily charged antibodies, although the average charge observed on the NISTmAb under native MS conditions is less than the predicted level of charge (*vide infra*). For the theoretically derived Ω_{He} and Ω_{N2} values, positive charges (+23 and +27, which corresponds to the charge range observed under native MS conditions) were added to the energy-refined and optimised NISTmAb structure through the IMoS mobility software (84-85), and were assumed to be geometrically centred within the intact mAb structure. No additional mass was added to the NISTmAb structure upon assignment of positive charge because the mass of the charges is negligible compared to the molecular weight of the mAb (148.7 kDa).

Results

Native MS Analysis

Over the past decade, the nESI-oaToF instrumentation have been the *de facto* instruments of choice for the analysis of native protein complexes due to high *m/z* range of the oaToF analyser and easy optimization of the preceding ion optics and pressure stages. However, over the past few years, the OrbiTrap (*68*, *69*) and Fourier transform ion cyclotron resonance (FT-ICR) (*86*) instruments have proven themselves to be more than capable of preserving, transmitting, and detecting large native protein complexes. On both the Synapt G2 and OrbiTrap EMR instrumentation, upon infusing the NISTmAb under native MS-buffer conditions (50 mM ammonium acetate), a narrow charge state distribution (+22 to +28) can be observed (Figure 4). Interestingly, the charge state distribution is very similar on both instruments Additionally, the predicted average charge (*z*) for a protein of molecular weight 148.7 kDa is +30 (*87*); the NISTmAb experimental results yield an average *z* value of +25. The predicted level of *z* for a protein

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

is based on the power relationship z = 0.0778 $Mw^{0.5}$, assuming the protein is spherical with a density of 1 g/cm^3, and a nESI-produced droplet possessing a surface tension value of 0.078 N/m (*87*). We know from the limited number of full crystal structure coordinate sets which exist for full mAb molecules (for example 1HZH and 1IGT) that the structures are far from spherical, and a level of gas-phase collapse has been observed (*88*). Therefore, the above theoretical power relationship (*vide supra*) may not be consistent for mAbs. Additionally, Benesch (*89*) demonstrated that depending on the position of the nanoflow ESI needle tip in relation to the mass spectrometer sample cone, the average charge can vary by up to seven charges for the HSP16.5 complex.

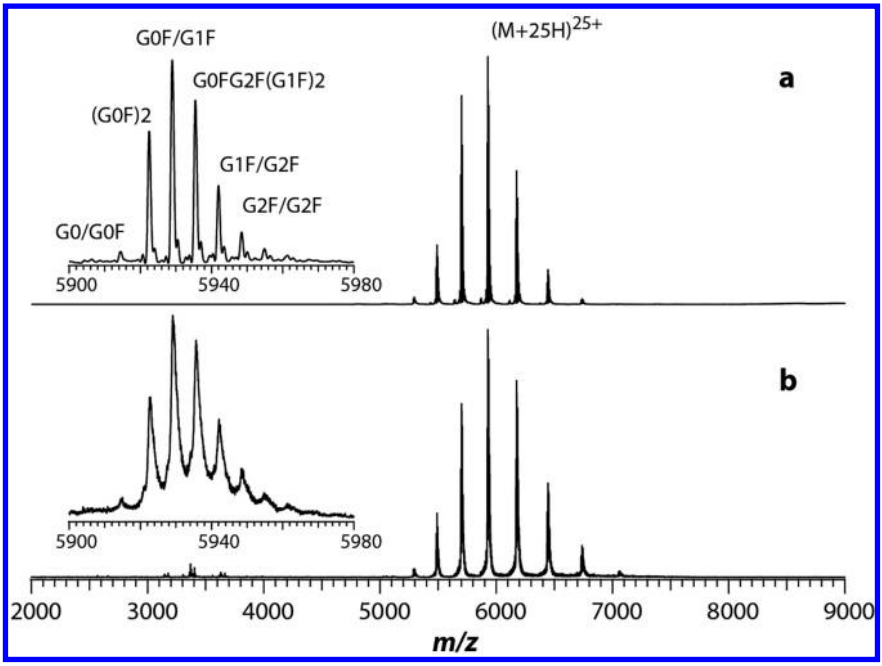

Figure 4. A comparison of NISTmAb mass spectrometry (MS)-spectra acquired on the OrbiTrap Exactive Plus extended mass range (EMR) instrument (Thermo Scientific; upper spectrum) and the Synapt G2 HDMS (Waters Corporation; lower spectrum) under native buffer and MS conditions. Glycan structure annotation: G0, $GlcNAc_2Man_3GlcNAc_2$; G0F, $GlcNAc_2Man_3GlcNAc_2Fuc$; G1F, $GalGlcNAc_2Man_3GlcNAc_2Fuc$; G2F, $Gal_2GlcNAc_2Man_3GlcNAc_2Fuc$.

Under native MS conditions, the individual glycoforms (G0/G0F, G0F/G0F, G0F/G1F, G0FG2F [G1F/G1F], G1F/G2F, G2F/G2F) of each charge state can be efficiently resolved by native MS, allowing for relatively easy identification (Figures 4a and b). Considering the vastly differing ion optics of both instrument platforms, dimeric NISTmAb only constitutes less than 5% of the monomer species (data not shown). The glycoform resolution is only achieved at significantly increased source voltages. Both instruments are operated with very high source voltages. For example, the Synapt G2 HDMS sample cone must be

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

operated at 150 V or above to obtain the demonstrated glycoform resolution. The OrbiTrap instrument also operates at an elevated in-source CID voltage of 125 V and an HCD trap voltage of 25 V. It is very clear from the Synapt G2 data in Figure 5b that at a cone voltage above 100 V, there is a significant amount of mAb unfolding. Increasing the instrument source and trapping voltages are clearly advantageous for improving spectral quality and mass resolution of the glycoforms within each charge state; however, it must be noted that the TWIM data indicates the mAb is clearly not in its native conformation and is in a gas-phase unfolded state. At a cone voltage of 50 V, a single ATD can be observed (Figure 5a); however at 200 V, where the glycoforms can be resolved close to baseline, the ATD is no longer single and symmetrical, but it is broad, extended, and bimodal, indicating a significant amount of protein unfolding (Figure 5d). Without an IM device proceeding the ion source in the OrbiTrap EMR instrument, it is challenging to understand whether the gas-phase structure of the protein under investigation is native-like or in an unfolded conformation. Given the increased voltages used to generate the levels of protein desolvation observed within the Orbitrap EMR instrument, it must be assumed that proteins are no longer in their native gas-phase conformation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

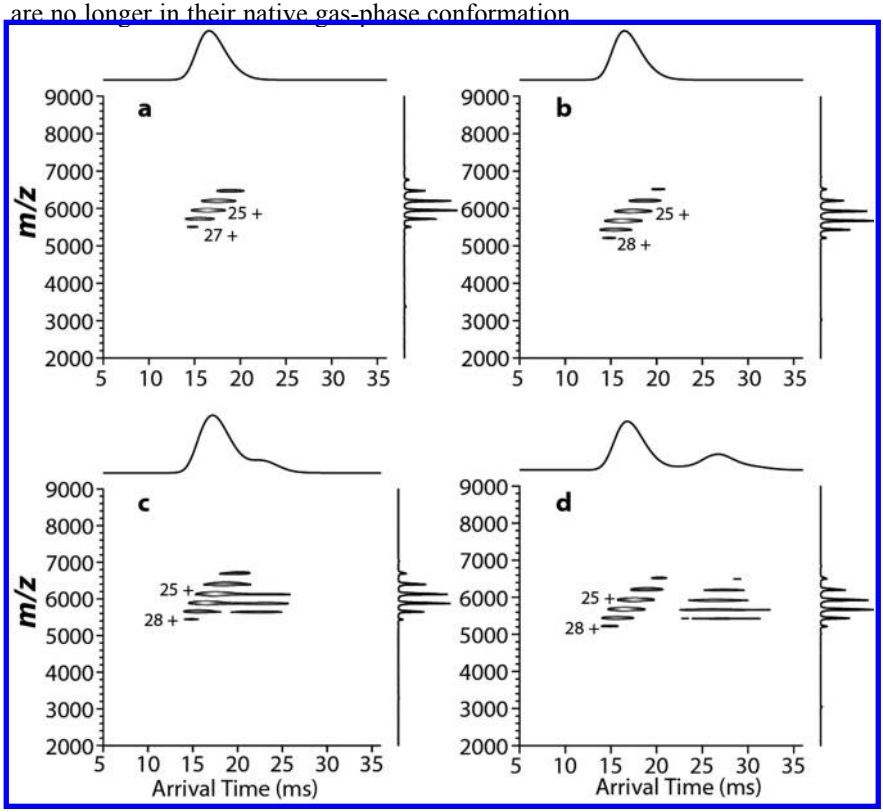

Figure 5. NISTmAb acquired over a range of sample cone voltages: a: 50 V; b: 100 V; c: 150 V, and d: 200 V, under native buffer conditions on the Synapt G2 HDMS instrument. Travelling wave ion mobility (TWIM) N_2 pressure 3.00 Torr (4 mbar).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

IM Analysis

Figure 6 shows a typical *m/z* versus mobility plot for analysis of the NISTmAb acquired on the RF-confining instrument. Under native MS conditions, the native NISTmAb charge states (+21 to +26) appear in the *m/z* range 5700 to 7100, and under a drift voltage of 75 V, the drift time is on the order of 16 to 20 msec, corresponding to a Ω_{N2} range, for all charge states, ranging from 7200 to 7400 Å^2 (Table 1). All major charge states (+21 to +26) display a single, uniformly shaped ATD (Figure A6, Appendix), which would indicate an ensemble of closely related gas-phase structures and very little change in gas-phase Ω as a function of charge state.

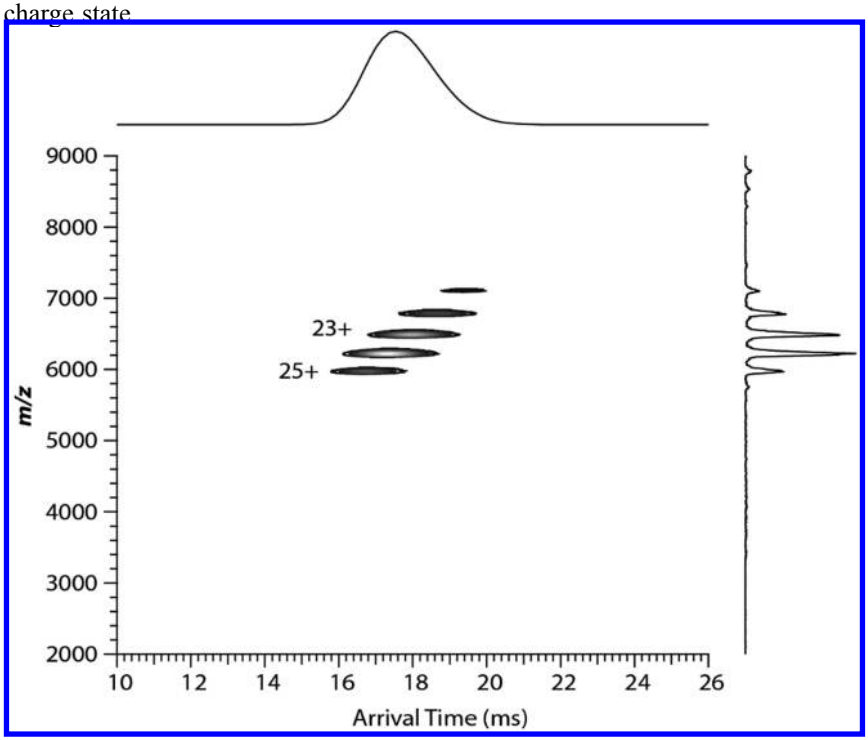

Figure 6. An RF-confining drift cell ion mobility (IM) plot (m/z versus drift time in msec) of the NISTmAb acquired under native buffer (50 mM ammonium acetate) and MS conditions, at a protein concentration of 5 μM. radio frequency (RF)-confining drift cell data displayed represents a drift cell voltage of 75 V (4.16 V/cm) and N_2 drift gas pressure of 1.55 Torr.

Based on the Ω values (both He and N_2) reported in Table 1, we can infer that the NISTmAb is more than likely in a native-like gas-phase conformation, because all measured Ω values are close to that of ADH, which is similar in molecular weight (147.5 kDa; Ω_{N2} value 7500 Å^2 for charge states +23 to +25; measured on our RF-confining drift cell instrument). Ruotolo et al. (*65*) have demonstrated that for a large range of proteins varying in molecular weight and subunit stoichiometry (5kDa to 200 kDa), a trend does exist between molecular weight and Ω value, based on the relationship $\Omega = 2.435 \times Mw^{2/3}$, thus adding

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

support to the hypothesis that the NISTmAb and ADH should, in theory, possess very similar Ω values. However, if we actually consider the widths (at half height) of the RF-confining, drift cell-measured ATDs, we find that for all charge states, the NISTmAb ATDs are consistently wider than the ADH ATDs by 24% to 43% in both He and N_2 drift gases. The higher charge states (more charge) displaying the greater difference than the lower charge states (less charge). This is consistent with observations made by Pacholorz (*88*), suggesting that the NISTmAb and other mAbs (*88*) display measureable levels of gas-phase flexibility (*vide infra*). The NISTmAb is therefore likely in an ensemble of native-like conformations with ATDs that may be characteristic of its gas-phase properties.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Table 1. RF-Confining Drift Cell and Travelling Wave Ion Mobility (TWIM) Theoretically Derived Ω_{He}, Ω_{N2}, and K_o Values

RF-confining drift cell					*TWIM*
Charge State z	Ω_{N2} *(Å²)*	K_{oN2} *(cm²/V.s)*	Ω_{He} *(Å²)*	K_{oHe} *(cm²/V.s)*	Ω_{N2} *(Å²)*
21	7223 (±51)	0.59	6696 (±7)	1.69	7310 (±12)
22	7275 (±29)	0.62	6759 (±55)	1.75	7306 (±7)
23	7257 (±28)	0.65	6820 (±4)	1.82	7301 (±15)
24	7310 (±0)	0.67	6866 (±14)	1.88	7279 (±12)
25	7376 (±24)	0.69	6910 (±4)	1.95	7246 (±6)
26	7403 (±9)	0.72	6892 (±28)	2.03	7247 (±9)
27	*7442 (±0)	0.74	–	–	7292 (±29)
28	*7533 (±149)	0.76	–	–	7333 (±124) 8151 (±23)
29	*7726 (±401) *8988 (±649)	0.75 0.66	–	–	8032 (±1)

* Denotes very low intensity for observed charge states. Charge states +27, +28 and +29 were not observed in the drift gas He. All Ω values are represented in units of Å². K_o values are represented in units of cm²/V·s. Values in parentheses represent the average standard deviation (Å²). Standard deviation values for the RF-confining drift cell instrument were derived from two separate measurements made over 10 different drift cell voltages. Standard deviation values for the Synapt G2 HDMS instrument were derived from separate measurements made over three different travelling wave amplitudes, 19 V, 20 V, and 21 V, all at wave speed of 250 m/sec. RF = radio frequency.

The high charge states (+28 and +29) display a range of different mobility profiles (Figure A6 and A7, Appendix). For example, if we consider the mobility data acquired in N_2 on the RF-confining instrument, the +28 charge displays one major conformation (7533 Å²) and an extended, very low intensity conformation. The +29 charge state shows a distinct bimodal distribution, one compact (7726 Å²) relative to the more extended conformation (8988 Å²). Note that these charge

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

states were present at very low ion intensity; therefore, the errors associated with the mobility measurements for the +28 and +29 charge states are high (Table 1). The bimodal and extended conformations for higher charge states are also mirrored on the Synapt G2 HDMS TWIM device (Figure A7, Appendix). However, the +29 charge state measured on the Synapt G2 HDMS instrument is predominantly the extended conformation (8032 Å^2). This would suggest that the conditions used within the Synapt G2 HDMS instrument results in an increased amount of ion activation/gas-phase unfolding. Charge states +27 to +29 are not observed when He is used as the drift gas on the RF-confining drift cell instrument. Differences in charge state distribution as a function of drift gas have been observed previously on an RF-confining drift cell instrument (*52*). If we compare the Ω_{N2} values derived on both instrument platforms (RF-confining drift cell and TWIM device) using very different IM-separation and Ω value derivation techniques, we can see there is a very high degree of correlation between all NISTmAb charge states. This gives us great confidence in the Ω values that we are deriving and subsequently reporting with this chapter.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

MD Simulations

If the unoptimized NISTmAb coordinate file (Higher Order Structure chapter/Volume 3, Chapter 2) was used directly for theoretical Ω calculations, the PA algorithm calculates a value of 8997 Å^2 for N_2 and 8637 Å^2 for He, both of which are significantly (20%) higher than the largest experimental value (7403 Å^2 in N_2 and 6829 Å^2 in He; both for the +26 charge state; Table 1). It is known that the PA routinely underestimates Ω values, so typically, a multiplication factor of 1.14 is applied (*40*) or more recently the value of 1.36 (*84*) to 1.38 (*90*) to account for ion-neutral scattering events. The resulting PA values for N_2 and He, using a multiplication factor of 1.36, would be 12,302 A^2 and 11,746 A^2, respectively. The initial solution-based structure must consequently undergo a significant compaction process to reach its native-like gas-phase conformation, which were accurately measured on both the RF-confining and TWIM instrumentation. This compaction process also must be relatively fast (sub-msec) because the mobility experiments take place over 18 msec (drift cell 60 V), and single ATDs are observed for the major charge states/structures (+21 to +26), suggesting that no gas-phase rearrangements, such as unfolding or additional compaction restructuring, is occurring during the time frame of the IM experiment. Figure 7 displays the MD evolution at times 0, 5, 10, 30, 50, and 100 psec. The process is particularly enlightening because it demonstrates how the NISTmAb structure collapses as much as 40% once it has reached its globular configuration (at 100 psec). The final MD configuration (100 psec) yields a PA Ω value of 5282 A^2 and 5125 A^2 for N_2 and He, respectively, significantly lower than the fully extended conformation Ω values *vide supra*. However, even using the multiplication factor of 1.14 (6021 A^2 and 5842 A^2; N_2 and He respectively) these new PA values for the 100 psec compacted NISTmAb structure, are still significantly lower than the instrument-measured Ω values, suggesting that the PA method is not ideal for mAb theoretical Ω calculation.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

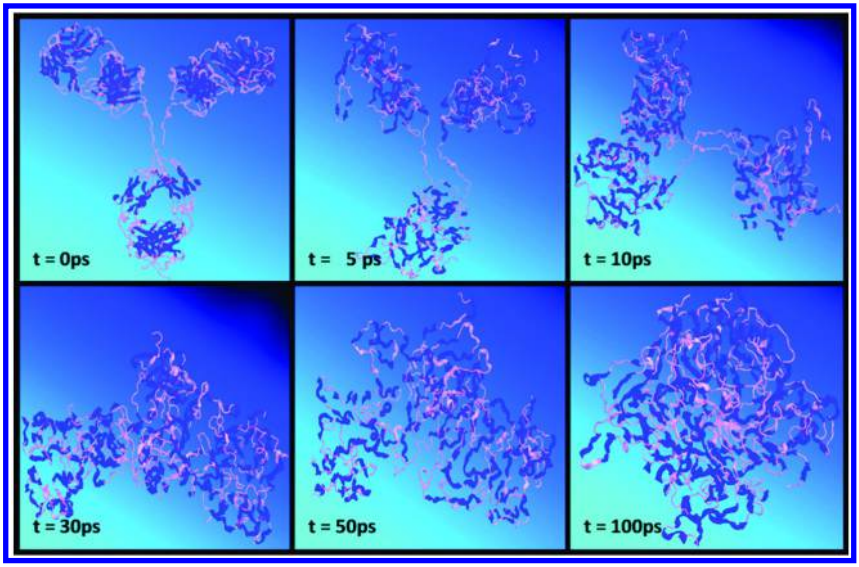

Figure 7. The temporal evolution of the NISTmAb theoretical structure during an "in vacuo" molecular dynamic simulation using the MM2 force-field in Chem 3D Pro, shown at 0, 5, 10, 30, 50, and 100 psec.

Additional theoretical Ω values derived with the IMoS software (EHSS, DHSS and TM) are shown in Table 2, which produce highly consistent Ω values compared to instrument-derived values (Table 1). Recently, Pacholarz et al. (*88*) have documented similar results for an IgG1 and IgG4 molecule, where they also demonstrated significant gas-phase compaction; however, their data does display a reasonable amount of discrepancy between their instrument and theoretically derived Ω values (He only). Their reported theoretical values are higher than their drift tube Ω_{He} values for charge states +22 to +25. It appears that they do not observe the same amount of MD-derived gas-phase compaction as we report herein, although they do face the same problem we experienced, in that there is no crystal structure coordinate file for the full intact mAb being analysed by IM. Based on our IM and MD studies, it is not actually surprising that we observe such a large amount of gas-phase compaction, because it has been reported 20 years ago using cryo-electron microscopy that mAbs are very flexible (*91*) with reported mean Fab-Fab angles for an IgG3 mAb of 136°. The MD simulation performed by Pacholarz (*88*) also suggest a 6 to 7% decrease in helical content, a 41 to 47% decrease in β-sheet content, and also major contractions around the hinge region.

Previously, groups (*45*, *88*) have demonstrated that theoretically derived Ω values for an intact IgG1 mAb is 8653Å^2 (PDB code 1HZH; PA-derived Ω value (*45*)) and an IgG2A mAb is 10,600 Å^2 (PDB code 1IGT; TM-derived Ω value (*88*)); these are significantly higher than instrument-derived Ω values. We also clearly demonstrate this finding with the NISTmAb; however, there is no X-ray

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

crystallographic-derived coordinate file for the NISTmAb. Nevertheless, it is clear a large amount of gas-phase compaction is occurring to mAbs once they transition from solution to the solution-free gas-phase environment. Based on all these findings reported herein, as well as those of others (*45*, *88*), it is very challenging, to say the least, to infer any real structure of a mAb in the gas phase by comparing it to the crystal structure coordinate file which has not undergone any prior *in vacuo* MD structure optimisation. The mAb structure within the crystal unit cell and potentially in solution is clearly dramatically different to the gas-phase conformation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Table 2. Theoretical Ω_{N2} and Ω_{He} Value Calculations of the NISTmAb as a Function of Different Charge States and the Different Ω Algorithms Used in the IMoS Suite of Software

Charge State z	*PA N_2 (Å²)*	*EHSS N_2 (Å²)*	*DHSS N_2 (Å²)*	*TM N_2 (Å²)*	*PA He (Å²)*	*EHSS He (Å²)*	*TM He (Å²)*
0	7183	6850	6952	7421	6971	6765	6730
23	–	7083	7189	7614	–	6758	6750
27	–	7162	7296	7688	–	6774	6722

Projected area approximation (projected area [PA] multiplied by correction factor 1.36); elastic/exact hard sphere scattering (EHSS); diffuse hard sphere scattering (DHSS) both with ion-induced dipole potentials (4–∞); trajectory method (TM) using Lennard-Jones potentials, including polarization (4-6-12 potential).

In order to interpret both RF-confining drift cell (Table 1) and theoretically derived Ω_{N2} values (Figure 7 and Table 2) further, Figure 8 plots theoretical and experimental Ω values as a function of z^2 for N_2. Out of the many features which stand out in Figure 8, the most notable is the well-defined slope that the experimental values (RF-confining drift cell) display as a quadratic function of the charge (z). The slope remains constant until the mAb charge states reach +27, where the slope immediately increases in value. Although surprising at first, this is a known feature of collapsed chain-like structures such as proteins and heteropolymers in the gas phase. The initial portion of the slope (charge states +21 to +26) can be attributed to a combination of the ion-induced dipole potential and a slight enlargement of the protein due to the repulsion of the surface charges. One could also interpret these data as follows: the lower charge states (and therefore lower Ω values) could potentially possess more favourable ion-pair interactions on the surface, leading to a more compact structure; this effect is lessened in the higher charge states (+28 to +29). Furthermore, the change which leads to the second, steeper slope (+28 to +29) simply corresponds to the unfolding of the protein due to having achieved the maximum possible charge on its surface. Once this maximum charge is achieved, the protein must undergo changes in its

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

structure to accommodate a larger number of charges in its surface, leading to much larger Ω value. Of all the theoretical methods performed for calculating the Ω_{N2} value of the MD energy-minimised NISTmAb structure, the algorithm that is most accurate is the DHSS calculation with polarization (< 2% error); followed by EHSS (< 3%); and finally TM with optimized Lennard-Jones parameters for N_2 (18), which overestimates the RF-confining drift cell Ω_{N2} values by only less than 5%. Considering the resolution of the experimental ATDs and previously reported errors (~4% (*52*)), excellent agreement is achieved between instrument (RF-confining and TWIM) and theoretically derived Ω_{N2} and Ω_{He} values.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

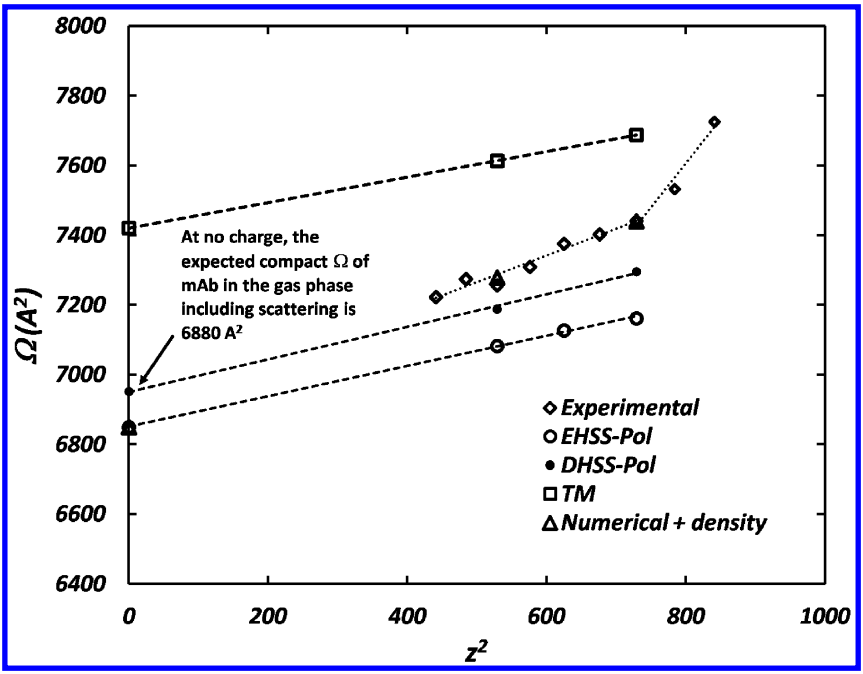

Figure 8. A comparison of experimental (radio frequency [RF]-confining drift cell) and theoretical Ω_{N2} values plotted as a quadratic function of ionic charge. EHSS-Pol refers to elastic hard sphere scattering with dipole potentials. DHSS-Pol refers to diffuse hard sphere scattering with dipole potentials. TM refers to trajectory method. Numerical + density corresponds to a coulombic enforced volume enlargement correction value determined from matching experimental data, where the Ω resulting from DHSS (EHSS) is increased by b = 0.20 (0.37) using equation $CCS = CCS_0 + (\alpha + \beta)z^2$ (provided by the IMoS code and inferred to be a value of 0.4405).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Collision-Induced Unfolding

A single Ω value (N_2 and He) for each charge state has been derived both the experimentally and theoretically (Table 1) for the NISTmAb. There is very good agreement between both experimental and theoretical Ω values; however, it is after all just a single value. A molecule as large as a mAb can potentially possess many subtly different gas-phase tertiary structures that are clearly unresolvable using current IM technologies unless a gross structural change occurs as a result of ligand binding, for example. Recent studies (*35*, *61*, *92*, *93*) describe CIU as a possible mechanism for mapping and identifying subtle structural differences between two or more related proteins (mAbs for example) or native protein gas-phase stability, where a single Ω value would fail to produce anything meaningful.

Figure 9 displays a CIU experiment performed on the +26 charge state of the NISTmAb. Between trap cell activation voltages of 4 V to 30 V, there is very little change in the ATD of the +26 charge state. At 40 V, there is evidence of a trailing edge, suggesting the unfolding process is beginning. At 50 V, there are two dominant gas-phase conformations. Between trap cell voltages of 60 V and 100 V, dramatic unfolding profiles can be observed, with measured Ω_{N2} values as high as 10,700Å^2 being detected. Analogous experiments performed on groups of related mAbs—such as the human immunoglobulin isoforms IgG1, IgG2A, and IgG2B isoforms—may prove beneficial as a method of producing low-resolution finger-printing data, enabling isoform differentiation and ultimately identification, based on gas-phase CIU (*94*). Recently, Zhong (*95*) has demonstrated a very strong correlation between the number of IM-measured unfolding events and the number of known protein domains within a group of 16 proteins. It is interesting to postulate that the initial unfolding events (40–60 V) we observe in Figure 9 may potentially take place around the flexible hinge region. Further unfolding events at higher collision energies (60–100 V) may be a result of more localised tertiary domain unfolding, for example, CIU of the variable and constant domains within the Fc (fragment crystallizable region) and Fab (fragment antigen binding) regions. The above inferences would require validation by MD simulation, similar to those previously performed by Hall (*96*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

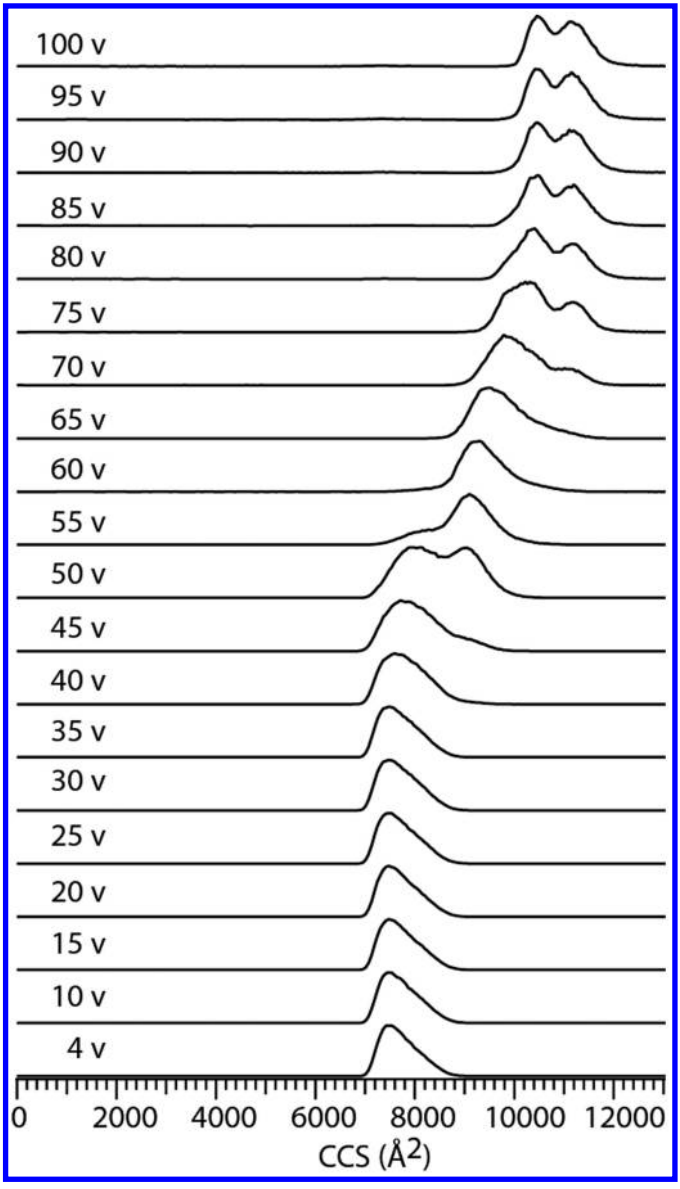

Figure 9. Collision-induced unfolding (CIU) of the NISTmAb. Quadrupole selection of charge state +26 (m/z 5720) and trap cell activation using a voltage range of 4 to 100 V. Sample cone was set to 50 V. Infusion was performed under native buffer conditions (50 mM ammonium acetate). Data are displayed as a function of normalised intensity versus calibrated travelling wave ion mobility (TWIM) collision cross-section scale (Ω_{N2}).

Conclusion

Using a quadrupole ion mobility time-of-flight (Synapt G2) and OrbiTrap (Exactive EMR) mass spectrometers, a narrow charge state distribution (+21 to +26) was observed. Well-resolved, and in the case of the OrbiTrap instrument, baseline-resolved glycoforms for each charge state (G0/G0F, (G0F)2, G0F/G1F, G0FG2F/(G1F)2, G1F/G2F, G2F/G2F) was achieved for the NISTmAb. Surprisingly, at an MS working concentration of 5 μM, the charge state distributions are incredibly similar under native-MS and buffer conditions, considering the dramatic differences in instrument ion optics in terms of both source and analyser. Additionally, the stoichiometry also was very similar; low levels of NISTmAb dimer were observed on both instruments. Glycoform resolution under native-MS and buffer conditions only can be achieved using a high sample cone voltage (150–200 V) on the Synapt G2 instrument. On the OrbiTrap, high activation voltages also are utilised: in-source CID voltage of 125 V and an HCD voltage of 25 V. Achieving glycoform resolution results in a significant amount of gas-phase activation, and therefore protein unfolding, which can only be demonstrated and quantified on the Synapt G2 system by utilising the TWIM device; at a sample cone voltage of 150 V and above, a significant increases in TWIM drift times can be observed. It is clear that the Exactive-EMR instrument is an attractive option for native protein analysis; however, the impressive resolution is only achieved at high declustering (activation) voltages (in-source CID and HCD voltage), which is likely causing significant gas-phase unfolding, the level of which cannot be quantified, due to the absence of an IM device.

Mobility and therefore gas-phase Ω measurements were derived for the NISTmAb on both an RF-confining drift cell and TWIM instruments. The Ω_{N2} values were highly consistent across both platforms despite fundamental instrument differences. For the dominant charge states, +21 to +26, RF-confining Ω_{N2} and Ω_{He} values covering the range 7223 to 7403Å^2 and 6696 to 6892 Å^2, respectively. The MD simulations indicate that over a very short period of time (100 psec), a significant amount of gas-phase compaction is observed. For example, the unoptimised starting structure has a Ω_{N2} value of 8997 Å^2 (PA-derived) compared to the 5282 Å^2 (PA-derived) for the 10 nsec optimised structure. This difference represents approximately a 40% reduction in theoretical Ω value. Additionally, it is clearly not advisable to compare gas phase-derived Ω values to those theoretically derived from unoptimised (by MD force fields) mAb coordinate files, which have been obtained from X-ray crystallographic data sets. Once the mAb is transferred from solution to the gas phase, there is clearly a large amount of structural reorganisation taking place in a very short space of time, which is obviously not reflected in an X-ray crystallographic data set.

The PA method is known to underestimate the Ω value for proteins because it does not account for the gas scattering and ion-induced dipole interactions between the charged protein and neutral gas atom (or molecule); therefore, more comprehensive theoretical Ω calculations were performed using the EHSS, DHSS, and TM algorithms, using both N_2 and He as the neutral drift gases, to compare to our instrument-derived Ω values. Our theoretical Ω values in both N_2

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

and He for the optimised gas-phase collapsed structure are highly consistent with both our RF-confining drift cell and TWIM measurements. If we consider the +23 charge state, for example, there is only a 0.9% difference between theoretical (DHSS) and RF-confining drift cell-derived Ω values for both and N_2 and He drift gases. Based on our theoretical Ω calculations, it is our opinion that the PA method is therefore not suitable for theoretical Ω value calculations for mAb structures and importantly, an X-ray crystallographic-derived 3-D coordinate set is not representative of the gas-phase structure of a mAb. The discrepancy between theoretical and instrument-derived Ω values has been demonstrated previously with the protein tryptophan-RNA-binding attenuation protein (*30*). It was shown that only the +19 charge state has a measured Ω value consistent with the crystal structure, and charge states +20 to +22 all display increasing levels of structural collapse. As demonstrated with the NISTmAb, all charge states display large levels of collapse when compared with theoretically derived and closely related IgG1 coordinate sets.

CIU is potentially a powerful method that can be used to identify subtle differences in closely related biological samples. The example described herein is a little contrived because there is no related mAb to compare to; however, it does demonstrate that a significant amount of structural information potentially can be inferred from a simple experiment. The CIU experiment described herein was performed over 5 V step increments, but to obtain more granularity, 1 V increments could be easily performed to "tease out" even more detail as a function of native mAb unfolding. The IM experiments combined with the MD simulation demonstrate very elegantly how instrument- and theoretically derived Ω values can match to within a few percent or better. However, the derived Ω value is only a single number; therefore, only limited inferences can be made for such a large protein molecule. Therefore, using an additional orthogonal technique such as CIU to infer structural detail potentially could be incredibly useful. CIU also could be used as a complimentary characterisation/profiling technique to the already existing nuclear magnetic resonance (NMR) profiling methodology (*97*) allowing for gas-phase CIU and solution-phase comparisons of mAbs to be made.

Monoclonal antibody isoforms and heterogeneity are common and can lead to a high level of sample complexity. Post-translational modifications such as glycosylations, deamidations, disulphide variants, oxidations, and N-terminal and C-terminal modifications may appear small, or even undetectable on the mass scale, but may have profound effects on the mAb structure and, therefore, potency. As techniques such as IM, native MS, MD simulations, and CIU establish themselves within the area of mAb research, they can potentially aid in the characterisation of the subtle aforementioned changes.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Acknowledgments

The authors would like to thank Kevin Giles and Michael Morris from the Waters Corporation, MS Technologies Centre in Manchester, UK, for their help in converting our production Synapt G1 TWIM instrument into an RF-confining IM device. As a result, the data presented and described within this manuscript is much stronger. We also greatly acknowledge Brandon Ruotolo from the University of Michigan, MI, USA, for his useful and open discussions regarding collision-induced unfolding of monoclonal antibodies. We also would like to thank Alistair Wallace from the Waters Corporation, MS Technologies Centre in Manchester, UK, for kindly allowing us to use the Synapt G2 HDMS instrument schematic displayed in Figure 1. Finally, we would also like to thank Detlef Schumann from Thermo Fisher Scientific, Bremen, Germany, for kindly allowing us to use the Exactive Pro EMR schematic displayed in Figure 3.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Appendix

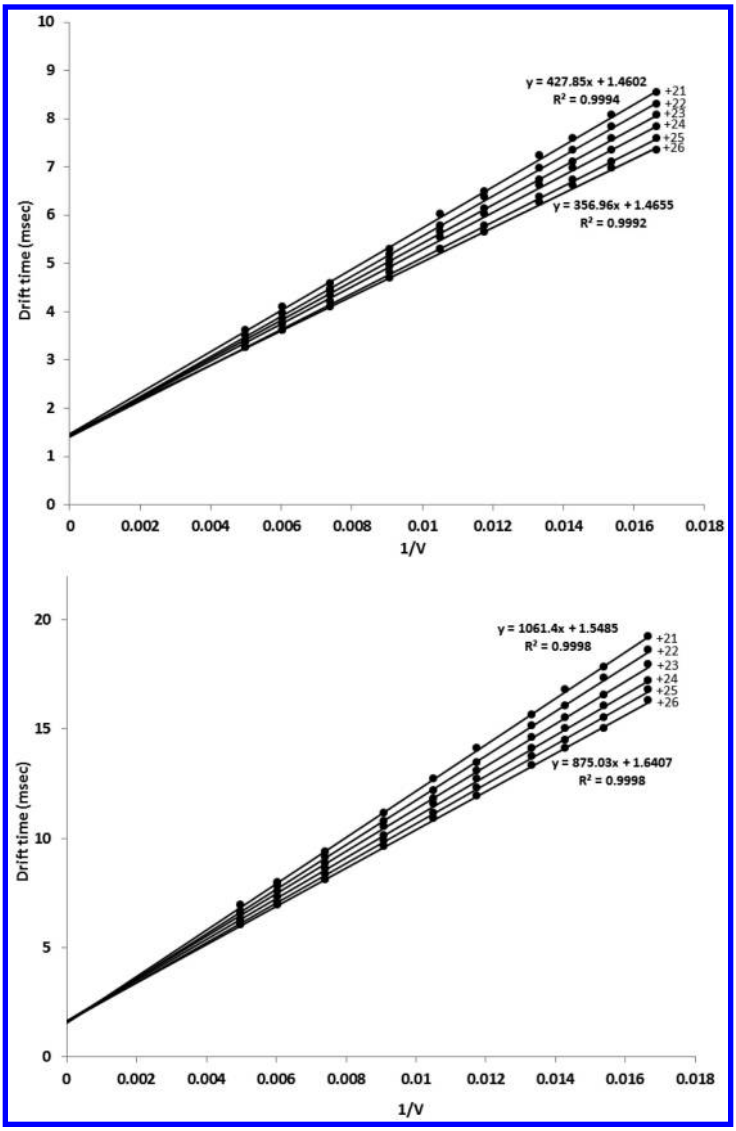

Figure A1. Radio frequency (RF)-confining drift cell drift time versus 1/V plot for the observed NISTmAb charge states +21 to +26 under native buffer (50 mM ammonium acetate) and mass spectrometry (MS) conditions in drift gases He (upper) and N_2 (lower). Drift voltages used were 60 V, 65 V, 70 V, 75 V, 85 V, 95 V, 110 V, 135 V, 165 V, and 200 V. Sample cone 25 V; ion mobility (IM) bias 14 V (He) and 16 V (N_2); extraction cone 1 V; trap collision energy 3 V; IMS RF-voltage 150 V peak-to-peak.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

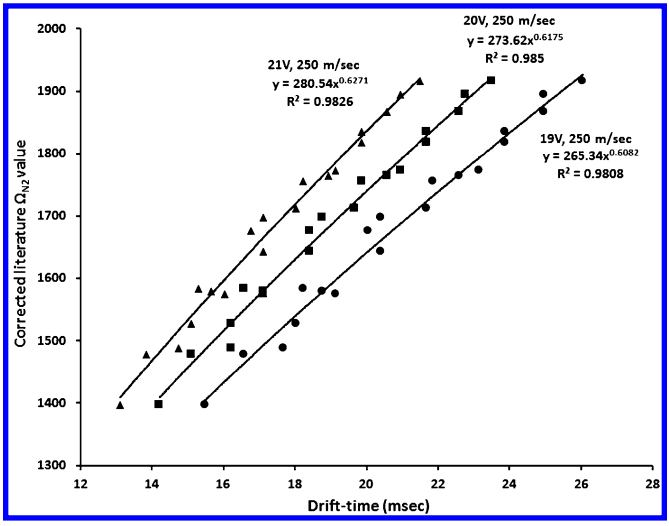

Figure A2. Travelling wave ion mobility (TWIM) calibration plot used to calculate Ω_{N2} values for the NISTmAb. Calibration proteins analysed under native buffer (100 mM ammonium acetate) and mass spectrometry (MS) conditions: bovine serum albumin (BSA), serum amyloid protein (SAP), alcohol dehydrogenase (ADH), pyruvate kinase (PK), and glutamate dehydrogenase (GDH). TWIM calibration was performed at three different travelling wave amplitudes: 19 V, 20 V, and 21 V, all at a speed of 250 m/sec.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

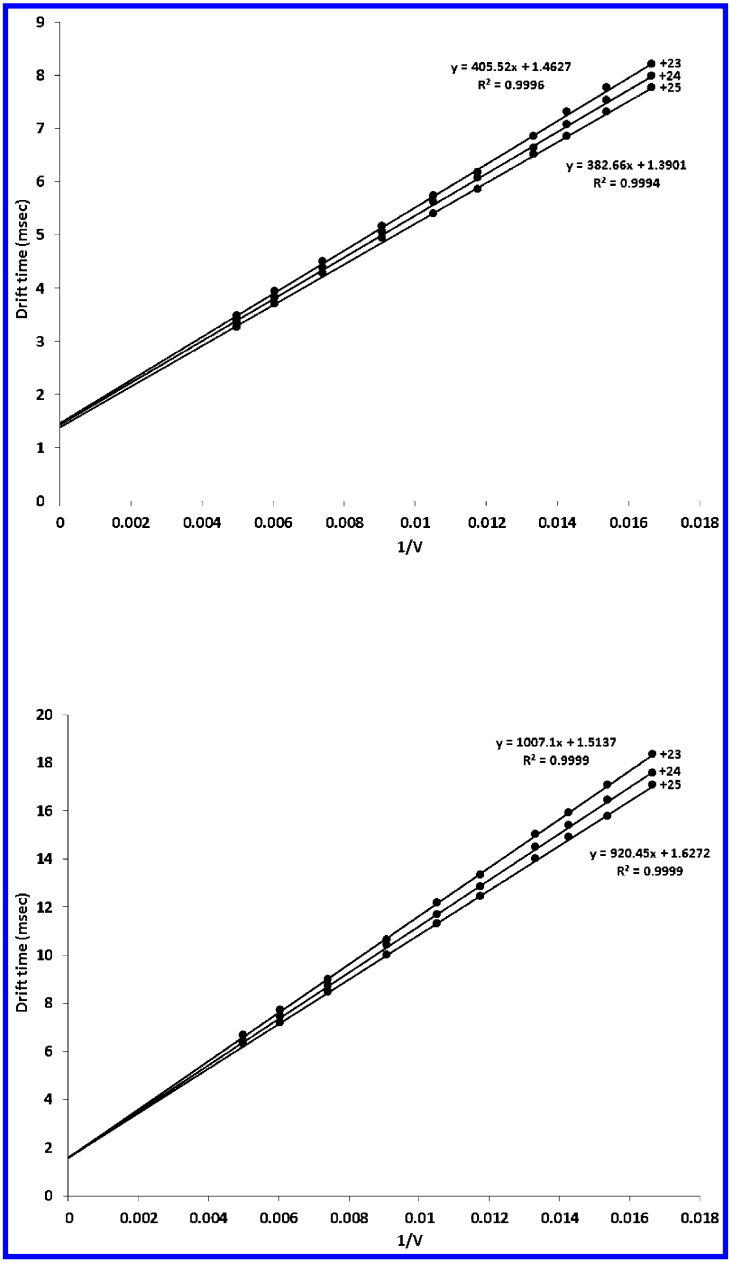

Figure A3. Radio frequency (RF)-confining drift cell drift time versus 1/V plot for the observed alcohol dehydrogenase (ADH) charge states +23, +24, and +25, under native buffer (100 mM ammonium acetate) and mass spectrometry (MS) conditions, in drift gases He (upper) and N_2 (lower). Drift voltages used were 60 V, 65 V, 70 V, 75 V, 85 V, 95 V, 110 V, 135 V, 165 V, and 200 V. Sample cone 25 V; ion mobility (IM) bias 14 V (He) and 16 V (N_2); Extraction cone 1 V; trap collision energy 3 V; IMS RF-voltage 150 V peak-to-peak.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

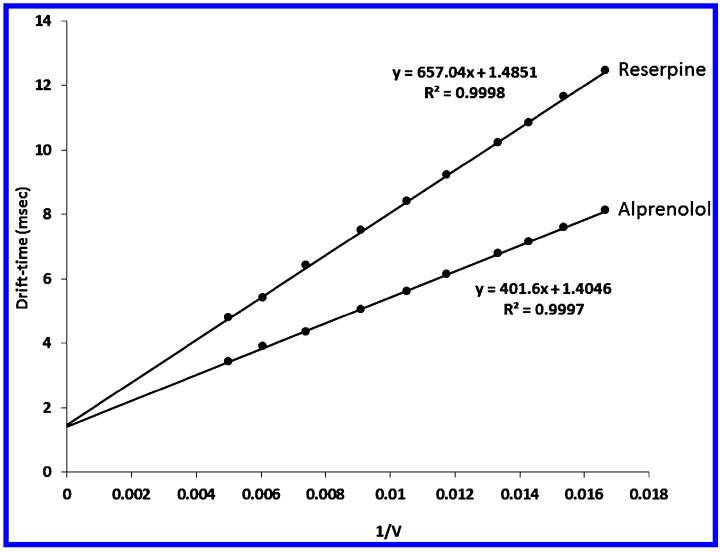

Figure A4. Radio frequency (RF)-confining drift cell drift time versus 1/V plot for the singly charged, small, drug-like molecules alprenolol and reserpine in the drift gas N_2. Drift voltages used were 60 V, 65 V, 70 V, 75 V, 85 V, 95 V, 110 V, 135 V, 165 V, and 200 V. Sample cone 25 V; ion mobility (IM) bias 14 V (He) and 16 V (N_2); extraction cone 1V; trap collision energy 3 V; IMS RF-voltage 150 V peak-to-peak.

Table A1. The RF-Confining Drift Cell Ω_{N2}-Derived Values for Alprenolol and Reserpine, and Ω_{He} and Ω_{N2} values for Alcohol Dehydrogenase (ADH)

	Ω_{He}	Ω_{N2}	$^{*}\Omega_{He}$	$^{*}\Omega_{N2}$
Alprenolol $z = 1$	–	159.8	96.9	157.5
Reserpine $z = 1$	–	253.9	178.8	254.3
ADH $z = 23$ $z = 24$ $z = 25$	 7008 7099 7110	 7526 7514 7510	 6940 6940 6830	 7420 7450 7440

* Denotes published literature Ω values. Mobility measurements in He were not made for alprenolol or reserpine. Values were obtained on the instrument described and used within this manuscript and compared to the published literature values (*18*, *52*).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

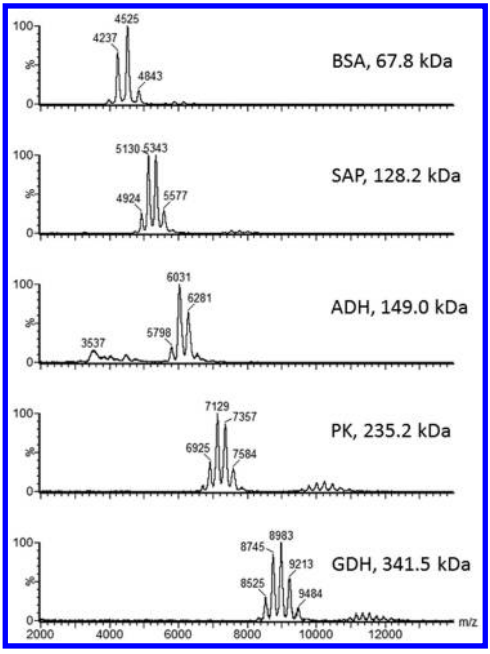

Figure A5. Representative native mass spectrometry (MS) spectra for the travelling wave ion mobility (TWIM) calibration standards bovine serum albumin (BSA), serum amyloid protein (SAP), alcohol dehydrogenase (ADH), pyruvate kinase (PK), and glutamate dehydrogenase (GDH). All data acquired over the m/z range 1000 to 14,000, under native buffer (100 mM ammonium acetate) and MS conditions. All proteins infused at a concentration of 5 μM.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

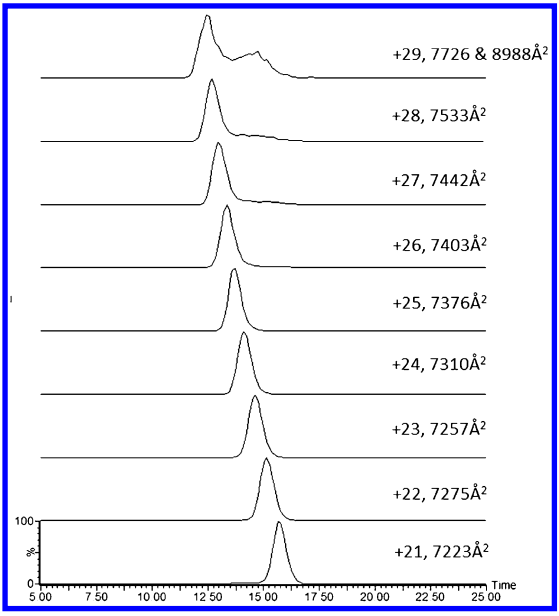

Figure A6. Individual arrival time distributions and annotated Ω_{N2} values, for all the observed NISTmAb charge states (+21 to +29) acquired under native buffer (100 mM ammonium acetate) and mass spectrometry (MS) conditions using a drift voltage of 75 V, in the drift gas N_2, on the radio frequency (RF)-confining drift tube instrument. Sample cone 25 V; ion mobility (IM) bias 14 V (He) and 16 V (N_2); extraction cone 1 V, trap collision energy 3 V; IMS RF-voltage 150 V peak-to-peak. Charge states +27, +28, and +29 are of very low intensity. X-axis in msec.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

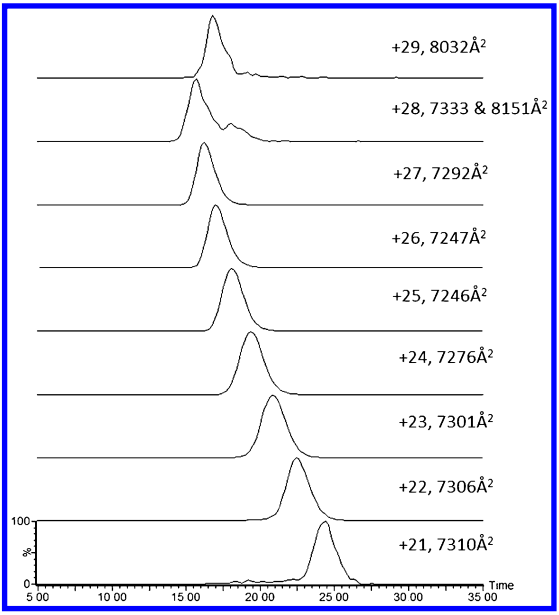

Figure A7. Individual arrival time distributions for all the observed NISTmAb charge states (+21 to +29) acquired under native buffer and mass spectrometry (MS) conditions using a travelling wave amplitude and speed of 19 V and 250 m/sec, respectively, on the Synapt G2 HDMS instrument using the travelling wave ion mobility (TWIM) gas N_2. The annotated Ω_{N2} values are an average derived from acquisitions made at 3 travelling wave amplitudes. 19 V, 20 V and 21 V. Sample cone 25 V; trap bias 45 V; ion mobility (IM) bias 3 V; Extraction cone 1 V, trap collision energy 3 V; IM RF-voltage 250 V peak-to-peak. Synapt G2 HDMS instrument only operated with N_2 as the drift gas. X-axis in msec.

References

1. McDaniel, E. W.; Martin, D. W.; Barnes, W. S. *Rev. Sci. Instrum.* **1962**, *33*, 2.
2. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Jr. *J Mass Spectrom* **2008**, *43*, 1–22.
3. Eiceman, G. A.; Karpas, Z.; Hill, H. H. *Ion Mobility Spectrometry*; CRC Press: Boca Raton, FL, 2010
4. Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. *Int. J. Mass Spectrom.* **2009**, *287*, 46–57.
5. Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. *Annu. Rev. Anal. Chem.* **2008**, *1*, 293–327.
6. Hill, H. H., Jr.; Siems, W. F.; St Louis, R. H.; McMinn, D. G. *Anal. Chem.* **1990**, *62*, 1201A–1209A.
7. Guevremenot, R. *J. Chromatogr. A* **2004**, *1058*, 3–19.
8. Kaufman, S. L.; Skogen, J. W.; Dorman, F. D.; Zarrin, F.; Lewis, K. C. *Anal. Chem.* **1996**, *68*, 1895–1904.

9. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401–2414.
10. Steiner, W. E.; Clowers, B. H.; Matz, L. M.; Siems, W. F.; Hill, H. H., Jr. *Anal. Chem.* **2002**, *74*, 4343–4352.
11. Shi, H.; Pierson, N. A.; Valentine, S. J.; Clemmer, D. E. *J. Phys. Chem. B* **2012**, *116*, 3344–3352.
12. Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2011**, *115*, 12266–12275.
13. Pierson, N. A.; Valentine, S. J.; Clemmer, D. E. *J. Phys. Chem. B* **2010**, *114*, 7777–7783.
14. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T. *Nat Chem* **2011**, *3*, 172–177.
15. Chan, E. C.; New, L. S.; Yap, C. W.; Goh, L. T. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 384–394.
16. Eckers, C.; Laures, A. M.; Giles, K.; Major, H.; Pringle, S. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1255–1263.
17. Bagal, D.; Zhang, H.; Schnier, P. D. *Anal. Chem.* **2008**, *80*, 2408–2418.
18. Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. *Anal. Chem.* **2012**, *84*, 1026–1033.
19. Momoh, P. O.; Hamid, A. M.; Abrash, S. A.; El-Shall, M. S. *J. Chem. Phys.* **2011**, *134*, 204315.
20. Moss, C. L.; Chamot-Rooke, J.; Nicol, E.; Brown, J.; Campuzano, I.; Richardson, K.; Williams, J. P.; Bush, M. F.; Bythell, B.; Paizs, B.; Turecek, F. *J. Phys. Chem. B* **2012**, *116*, 3445–3456.
21. Ridenour, W. B.; Kliman, M.; McLean, J. A.; Caprioli, R. M. *Anal. Chem.* **2010**, *82*, 1881–1889.
22. Fenn, L. S.; McLean, J. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2196–2205.
23. Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *PLoS One* **2010**, *5*, e12080.
24. Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem.* **2007**, *76*, 167–193.
25. Taverner, T.; Hernandez, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinkovic, D.; Devos, D.; Russell, R. B.; Robinson, C. V. *Acc. Chem. Res.* **2008**, *41*, 617–627.
26. Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J. *Nat. Chem.* **2011**, *3*, 126–132.
27. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141–10142.
28. Wyttenbach, T.; von Helden, G.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, *118*, 8355–8364.
29. Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom* **1997**, *32*, 577–592.
30. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658–1661.
31. Barrera, N. P.; Isaacson, S. C.; Zhou, M.; Bavro, V. N.; Welch, A.; Schaedler, T. A.; Seeger, M. A.; Miguel, R. N.; Korkhov, V. M.; van Veen, H. W.; Venter, H.; Walmsley, A. R.; Tate, C. G.; Robinson, C. V. *Nat. Methods* **2009**, *6*, 585–587.
32. Alverdi, V.; Mazon, H.; Versluis, C.; Hemrika, W.; Esposito, G.; van den Heuvel, R.; Scholten, A.; Heck, A. J. *J. Mol. Biol.* **2008**, *375*, 1380–1393.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

33. van Duijn, E.; Barendregt, A.; Synowsky, S.; Versluis, C.; Heck, A. J. *J. Am. Chem. Soc.* **2009**, *131*, 1452–1459.
34. Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. *Chem. Soc. Rev.* **2010**, *39*, 1633–1655.
35. Hopper, J. T.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1851–1858.
36. Schenauer, M. R.; Leary, J. A. *Int. J. Mass Spectrom.* **2009**, *287*, 70–76.
37. Schenauer, M. R.; Meissen, J. K.; Seo, Y.; Ames, J. B.; Leary, J. A. *Anal. Chem.* **2009**, *81*, 10179–10185.
38. Leary, J. A.; Schenauer, M. R.; Stefanescu, R.; Andaya, A.; Ruotolo, B. T.; Robinson, C. V.; Thalassinos, K.; Scrivens, J. H.; Sokabe, M.; Hershey, J. W. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1699–1706.
39. Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C.; Oldham, N. J. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 8291–8294.
40. Benesch, J. L.; Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011**, *21*, 641–649.
41. Hyung, S. J.; Ruotolo, B. T. *Proteomics* **2012**, *12*, 1547–1564.
42. Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. *Expert Rev. Proteomics* **2012**, *9*, 47–58.
43. Marcoux, J.; Robinson, C. V. *Structure* **2013**, *21*, 1541–1550.
44. Sharon, M. *Science* **2013**, *340*, 1059–1060.
45. Bagal, D.; Valliere-Douglass, J. F.; Balland, A.; Schnier, P. D. *Anal. Chem.* **2010**, *82*, 6751–6755.
46. Beck, A.; Sanglier-Cianferani, S.; Van Dorsselaer, A. *Anal. Chem.* **2012**, *84*, 4637–4646.
47. Valliere-Douglass, J. F.; McFee, W. A.; Salas-Solano, O. *Anal. Chem.* **2012**, *84*, 2843–2849.
48. Yamaguchi, Y.; Nishima, W.; Re, S.; Sugita, Y. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2877–2884.
49. Rosati, S.; Rose, R. J.; Thompson, N. J.; van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. *Angew. Chem., Int. Ed. Engl.* **2012**, *51*, 12992–12996.
50. Jones, L. M.; Zhang, H.; Cui, W.; Kumar, S.; Sperry, J. B.; Carroll, J. A.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 835–845.
51. Thompson, N. J.; Rosati, S.; Heck, A. J. *Methods* **2014**, *65*, 11–17.
52. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010**, *82*, 9557–9565.
53. Taraszka, J. A.; Gao, X.; Valentine, S. J.; Sowell, R. A.; Koeniger, S. L.; Miller, D. F.; Kaufman, T. C.; Clemmer, D. E. *J. Proteome Res.* **2005**, *4*, 1238–1247.
54. Taraszka, J. A.; Kurulugama, R.; Sowell, R. A.; Valentine, S. J.; Koeniger, S. L.; Arnold, R. J.; Miller, D. F.; Kaufman, T. C.; Clemmer, D. E. *J. Proteome Res.* **2005**, *4*, 1223–1237.
55. Valentine, S. J.; Ewing, M. A.; Dilger, J. M.; Glover, M. S.; Geromanos, S.; Hughes, C.; Clemmer, D. E. *J. Proteome Res.* **2011**, *10*, 2318–2329.
56. Sundarapandian, S.; May, J. C.; McLean, J. A. *Anal. Chem.* **2010**, *82*, 3247–3254.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

57. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. *Int. J. Mass spectrom.* **2007**, *261*, 1–12.
58. Campuzano, I.; Giles, K. Nanospray Ion Mobility Mass Spectrometry of Selected High Mass Species. In *Nanoproteomics: Methods and Protocols*; Methods in Molecular Biology; Toms, S. A., Weil, R. , Eds.; Humana Press: New York, 2011, Vol. 790, pp 57−70.
59. Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. *Analyst* **201yes1**, *136*, 3534–3541.
60. Lorenzen, K.; Versluis, C.; Van Duijn, E.; Van den Heuvel, R.; Heck, A. J. R. *Int. J. Mass Spectrom.* **2007**, *268*, 198–206.
61. Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chem., Int. Ed. Engl.* **2007**, *46*, 8001–8004.
62. Chernushevich, I. V.; Thomson, B. A. *Anal. Chem.* **2004**, *76*, 1754–1760.
63. Wu, C.; Siems, W. F.; Hill, H. H., Jr. *Anal. Chem.* **2000**, *72*, 396–403.
64. Liu, Y.; Clemmer, D. E. *Anal. Chem.* **1997**, *69*, 2504–2509.
65. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nat. Protoc.* **2008**, *3*, 1139–1152.
66. Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 6247–6251.
67. Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. *Angew. Chem., Int. Ed. Engl.* **2013**, *52*, 4020–4023.
68. Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. *Nat. Methods* **2012**, *9*, 1084–1086.
69. Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. *Anal. Chem.* **2013**, *85*, 11163–11173.
70. McDaniel, E. W.; Mason, E. A. *The Mobility and Diffusion of Ions in Gas*; Wiley Interscience: New York, 1973.
71. Giles, K.; Williams, J. P.; Campuzano, I. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559–1566.
72. Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. *Analyst* **2011**, *136*, 3534–3541.
73. Shvartsburg, A. A.; Smith, R. D. *Anal. Chem.* **2008**, *80*, 9689–9699.
74. Giles, K.; Wildgoose, J. D.; Langridge, I.; Campuzano, I. *Int. J. Mass Spectrom.* **2010**, *298*, 10–16.
75. Clemmer, D. E. Clemmer Group. http://www.indiana.edu/~clemmer/Research/cross%20section%20database/cs%20database.htm (accessed June 2015).
76. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3297–3304.
77. Bush, M. F.; Campuzano, I. D.; Robinson, C. V. *Anal. Chem.* **2012**, *84*, 7124–7130.
78. Salbo, R.; Bush, M. F.; Naver, H.; Campuzano, I.; Robinson, C. V.; Pettersson, I.; Jorgensen, T. J.; Haselmann, K. F. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181–1193.
79. Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889–896.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

80. Bush, M. F.; Campuzano, I. D.; Robinson, C. V. *Anal. Chem.* **2012**, *84*, 7124–7130.
81. Merenbloom, S. I.; Flick, T. G.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 553–562.
82. Makarov, A. *Anal. Chem.* **2000**, *72*, 1156–1162.
83. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430–443.
84. Larriba, C.; Hogan, C. J., Jr. *J. Phys. Chem. A* **2013**, *117*, 3887–3901.
85. Larriba, C.; Hogan, C. J., Jr. *J. Comput. Phys.* **2013**, *251*, 344–363.
86. Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. *Anal. Chem.* **2014**, *86*, 317–320.
87. de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406*, 93–104.
88. Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. *Angew. Chem., Int. Ed. Engl.* **2014**.
89. Benesch, J. L.; Ruotolo, B. T.; Sobott, F.; Wildgoose, J.; Gilbert, A.; Bateman, R.; Robinson, C. V. *Anal. Chem.* **2009**, *81*, 1270–1274.
90. Hewitt, D.; Marklund, E.; Scott, D. J.; Robinson, C. V.; Borysik, A. J. *J. Phys. Chem. B* **2014**, *118*, 8489–8495.
91. Roux, K. H.; Strelets, L.; Michaelsen, T. E. *J. Immunol.* **1997**, *159*, 3372–3382.
92. Niu, S.; Rabuck, J. N.; Ruotolo, B. T. *Curr. Opin. Chem. Biol.* **2013**, *17*, 809–817.
93. Han, L.; Hyung, S. J.; Mayers, J. J.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011**, *133*, 11358–11367.
94. Ruotolo, B. T. PITTCON Conference 2014, March 2−6, Chicago, IL, U.S.A.
95. Zhong, Y.; Han, L.; Ruotolo, B. T. *Angew. Chem., Int. Ed. Engl.* **2014**, *53*, 9209–9212.
96. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134*, 3429–3438.
97. Poppe, L.; Jordan, J. B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P. D. *Anal. Chem.* **2013**, *85*, 9623–9629.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch004

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Chapter 5

Characterization of Monoclonal Antibody Aggregates and Emerging Technologies

Richard L. Remmele, Jr.,[*,1] Jared S. Bee,[1] Jonathan J. Phillips,[2] Wenjun David Mo,[1] Daniel R. Higazi,[2] Jifeng Zhang,[1] Vivian Lindo,[2] and Alistair D. Kippen[2]

[1]Analytical Biotechnology, Biopharmaceutical Development, MedImmune One MedImmune Way, Gaithersburg, Maryland 20878, United States
[2]Analytical Biotechnology, Biopharmaceutical Development, MedImmune Granta Park, Cambridge CB21 6GH, United Kingdom
[*]E-mail: remmeler@medimmune.com

The formation and assembly of soluble monoclonal antibody (mAb) aggregates (nanometer size range) has been a challenging area of research for many decades. It is integral to acquiring a fundamental understanding of antibody structure, mechanism, propagation, and pathways to particle formation. There are many factors at play that can contribute to aggregation, including a variety of stresses, but it is the physico-chemical properties of these multidomain molecules that determine their type and variety of oligomeric forms. As technology has advanced, our ability to elucidate greater detail pertaining to assembly, conformational alteration, and molecular properties has opened the gate to more sophisticated approaches that help to visualize and predict the factors responsible. In this chapter, we describe how a variety of technologies, ranging from the conventional biophysical to an emerging world of promising upcoming applications, can be used to characterize and understand the aggregation phenomena.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Common Causes and Mechanisms of Antibody Aggregation

There are multiple chemical and physical instabilities of proteins (*1*). In general, aggregation of proteins in bulk solution is often described by mechanisms requiring the generation of an aggregation-competent species followed by assembly into aggregates (*2–4*). In the simplest models of aggregation, conformational instability is related to the generation of aggregation-competent species. These aggregation-competent species can then assemble into larger aggregates. Aggregate species ranging in size from soluble oligomers to dispersed nanoparticles to visible precipitates (see the Protein Particulates chapter/Volume 2, Chapter 8 on particle morphologies and detection) may be formed (*3*). This chapter is focused on nanometer size aggregates, but we note that the same basic mechanisms can result in aggregates of all sizes, including visible particles (i.e., solubility limit).

Multiple models and categorizations of aggregation mechanisms have been proposed. Philo and Arakawa (*5*) describe five general mechanisms that can be used to categorize common causes of aggregation:

- Reversible self-association of the native monomer (i.e., interactions of the structurally unaltered active form).
- Aggregation of conformationally altered monomer (i.e., protein unfolding and misfolding).
- Aggregation of chemically modified product (e.g., oxidation, deamidation, disulfide scrambling).
- Surface-induced aggregation (i.e., interfaces and agitation).
- Nucleation-controlled aggregation (i.e., leachables, metal catalyzed).

Figure 1 provides a brief illustration of these most frequently encountered aggregation pathways for antibodies. These mechanisms are not necessarily mutually exclusive; however, an understanding of the rate-limiting steps of aggregation can be used to determine a control strategy. For instance, if the rate-limiting step is the generation of a conformationally altered monomer, then adding excipients, such as sucrose, that are conformational stabilizers can reduce aggregation rates (*5*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

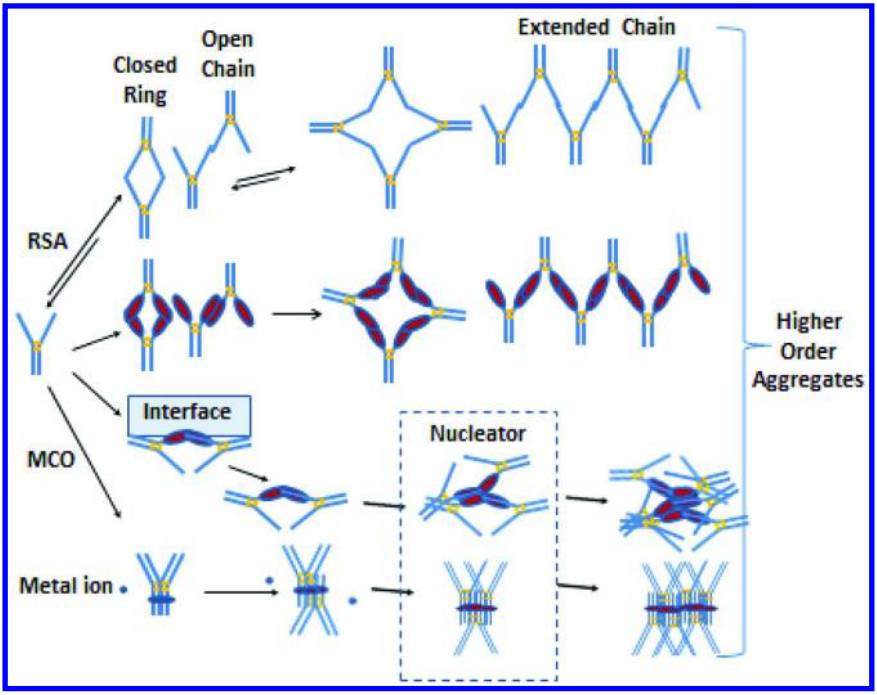

Figure 1. Illustration of possible antibody aggregation pathways. Pathways from top down show reversible self-association (RSA), unfolding-mediated (solid ovals represent unfolded domains) and interfacial-mediated unfolding, and metal ion-catalyzed oxidation (MCO)-mediated aggregation. Unfolded domains can be aggregation competent, whereas aggregates that reach a critical size are nucleators.

Native Protein-Protein Self-Association

Reversible self-association of the native monomer (unaltered conformationally and active) occurs when conformationally stable monoclonal antibody (mAb) molecules form complexes through non-covalent interactions (e.g., electrostatic or/and hydrophobic forces) at a given solution condition. Self-association can be reversible in that, for example, associated species revert back to monomer as the protein concentration is reduced or temperature is increased. Reversible self-association is generally found to result in minimal structural changes in the protein, therefore allowing the recovery of the original conformation upon a change in conditions. However, depending on the nature of the native protein, self-association also may lead to irreversible aggregation. For example, self-association may induce conformational changes in the molecule that cannot be reversed by a change in solution conditions.

In bulk solutions, the direct biophysical consequence of protein-protein self-association at high protein concentrations could manifest itself as phase transition events, for example, opalescence, liquid-liquid phase separation (*6–8*), gel formation, crystallization, or visible particle formation. At the air-liquid

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

interface, the formation of a thick film as a result of protein self-association could significantly affect the rheological behavior of protein solutions (*9*). From a biopharmaceutical development point of view, controlling and minimizing antibody self-association in bulk solutions and at the interfaces are important to ensure stability during shelf life. Even reversible self-association in a relatively short time frame could result in a slow conversion toward irreversible aggregates over an extended period of shelf-life storage or external stress.

Attractive protein-protein interactions result in protein-protein self-association, whereas repulsive protein-protein interactions can minimize it. However, it is difficult to predict protein-protein self-association behavior in both thermodynamic and kinetic terms. First, the exact type of force(s) driving protein self-association, which could be influenced by sequence and fold characteristics, needs to be elucidated. Because of the large size and flexible domain structure of mAb molecules, they can adopt diverse structures with anisotropic orientations. Second, the influence the solution environment and excipients can have a profound effect on protein-protein interactions. Although there are few established experimental approaches described in this book to study protein-protein self-association, it remains to be seen whether they can predict the solution behavior for antibodies at high protein concentrations.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Protein Unfolding and Misfolding

Antibody molecules have complex, multidomain structures. Studies of the unfolding process of antibody molecules specifically at the domain level can potentially facilitate a greater mechanistic understanding of antibody aggregation. When the unfolding step exposes hydrophobic surfaces, it is typically implicated in non-native (conformationally altered) aggregation, which could lead to the formation of subvisible and visible particles, loss of efficacy, and potential safety concerns. This is because a large percentage of hydrophobic residues are generally buried in the interior of the molecule when correctly folded. Exposure of hydrophobic surfaces to the solution is not favorable thermodynamically because of the penalty for ordering the surrounding water molecules (unfavorable entropy). Therefore, these conformationally altered proteins with exposed hydrophobic surfaces will tend to aggregate together to reduce the penalty for the exposed hydrophobic surface.

Elucidation of the detailed unfolding process of mAbs still remains a challenging task due to the structural complexity of mAbs, for example, different subclasses and differences in the complementarity-determining region (CDR) sequence composition. It has been revealed by differential scanning calorimetry (DSC) that the thermal unfolding temperature of a Fab fragment that includes the CDR region could have a melting temperature (T_m) below, overlapping, or exceeding those of the C_H2 and C_H3 domains of the Fc (where the C_H2 domain is generally less thermally stable than the C_H3 domain), depending on the physical properties of the CDR region. DSC measures the heat capacity required to unfold an antibody sample. DSC scans can be useful to determine the temperature at which a given domain unfolds and inform on individual domain unfolding events. A higher unfolding temperature correlates with a greater thermal stability.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Although for other multidomain proteins there is the potential for the least conformationally stable domain to be responsible for initiating aggregation, there is limited characterization on the relationship between the thermal unfolding for a given domain and participation in aggregation. For example, the details of a complex Fc fragment (without Fab attached) and regions specifically involved in aggregation have yet to be elucidated (*10*). For the case where the C_H2 domain of an intact IgG1 mAb was thermally unfolded, aggregation was the consequential outcome (*11*). Inferred from this latter case, whatever region unfolds in the C_H2 domain, it is this region that is responsible for the aggregation that results, yet there is no detailed characterization to prove the point. Another important question to be addressed is the cooperativity of antibody unfolding. Studies with a mouse IgG2 using calorimetric and circular dichroism methods suggested that individual domains unfold independently (*12*), but whether this scenario is a rule rather than an exception remains to be evaluated for different mAbs and quite possibly may depend upon intramolecular domain properties of interactions.

Other aspects when considering the unfolding of mAbs are the sensitivity of each domain to solution environmental conditions such as pH, temperature, and solutes (or excipients), as well as proneness to form aggregates. The correlation of the unfolded state in different solution environments to kinetically form aggregates with temperature can be evaluated using thermal reversibility experiments on the timescale of repeated DSC scans. For example, it was shown that the unfolding of the C_H2 domain of an IgG1 Fc fragment (*Escherichia coli* [*E. coli*] expressed) exhibited high thermal reversibility in a KF salt solution at pH 4.8 (i.e., was less prone to aggregate) in contrast to KCl, KSCN, and Na_2SO_4 solutions (*10*).

Misfolds can occur when more stable but incorrectly paired or twisted disulfides are involved. Aggregation of recombinant misfolded proteins can occur during the cell culture process through hydrophobic interactions or through the formation of incorrectly paired intermolecular disulfide bonds (*13*). The initial glycosylation and folding of mAbs takes place in the endoplasmic reticulum (ER) and Golgi apparatus, and correctly folded proteins are released from the ER. In high-producing cell lines, however, the capacity of chaperones (responsible to correctly fold misfolded protein forms within the cell) may become overloaded with recombinant mAbs. Misfolding of mAb light chains and heavy chains can accumulate as intracellular aggregates and be released to the culture medium through induced dilation of the ER rather than being degraded in the endosome. Ho et al. (*14*) reported that mAb aggregation at the cell culture stage could be minimized by controlling the light chain:heavy chain ratio. Apart from differences in size, misfolded mAbs may lead to differences in properties (e.g., surface hydrophobicity) from correctly folded monomers that can impact the yield and potency of the drug substance.

Chemical Modifications

Post-translational modifications of a mAb can occur during manufacturing and storage (*15*). Common modifications include, but are not limited to, disulfide scrambling, oxidation, fragmentation, deamidation, isomerization, and glycation. The molecular properties of mAbs will be impacted by these modifications;

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

the altered properties in turn either have been reportedly linked to, or may potentially contribute to, mAb aggregate formation. The mechanism of each of the modifications and its potential correlation to aggregation will be reviewed.

Disulfide Scrambling

Disulfide-bonded species tend to be the dominant covalently linked reducible aggregates (*16*). The formation of reducible intermolecular disulfide bonds may derive from process parameters, and the unique structural traits of IgG isotypes. Buchanan et al. (*17*) reported that unpaired cysteines (Cys) in the variable region of an IgG2 induced rapid aggregation. Later, by engineering out the unpaired Cys and switching the isotype from IgG2 to IgG1, the aggregation rate was significantly reduced. The non-reducible aggregates involved with the cross-linking of Cys residues may be related to the formation of thioester bonds (*15*), as shown by Tous et al. (*18*), who carried out detailed studies of a non-reducible humanized monoclonal IgG1 species in an accelerated stability setting using reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), reduced capillary gel electrophoresis (rCGE), and liquid chromatography-mass spectrometry (LC-MS) to elucidate the mechanism. Characterization of the non-reducible entity revealed it was a cross-linked adduct of a thioether bond between Cys223 of the heavy chain and the c-terminus cysteine residue of the light chain.

Manufacturing-induced disulfide reduction has been observed for IgG antibodies. For example, Hutterer et al. (*19*) concluded that susceptibility to reduced disulfides depended on both antibody class and light chain type. For example, the trend in dithiothreitol reduction follow the order IgG1λ > IgG1κ > IgG2λ > IgG2κ. On the process side, the formation of scrambled disulfides may originate from the reduction of the interchain disulfide by certain host cell proteins, such as thioredoxin and thioredoxin reductase (*20–22*), or certain process parameters, such as temperature (*23*).

Oxidation

Methionine (Met) and Tryptophan (Trp) are the amino acid residues most susceptible to oxidation. For Met oxidation, the residue is oxidized to form sulfoxide, increasing the mass by 16 Da. This makes the side chain of Met more polar (*24*, *25*). In the case of Trp oxidation, the intermediate species and final oxidation product are more complex and generally follow the pathways that are initialized with the formation of hydroxy-tryptophan or N-formylkynerenine (*26–29*).

Tyrosine amino acids also can form oxidative adducts. In a molecular pathology study of several globular proteins, light stress-induced oxidation resulted in both intra- and intermolecular dityrosine covalent bond formation, which compromised the proteins' structural stabilities and activities (*30*). In a biopharmaceutically relevant case, light stress-related oxidation resulted in

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

covalent dimerization of insulin due to dityrosine formation with a concomitant loss in structure and activity (*31*).

Oxidation to form carbonyl groups on the protein (carbonylation) also can occur but is more challenging to characterize because of its variable occurrence across a range of amino acid residues in the protein (*32*). Because protein carbonylation of antibodies is a relatively new emerging area of investigation, there is little published on its influence on aggregation propensity. However, the consensus is that an elevated level of various forms of oxidation of a recombinant protein may increase the susceptibility to aggregation by destabilizing the protein structure (*33*). In one example, the results implied a correlation between Trp oxidation and the aggregation of a mAb exposed to stainless steel and its leached ions (*34*). Metal-catalyzed oxidation (MCO) of proteins in vivo is a phenomenon that has been linked to cellular aging processes (*35*). Similarly, MCO has been implicated as a cause of oxidation and subsequent aggregation of biopharmaceuticals (*36*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Fragmentation

Fragmentation of IgGs includes the previously described Cys reduction mechanism that resulted in the loss of light chain(s) (*20–22*); non-enzymatic hydrolysis at the location of Xaa-Asp, Xaa-Ser, Asn-Xaa, or Cys-Cys (*37*); fragmentation induced by proteases that are derived from host cell proteins (*38*); and hinge region cleavage that is catalyzed by Cu or Fe ions in solution (*39*). It is conceivable that IgG fragments may have different physico-chemical properties and can potentially serve as nucleators of aggregation (*40*).

Deamidation, Isomerization, and Glycation

Unlike previously described chemical modifications, there have been no reports correlating deamidation (*41–43*), isomerization (*44*, *45*), and glycation (*15*, *46*) to mAb aggregation. Nevertheless, these modifications will alter the physical and chemical properties of mAbs and may destabilize the native conformation, possibly leading to aggregation under certain circumstances.

Interfaces and Agitation

Proteins can adsorb to both solid-liquid and liquid-air interfaces and, in some cases, adsorption has caused aggregation of biopharmaceuticals (*47*). There are multiple interface exposures under a variety of solution conditions during antibody production, filling, transportation, storage, and delivery. Contact with manufacturing equipment (e.g., pumps, lines, containers) could potentially cause aggregation of antibodies (*48*, *49*). For instance, contact with stainless steel was shown to cause aggregation of several antibodies in forced degradation stress studies (*16*, *50*, *51*).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

In manufacturing operations, accumulation of solid materials such as aggregates or precipitates at interfaces can cause "fouling." Fouling of membranes due to aggregation can be a challenge during filtration or diafiltration operations because the membrane can become physically clogged and complicate production of drug substance (*52*).

Protein adsorption to interfaces is a highly complex phenomenon that has been the subject of many studies using a wide variety of methods (*53*, *54*). The properties of the protein, the interface, and the solution conditions are some factors that can impact the extent and rate of adsorption and also the degree of structural perturbation that occurs (*54–56*). Multiple theoretical models of varying complexity have been developed to describe the observed kinetics of adsorption and desorption and the final saturated coverage of proteins at interfaces (*56*). Recent use of single molecule tracking experiments show promise as a way to overcome some of the challenges in the interpretation of adsorption phenomena related to the assumptions in using macroscopic observations for modeling the behavior of proteins at interfaces (*57*). Antibodies can adsorb to solid surfaces in different orientations depending upon the surface and protein properties (*58*). Multidomain proteins such as antibodies may bind to surfaces with unfavorable net electrostatics by orienting a patch of opposite charge to the surface and also by structural rearrangement after adsorption (*55*, *56*). Displacement of surface-adsorbed water, rearrangement of ions around the surface and protein, and structural rearrangements of the protein are key drivers of adsorption (*47*, *55*). The saturated adsorbed amount for a typical monolayer coverage can range from 0.5 to 4 mg/m^2 (*58*). A rough calculation for antibody drug formulations in the range 5 to 100 mg/mL suggests that a negligible adsorption of 0.01 to 0.1% of the total protein could occur in a container with a surface area of about 10 cm^2/mL. In these situations, the potential for aggregation and particle formation as a consequence of adsorption can be a more important consideration than simple adsorptive losses of product. For example, adsorption followed by desorption has been shown to lead to structural changes in certain cases (*59*). Structurally altered protein molecules could potentially act as nuclei for further aggregation on the surface or in the bulk solution after desorption or homomolecular (self) exchange (*54*, *60*). In a study of adsorption to hydrophilic silica, the Fc domain of an antibody was shown to undergo more rapid adsorption with a greater structural change than the antigen-binding fragment domain (*61*). The increased adsorption of a mAb to silica near its isoelectric point, studied using total internal reflectance fluorescence and neutron reflectometry, was consistent with a non-uniform triple layer containing oligomers (*62*).

Antibodies are large proteins that are considered "soft," meaning that they are easily altered in structure upon adsorption (*55*). Adsorption to hydrophobic surfaces has been reported to induce the greatest structural rearrangements of adsorbed proteins (*55*, *56*). The adsorption of an antibody to Teflon® surfaces was shown to result in large changes in secondary structure compared with the native form (*12*). In one study, freezing of an antibody in Teflon® containers was reported to result in high levels of aggregation when compared with polypropylene or glass containers (*63*). Atomic force microscopy (AFM) measurements of an adsorbed mAb found a lower adhesion force to hydrophobic surfaces than to

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

hydrophilic surfaces (*62*). In the same study, however, loss of secondary structure was measured by circular dichroism (CD) for the mAb when adsorbed to both hydrophilic and hydrophobic surfaces (*62*).

Silicone oil is a relevant hydrophobic liquid contact interface because it often is used to lubricate the barrel of glass prefilled syringes (PFS). The exposure of Abatacept (an Fc-Fusion protein) to siliconized PFS resulted in the formation of silicone droplets entangled in a fibrous structure of aggregated protein (*64*). The fact that aggregates and silicone oil droplets were entangled suggests that the silicone oil droplets were the most likely cause of the observed aggregation. In another case, the presence of salt was shown to reduce the unfolding of an antibody adsorbed to silicone oil, resulting in a reduction in its propensity for aggregation during agitation (*65*).

The air-water interface, which is ubiquitous for liquid dose formulations of antibodies, is one of the most commonly cited causes of protein aggregation. Because the air at the air-water interface has a dielectric constant of ~1, it is a hydrophobic environment and is capable of inducing unfolding during agitation. Agitation-induced aggregation of liquid antibody formulations can occur during processing, filling operations, and transportation. Exposure to the air-water interface during agitation has been shown to induce aggregation of antibodies (*66*, *67*). For example, pressure changes and vibrations during air and ground transportation can lead to sloshing of liquid or generation and movement of bubbles in the container closure, resulting in interfacial stress-induced aggregation. When proteins adsorb at the air-water interface, they can form an aggregated film that may be susceptible to mechanical disruption with subsequent release of aggregated protein into the bulk (*68–70*). Adsorption to the air-water interface, enhanced by the turnover due to shear, is generally considered to be the major cause of aggregation during agitation of antibody formulations rather than direct impact of bulk shear alone (*48*, *71*, *72*).

Formulating with surfactant is a common strategy employed to protect proteins from aggregation at interfaces, including solid surfaces, the air-water interface, and ice-water interfaces (*47*, *72–74*). The two main mechanisms commonly proposed for stabilization of proteins by surfactants are that the surfactant can either compete with the protein for adsorption at the interface or that the surfactant can "decorate" the protein and thereby cover its aggregation-prone surfaces (*73*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Leachables

Leachables have been implicated in various degradation pathways of antibodies (*36*). Metal ions, particularly iron ions (Fe^{2+} and Fe^{3+} ions) leached from stainless steel equipment, can cause oxidation, aggregation, and fragmentation (*75*). In one example, Fe ions leached from stainless steel tanks at low pH in sodium chloride formulations resulted in the oxidation of recombinant humanized mAb HER2 (*76*). In another case, aggregation of an anti-IL8 antibody was most likely caused by MCO due to Fe ions leached from stainless steel tanks during freeze-thaw cycles for a formulation with high levels of histidine hydrochloride (*77*). There also are many examples where metal ions have caused

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

fragmentation of antibodies (*78*, *79*). The relative leaching rates of Fe ions from stainless steel under various conditions have been reported (*80*, *81*).

Leachables from the final container closure, delivery device, or administration supplies also can cause aggregation of biopharmaceuticals. Tungsten residues left over from the PFS-forming process were implicated in aggregation of a protein (*82*). High levels of soluble polyanionic forms of tungsten have been shown to cause coagulation of soluble monomers, resulting in particle formation and precipitation for several proteins, including an antibody, Fc-fusion protein, and α-helical protein (*83*, *84*). These reports have prompted improvements in manufacturing of next-generation PFS with greatly reduced tungsten levels (*82–84*). Although not necessarily specific for antibodies, there have been cases reported where leachables from contact materials such as rubber stoppers, glass vials, or intravenous administration components have impacted the stability of biopharmaceuticals (*47*). Forced degradation, accelerated, and long-term stability studies, as well as in-use stability studies, can inform on the potential for contact surfaces to impact aggregation of antibodies.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Kinetics and Formulation Section

Aggregation can involve the association of either native, conformational, and/or chemically altered protein monomers and/or the generation of a critical aggregate nucleus (Figure 1). Aggregation can be rate-limited by protein conformational changes, the assembly of aggregates, or nucleation that is concentration-dependent relative to monomer loss, provided that there are no other competing pathways (e.g., fragmentation) and aggregation is the predominant degradation (*85*). For instance, if a conformational change in the antibody is the rate-limiting step in the aggregation process, then the kinetics could be first order in protein concentration. In contrast, if the assembly of two molecules is rate-limiting, then the kinetics of aggregation could be second order in protein concentration. Nucleation-controlled aggregation can be more complex, with a lag phase in time that occurs while the assembly of a critical nucleus or cluster (nucleator) of aggregated protein is formed followed by a sudden rapid growth in aggregation (Figure 1). In some more complex cases, there may be multiple competing aggregation pathways occurring simultaneously, and a change in the dominant pathway can occur with changes in protein concentration, solution conditions, and temperature. These complex factors together make determination of the fundamental causes and kinetic behavior of soluble aggregation quite challenging.

The different aggregation pathways each can have different kinetics and temperature dependencies (*85*, *86*). The temperature dependence of the aggregation of proteins can be complex because of the fact that both conformational stability and the assembly rate of aggregates are both impacted by temperature to different extents. Competing mechanisms for aggregation can lead to a change in the dominant mechanism with temperature, resulting in non-Arrhenius behavior (*85*). At lower temperatures, for example, the partial unfolding of proteins may be rate-limiting, whereas at higher temperatures,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

their assembly may become rate-limiting. In one example, the aggregation rate increased due to a lower activation energy threshold, deviating from linearity near the melting transition of one therapeutic protein (*87*). A simulation model was generated that could predict the aggregation rates based on the temperature, concentration, and estimated thermodynamic and kinetic parameters of the aggregation process (*87*).

Either conformational stability or assembly processes can be rate-limiting for the aggregation of proteins (*3*). Noted above, temperature can have a strong impact on both these factors. The proportion of structurally perturbed and partially unfolded protein molecules increases as the temperature approaches the melting temperature of a protein domain. The processes involved in the assembly of aggregates from monomers also usually increase with temperature. Conformational stability of proteins (defined by the free energy change) is temperature-dependent, often following a parabolic profile with the maximum temperature centered at about 21 °C for several types of proteins (*88*). The temperature dependence of the free energy of protein unfolding is a result of a change in the heat capacity (*89*). This change in heat capacity can lead to deviations in the expected aggregation kinetics (i.e., non-Arrhenius behavior) when extrapolating from one temperature to another. Upward curvature in the Arrhenius plots for aggregation of antibodies can lead to incorrect overestimation of the stability at lower temperatures (*85*). This type of behavior can be a challenge when performing an extrapolation from accelerated and stressed temperature storage conditions to the stability of refrigerated storage (*85*). In one study of five different antibodies, the observed non-Arrhenius aggregation kinetics could be best fitted under certain conditions using two parameters, an activation energy and a reference temperature (*90*). Scanning or ramping the temperature of samples instead of conducting isothermal kinetic studies is a recent development with the potential for saving time and material (*91*).

High temperature is not the only concern for antibody aggregation. Freezing and thawing also can apply multiple stresses to antibodies, including cold denaturation, freeze concentration, excipient crystallization, and exposure to ice-water interfaces (*92*). Freezing can increase the concentration of both the protein and excipients as ice forms. Freeze concentration can change the solution environment around the protein, result in the crystallization of excipients, and permit protein concentration-dependent aggregation steps (*92*). Crystallization of buffer species such as sodium phosphate at subfreezing conditions can result in large shifts in the pH of the liquid phase that can destabilize the protein and accelerate aggregation (*93*).

Formulation strategies generally attempt to stabilize a folded state of the antibody while minimizing the potential for aggregation. In addition, the role of processing steps such as freeze-thaw or lyophilization should be considered with regard to role of the excipient and stresses involved. Disaccharide or polyol excipients often are used because they can stabilized proteins in solution and during dehydration (e.g., lyophilization). In some cases, the mechanism of stabilization of proteins by polyol excipients in solution is consistent with preferential exclusion (*94*). Preferential exclusion refers to the situation where excipients are “excluded” from the protein surface and are therefore at a lower

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

concentration near the protein surface than in the bulk solution. This causes a preferential hydration of the protein surface. The result of this is a perturbation of the native protein by an increased chemical potential of the excluded co-solute molecules (e.g., disaccharide). Exclusion of the disaccharide from the protein surface increases the chemical potential, but the chemical potential of the unfolded state is more unfavorable than that of the unfolded state in the absence of excluded co-solute molecule. Because the unfolded state is thermodynamically more unfavorable, this results in a greater conformational stability of the native protein with respect to unfolding or partial unfolding. Disaccharides are excluded from protein surfaces due to their greater size compared with water molecules. Other excipients also can be excluded from the protein surface and will exert a similar effect.

Another formulation advantage of disaccharides is that they also can form a viscous glassy state that immobilizes the antibody in a native-like state upon freezing, mitigating the impact of freeze concentration and unfolding propensities that lead to aggregation (*92*). Successful strategies to stabilize protein in an amorphous phase have focused on excipients that contribute to glassy properties with high glass transition temperatures (Tg) in lyophilized products or storage of the frozen material at a subzero temperature below the Tg' to avoid viscous flow and potential crystallization of stabilizing excipients. The glass transition temperature is the temperature below which a glassy phase of the excipient is stable and will not crystallize. The glassy state immobilizes the antibody, keeps all stabilizing excipients in the amorphous state with the antibody, and restricts molecular motion that could lead to aggregates.

Storage at a temperature below the glass transition temperature is important to maintaining the stabilizing environment surrounding the protein in the glassy state. Sometimes, the phase behavior can result in a loss of the stabilizing environment in the glassy state. For example, crystallization of excipients can result in a loss of their cryoprotective impact on protein stability. Crystallization results in phase separation from the amorphous phase removing stabilizing excipients from the phase where the protein resides. For example, aggregation of an Fc-fusion protein was linked to the crystallization of sorbitol in frozen formulations where the storage temperature was above the Tg' (*95*). Similarly, crystallization of trehalose from frozen formulations of an antibody stored above the glass transition temperature resulted in an increase in aggregation (*60*). In both cases, a stabilizing agent was removed from the phase of the protein.

Interfacial unfolding of the protein by adsorption to the ice surface during freezing can lead to pathways of aggregation and often has been addressed with the addition of surfactants (*96–98*). The type and size of the container used for freezing also can be important. For instance, more aggregation occurred for an antibody that was frozen and thawed in Teflon® than in polypropylene containers (*63*). Thus, the surface properties of the container itself can play a role and need to be taken into consideration for frozen storage.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Physico-Chemical Properties of Antibodies Affecting Aggregation Propensity

Electrostatic Properties and the Isoelectric Point (pI)

In the previous sections, certain biochemical and biophysical properties of antibodies have been noted to be involved with the propensity toward aggregation. Electrostatic interactions play an important role in protein-protein self-association. The net charge for a mAb is the sum of the total number of surface-exposed charged side chains, which is determined by the pH of the solution and the pKa of the protein side chains. The pH value where a protein carries an equal number of positive and negative charges is designated as the isoelectric point or "pI." The pI is a fundamental biophysical property for a protein and can be a reference to delineate the net charge change as a function of pH. Specifically, an antibody will carry a net positive charge at pH conditions below the pI, which is a common situation for antibody formulations. The net charge on proteins can impact solubility and association behavior. At low salt concentrations, proteins are generally more soluble at pH conditions away from their pI because of repulsive electrostatic interactions.

Another biophysical property to consider during the protein self-association process is the charge density, defined as the ratio of the net charge to the surface area. At high charge density, the electrostatic repulsion globally dominates protein-protein interactions, and the antibody molecules are less likely to self-associate. At conditions of low charge density and net charge neutral, the antibody molecules tend to self-associate. Third, it is important to consider the chemistry of the charged side chains based on the Law of Matching Water Affinities (*99*). Specifically, the positive charges are from weakly hydrated Arg and Lys side chains while the negative charges are from strongly hydrated side chains of Asp and Glu. The hydration properties of these charged side chains dictate how the electrostatic interactions are specifically modulated by the ionic solutes.

Salts are powerful ionic solutes for modulating electrostatic interactions between antibody molecules. Currently, the precise mechanisms for how salt ions interact with proteins remains an active area of research. Empirically, the effect of salt on protein solubility was defined by the Hofmeister series. Now more experimental evidence in the literature suggests that salt ions affect protein-protein interactions through specific interactions with surface-exposed side chains and backbone residues on the protein surface (*100*). For a highly charged protein where the pH is far away from its pI, it is generally acknowledged that the initial addition of salts tends to reduce the electrostatic repulsion and leads to protein self-association (*101*). Weakly hydrated anions such as Cl^- tend to interact more strongly with weakly hydrated, positively charged side chains on the protein surface and therefore more effectively minimize the electrostatic repulsions between two positively charged protein molecules than strongly hydrated anions such as acetate. On the other hand, salt addition typically makes protein-protein interactions less electrostatically attractive and, therefore, minimizes protein self-association when a protein is net charge neutral.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Surface Hydrophobicity

Besides the charged side chains, there are exposed hydrophobic side chains on the antibody surface. The hydrophobic feature of an antibody surface not only affects the tendency for the antibody self-association but also dictates how the protein responds to the solution condition changes (e.g., salt addition). For example, it has been shown that a strongly hydrophobic and positively charged surface could demonstrate charge inversion (i.e., from positive to negative charge) (*102*) as the concentration of chaotropic anions (e.g., Cl^-) is increased. This behavior suggests that the addition of Cl^- would initially decrease the electrostatic repulsive interactions between protein molecules and then begin to increase such interactions with additional anion concentration.

Surface hydrophobicity properties of the protein molecule can impart sites of protein-protein binding, depending on the charge and surface area of exposure (*103*). The hydrophobic force of attraction will depend on electrostatics (coulombic attraction or repulsion) and solution entropic properties (*104*). A hydrophobic surface in water can have a large surface energy. For example, CH_2 groups in water have reported energies of approximately 50 mJ/m^2 (*105*). This means that such surfaces in aqueous solutions can be quite "sticky" and can be potential sites where protein molecules can aggregate (*106*). Currently, although postulated (*107*), there is not a good elucidation of a hydrophobic force in such solutions, but if there were, it would be possible to evaluate such a force with respect to forces that repel (e.g., coulombic forces). This might enable a way to predict when hydrophobic patches found on native folded proteins might be sites of hydrophobic interactions. In light of this knowledge, there has been reported efforts to identify aggregation-prone regions based on the dynamic exposure of spatially adjacent hydrophobic amino acids using in silico algorithms (*108*, *109*).

From adsorption studies of mouse anti-human IgG onto Teflon® (a hydrophobic surface), it is possible to gain some insights regarding the differences between Fab and Fc domains pertaining to surface hydrophobic properties. For example, it was shown that adsorption leads to the denaturation of the Fab but not the Fc (*110*). The reason cited was that the Fab hydrophobicity was greater than the Fc, and thus, the Fab adsorbed more strongly to the Teflon® than the Fc. Furthermore the Fab was conformationally altered in the adsorption process. Hence, the properties of the protein solvent-accessible surface hydrophobicity needs to be factored in when considering the mechanism of protein aggregation in context with cause and effect (e.g., mechanisms and propensity of surface adsorption or solution environment protein-protein aggregation).

Glycosylation

Glycosylation of antibodies has the potential to influence aggregation propensity. Glycosylation of the Fc region of antibodies is heterogeneous, with G0F, G1F, and G2F being the predominant glycoforms corresponding to the number of terminal galactose units. For more detailed information on glycosylation, see the Glycosylation chapter/Volume 2, Chapter 4.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Oligosaccharides also can be located within the Fab region of antibody structures. In Kabat's database, ~18% of the variable region sequences contain a potential N-linked glycosylation site in heavy chains. Approximately 15 to 25% of the Fab fragments and 15% of the light chains isolated from human myeloma proteins were found to contain N-linked oligosaccharides (*111*).

The correlation between mAb glycosylation (Fc and Fab) and aggregation is still being elucidated, and a detailed understanding of this relationship is important for understanding the stability risks involved in drug development, drug storage, and the development of stable drug formulations (*112*). There has been reported evidence that glycosylated IgG1 Fc affords greater thermal stability than the aglycosylated form (*113*).

Enzymatic deglycosylation has been shown to cause conformational destabilization and lead to increased aggregation in humanized IgG1 mAb Fcs. Aglycosylated mAbs were less stable and therefore aggregated more readily than the glycosylated forms (*114*). Arnold et al. were able to show that deglycosylation of the Asn-linked glycans (located at the C_H2 domain dimer interface) in a murine mAb caused structural perturbations. The two C_H2 domains moved with respect to each other during molecular dynamic simulations, thereby perturbing the quaternary structure of the deglycosylated mAb (*115*, *116*).

Using spatial aggregation propensity (SAP) modelling, which attempts to determine exposed surface patches at high risk of being associated with aggregation (see also the "In Silico Aggregation Prediction" section of this chapter), Kayser et al. were able to demonstrate that the Fc glycan shields a region of hydrophobic residues that have a high propensity to aggregate once the molecule is deglycoyslated. Under stress conditions, deglycosylated antibodies were more prone to aggregation than native glycosylated molecules (*112*). Conversely, hyperglycosylation has been shown to improve properties such as molecular stability, solubility, and in vivo biological activity and seemed to reduce immunogenicity for some therapeutic proteins (*117*). It also has been reported that the type of glycosylation present may play an important role in the likelihood of protein aggregation (*118*). Ho et al. (*14*) were able to generate mAbs with different levels and types of Fc glycosylation by varying expressed levels of heavy and light chains. Antibodies could be generated containing sialylated bi- and tri-antennary structure, as well as significantly higher levels of mannose sugars. These mAbs with altered Fc glycans were less stable and aggregated more readily than typical non-altered glycosylated mAbs and showed lower conformational stability. Although the findings may point to altered glycans destabilizing antibody structure, the presence of these altered Fc glycoforms also may indicate the formation of the antibody by an aberrant folding pathway due to overwhelming the folding and post-translational modification machinery of the cells in the case of high mannose sugars. The presence of sialylated bi- and tri-antennary structures may point to structural aberrations causing increased exposure of the glycosylation site, allowing for a greater level of glyco-modifications within the Golgi apparatus and resulting in the generation of more aggregation-prone molecules.

More representative data generated by Schaefer and Pluckthun (*119*) were obtained from IgG molecules expressed in *Pichia pastoris* that were found to

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

possess mannose-rich Fc glycans. Although showing lower thermal stability, these molecules showed superior aggregation resistance when compared with the same molecule expressed in HEK293, containing normal expected Fc glycans. The physical stability of an antibody also was increased, making it less aggregation prone by reintroducing an N-linked carbohydrate moiety within a CDR sequence shielding an aggregation "hot-spot" (*119*). This suggests that Fc or CDR glycosylation can increase antibody stability, although the possibility of conformational changes due to IgG production in different expression systems cannot be excluded.

In general, glyco-engineering is a strategy that can be used to modulate certain properties of an antibody, including solubility, physical stability, and aggregation propensity. For typical antibody Fc glycosylation (G0, G1F, and G2F structures) and in the absence of aberrant protein folding mechanisms, glycosylation seems to enhance solubility and physical stability of antibody therapeutics.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Antibody Isotype and Fv and Fc Domain Structure

Although formulation has some impact (*120*), clearly a vital consideration is the primary sequence of the protein itself. Although the modular nature of mAbs is such that the Fc region (and majority of the molecule) is largely identical between antibodies of a particular isotype, the Fab region differs greatly. Consequently, there are significant differences in stability and aggregation propensity among antibodies related to Fab differences and the particular antigen specificity of the mAb (*121*). These effects may be highly localized within the variable domains or potentially involve longer distance domain-domain interactions (*122*).

Although broadly similar in structure, the IgG isotypes (IgG1, 2, 3, and 4) differ in structural detail (*123*). Most therapeutic mAbs are either based upon IgG1 or IgG2 and only a handful of IgG4 molecules in the clinic (*124*). Furthermore, all seven currently marketed Fc-fusion proteins contain IgG1 Fc, with a smaller number of IgG2/4 Fc fusions in clinical development (*125*). Research into the comparison of IgG isotypes in terms of stability and the propensity to aggregate have yielded variable results. For example, a study examining the stability of an anti-streptavidin mAb as IgG1 and IgG2 isotypes showed that the IgG2 variant was more prone to aggregation (*126*). IgG2 molecules contain two additional disulfide bonds in the hinge region compared to IgG1 that are capable of forming covalent intermolecular bonds between molecules. By contrast, another group investigated an anti-LINGO-1 (LRR and Ig domain-containing, Nogo receptor-interacting protein) antibody using a range of both wild-type and mutated IgG1 and IgG2 variants (*62*, *127*). This study showed that there were both isotype-specific and Fab-specific factors contributing to aggregation propensity and that the wild-type IgG1 was less soluble than the wild-type IgG2. This work also demonstrated that the different isotypes exhibited different aggregation mechanisms, both reversible and irreversible. Such studies indicate that isotype switching can be a useful tool to improve stability and reduce aggregation propensity but, given the complex nature of IgG molecules, it is difficult to predict beforehand which will be most successful. For instance, IgG4-based molecules have been shown to exhibit Fab arm switching, which

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

may affect their in vivo properties (*128*). A comprehensive analysis of Fab arm switching across the IgG isotypes was conducted by Rispens and co-workers with the insight that a defined set of amino acids at the C_H3-C_H3 interface modulate inter-heavy chain interaction (K_d) by six orders of magnitude (*129*). Thus, ordinarily solvent-inaccessible regions in the Fc may be exposed more or less in different isotypes. This would potentially lead to isotype-specific aggregation if the isotypes were prone to self-association via the variable domains.

Mutations in constant domains have been reported that increase stability or reduce aggregation (*130*). More often, however, the variation in the biophysical properties of human therapeutic antibodies relates to their variable domains. This can present problems in the development of IgG monoclonals as it often cannot easily be overcome by isotype switching and formulation approaches. This is probably not surprising as variable domains (as their name indicates) encode most of the variation in the antibody sequence. As a result, numerous studies aiming to improve the biophysical properties of antibodies through variable domain engineering have been described. One powerful example of variable domain engineering to eliminate aggregation comes from Tessier and co-workers (*131*). Mutational and predictive algorithm approaches built a case for CDR1 being responsible for driving aggregation. Thus, CDR1 was grafted from an anti-hen egg lysozyme construct (Hel4) V_H domain into a domain antibody (dAb) construct, Dp47d (aggregation-prone wild-type). This resulted in the mitigation of aggregation and poor solubility. "Nanocharging" the CDRs with charged (typically acidic) residues also has met with some success in isolated variable domains (*132*). Christ and co-workers were able to engineer CDR1 of the heavy chain and CDR2 of the light chain in the anti-human epidermal growth factor receptor 2 (anti-HER2) antibody Trastuzumab. This was done initially in isolated variable domains and then combined into a single chain variable fragment (scFv) format with limited data on IgG1 mutants. With the relatively low stability of scFvs and without the presence of constant domains to complicate interpretation, the impact of the variable domain engineering on aggregation and thermal unfolding was readily determined. Some non-paratope mutants were successful in improving these biophysical properties while retaining binding affinity to HER2.

Another approach to improving solubility where there is an aggregation-prone site in a variable domain is the introduction of N-linked glycosylation (*127*, *133*). This is a complex strategy that is covered in the "Glycosylation" section of this chapter.

Conventional Approaches To Elucidate and Characterize Antibody Aggregates

Process Characterization of Antibody Aggregation

Strategy To Characterize Soluble Antibody Aggregates

Aggregates of a therapeutic protein are generated during its bioprocessing, which include cell culture, purification, formulation, filling, and finishing (*16*). This section is focused on soluble aggregates; we note that the same mechanisms

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

can result in aggregates of all sizes, however, including visible particles. Protein particle formation and detection is discussed in more detail in the Protein Particulates chapter/Volume 2, Chapter 8. For protein therapeutics, the presence of aggregates of any type typically is considered to be undesirable because of the concern that the aggregates may lead to immunogenic reactions or may cause adverse events on administration (*134*).

From the perspective of patients and regulatory agencies, the foremost concern is about the level and properties of the aggregates in a drug substance. Unlike the plethora of literature describing the mechanism of aggregate formation and mechanisms in each of the processes, there is a lack of data on the detailed characterization of aggregates of final protein therapeutics. In the following sections, the characterization of aggregates of a therapeutic IgG1 at the BLA (Biologics License Application) or equivalent MAA (Marketing Authorization Application) level, is described. The analytical strategy involved isolation of soluble aggregates from the drug substance using size exclusion chromatography (SEC) fractionation that were then tested using various biochemical and biophysical methods.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Aggregate Control during Production

Aggregates are one of the most common product-related variants in recombinant antibody production. Quantification and characterization of aggregates during production and purification is critical to the control and understanding of aggregate formation. This can affect the choice of appropriate manufacturing cell lines and fermenter conditions, viral removal steps (such as low pH, heat, UV), antibody capture and purification columns, and appropriate separation (removal) of aggregate structures during the process.

Typically, aggregation is monitored and quantified under native (non-denaturing) conditions using SEC with analytical ultracentrifugation (AUC) as a confirmatory orthogonal method. SEC is predominantly used due to the need to have a routine, accurate, sensitive, and validated method for soluble aggregates. Because SEC is often the method of choice for lot release testing to quantify soluble aggregates, its application to process monitoring and control can help to avoid process comparability issues. There are a limited number of emerging bioprocess analytics for more "online" measurement of soluble aggregates involving light scattering and dye binding. However, these tend to be relatively insensitive (the difference of a few percent aggregate can be significant) and thus are not yet in routine use for aggregate evaluation. A routine device (for process analytical testing) attached to the purification/chromatography instruments that enables direct detection of aggregate measurement online (in-process) remains under development.

Beyond protein engineering, where sequences can be screened for aggregation propensity, the key areas where aggregation can be controlled are during expression, purification, formulation, and storage. During the course of cell expression development, the most typical type of aggregation occurs through incorrect protein folding. This type of contaminant can be minimized through screening of cell lines and subsequent fermentation conditions, including media

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

development. Post-expression, aggregated protein is removed during purification with the use of a number of chromatography steps that take advantage of size/charge differences between multimeric material and monomer.

Conventional Assays To Characterize Aggregates

This section describes more routine assessments of aggregation by traditional methods, using general elucidation of the NIST reference antibody as examples. The analytical strategy (for all assays other than AUC) involved isolation of soluble aggregates from the drug substance using SEC fractionation that were then tested using various biochemical and biophysical methods. Fraction collection and evaluation of concentrated dimer species allows direct evaluation of the aggregate in question without being masked by other aggregate fractions.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Capillary Electrophoresis (CE) and Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is a robust and common method that is routinely implemented to obtain the information of approximate molecular weight and quantity. However, the presence of SDS means that non-covalent aggregates are disrupted, so the method only detects covalent aggregates. If operated under reducing conditions, SDS-PAGE can differentiate the aggregates formed between the reducible disulfide bonds and those held together by the non-reducible covalent bonds (*18*).

SDS-PAGE in recent years has been eclipsed by the increasingly more popular CE counterpart in industrial settings, as the latter is better suited for the needs of quantitative robustness, automation, and high throughput. However, the capability of detecting aggregates is limited by the specifications of commercial kits of the molecular weight range, which is normally less than 250 kDa and thus lower than the molecular weight of dimeric IgGs.

Size Exclusion Chromatography-Multi-Angle Laser Light Scattering (SEC-MALLS)

SEC is widely used for molecular weight estimations of proteins in their native state (*135*). SEC-MALLS has found routine applications for mAbs in studies on purity, protein-protein interactions, and aggregation.

For all subsequent analytical evaluations (data presented below for SEC-MALLS, AUC, CD, Fourier transform infrared spectroscopy [FTIR], and fluorescence) of isolated soluble aggregate samples of the NISTmAb, aggregate fractions were collected using an Agilent 1100 high-performance liquid chromatography (HPLC) system equipped with two tandem SEC columns (G3000SWXL column, 7.8 × 300 mm). The NISTmAb at 100 mg/mL was injected and eluted from the tandem SEC columns isocratically with 0.1 M disodium phosphate buffer (pH 6.8) containing 0.1 M sodium sulfate and 0.05% sodium azide at a flow rate of 1 mL/min. Elution of the protein was monitored

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

by UV absorbance at 280 nm. The collected pools of the fractionated aggregate (dimer in this case) were combined and concentrated using a Microcon-YM-10 10 kDa molecular weight cut-off centrifugal filter. Protein concentration of the isolated aggregate was measured using the A280 method, which assumes that the molar extinction coefficient is unaltered in the aggregate. This isolated aggregate sample was then analyzed for purity with an injection volume of 10 µl for samples at 0.5 to 1.5 mg/mL concentration by SEC.

The adaptation of light scattering to eluting components from SEC for detection of soluble aggregates has permitted a more accurate understanding of size distributions, aggregate heterogeneity, and mechanistic details of net-irreversible processes (*136*) than from SEC alone. Figure 2 presents the SEC-MALLS results of isolated aggregate for the NISTmAb for heat-stressed (40 °C for 12 months) and frozen (−70 °C) samples. Heat stress generally increases aggregation (as well as chemical modifications). The −70 °C sample is a reference point of isolated unstressed control dimer and shows some reversion occurred to monomer. Noted in the figure is the appearance of the aggregate (dimer) as measured by absorbance at 280 nm and light scattering data. Higher order aggregates are obscured by the presence of the column shedding that had occurred and can be seen to dominate the light scattering signal (a consequence of large particle scattering). Column shedding may be minimized by the choice of elution buffers, column conditioning, and type of column material used (e.g., low particle shedding stationary phase), but it is generally unavoidable. The data show minor amounts of monomer, and a majority of the sample is dimer.

The molecular weight of the isolated aggregate (dimer peak) was found to be 343 ± 7 kDa. The result suggests a slightly greater complex mass than the expected 296 kDa for the dimerized version of the NISTmAb. The source of the current mass disparity is unreconciled. A hydrodynamic radius based on light scattering for aggregates that were non-stressed was found to be 10.0 ± 0.5 nm. Using a shape factor parameter derived from the ratio of the hydrodynamic and gyration radii (R_h/R_g), a value of 1.53 was observed, suggesting an elongated shape for the aggregates (*137*). The radius of gyration, R_g, can be determined from the Zimm equation using static light scattering methods. Dynamic light scattering can be used to measure the hydrodynamic radius, R_h, using the Stokes-Einstein equation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

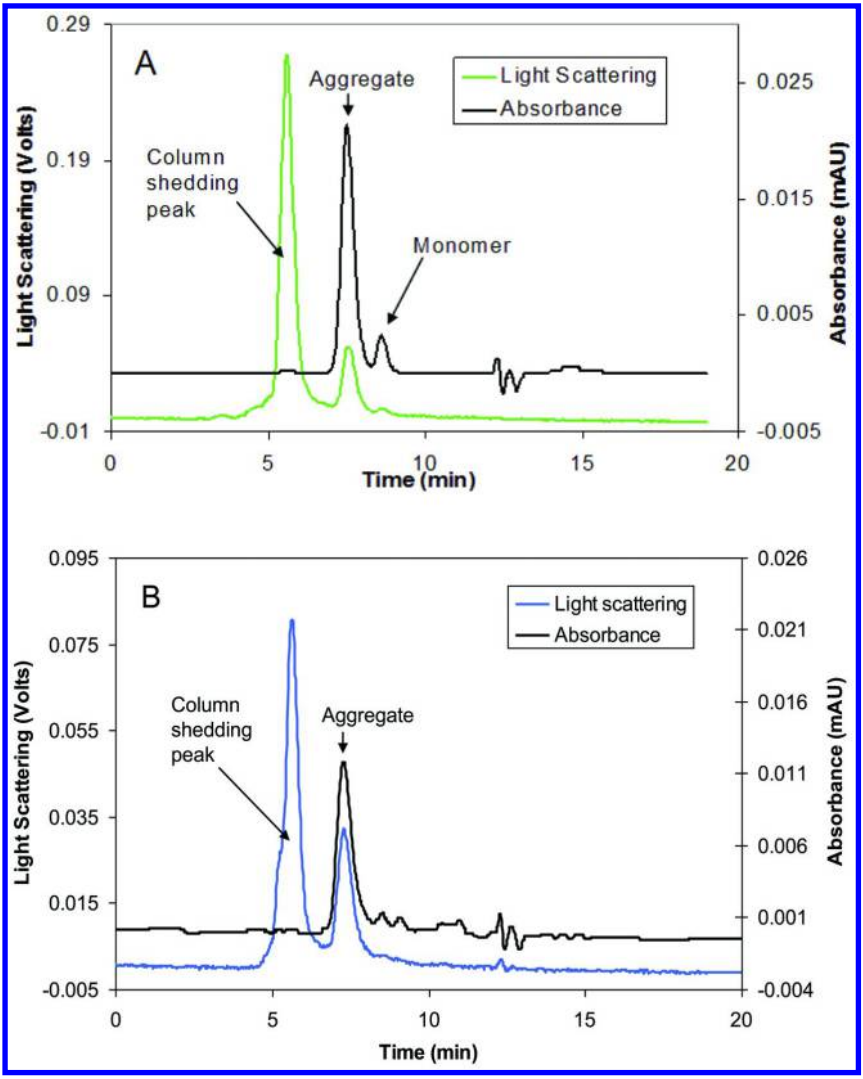

Figure 2. Size exclusion chromatography–multi-angle laser light scattering (SEC-MALLS) analysis of aggregate (dimer) purified from NISTmAb stored at (A) −70 °C and (B) heat stressed at 40 °C for 12 months.

Analytical Ultracentrifugation (AUC)

The size distribution of solution phase aggregates prior to dimer isolation can be ascertained by AUC. See the Biophysical chapter/Volume 2, Chapter 6 for more details about this technique. In general, this method has played an important role with regard to the assessment of true solution phase characterization and identifying artifacts from conventional chromatography methods like SEC (*138–140*). In Figure 3, a thawed, frozen NISTmAb sample (control) is compared

to a heat-stressed NISTmAb sample in the same solution environment by sedimentation velocity AUC. Depicted in the data, heat stress material gives rise to peaks of fragments, dimer and higher order soluble aggregates. In the heated sample it is apparent that the dimer has grown in intensity and the subtle appearance of higher order aggregates is observed. Thus, the dimer is the predominant aggregated form of the heat stressed NISTmAb.In some cases broadening may be observed and could arise from at least two important areas: (1) heterogeneity of size/shape in a given distribution (e.g., dimers and higher order aggregates); and (2) from the treatment of diffusion pertaining to shape and intermolecular interactions in the distribution analysis (*141*).

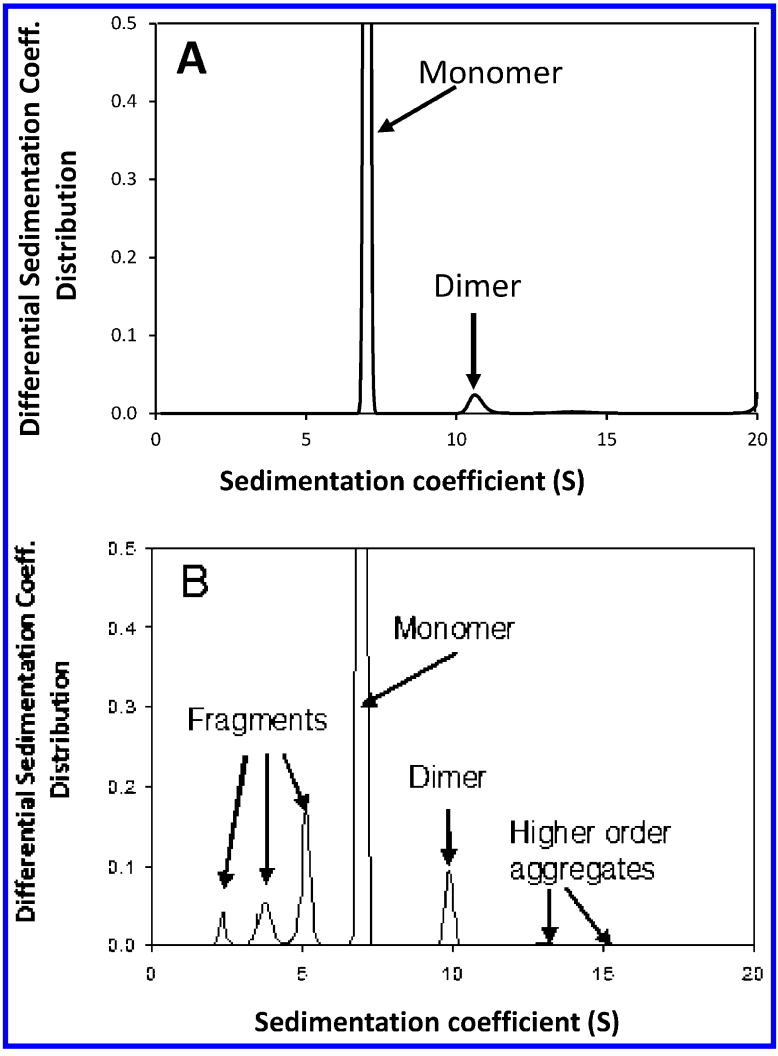

Figure 3. Distribution of sedimentation coefficient in NISTmAb samples stored at (A) −70 °C(control) and (B) heat stressed at 40 °C for 12 months.

Circular Dichroism (CD)

Far-UV CD measurements of the NISTmAb were performed on a Jasco J-810 spectropolarimeter. The purified aggregate samples were diluted to 0.1 mg/mL with 1 mM phosphate buffered saline and placed into a 1 mm quartz cell. The measurements were performed at a medium sensitivity range (100 mdeg) with a 0.1 nm data pitch and 8 sec integration time in the range from 190 to 260 nm. Five independent scans were averaged for the data presented.

The CD method or differential asymmetric absorbance of circularly polarized light of opposite rotations (reflected by the change in refraction indices causing ellipticity) is commonly used to assess mAb structure and aggregative motifs. See the Biophysical chapter/Volume 2, Chapter 6 for more details about this technique. Whereas amide bond absorption in the far-UV (190–250 nm) provides typical secondary structural assessment, aromatic chromophore absorption in the near-UV (250–320 nm) provides evaluation of tertiary structure. Examination of the far-UV structure for the isolated NIST-mAb aggregate (dimer) is depicted in Figure 4. Below 210 nm, there are delineating spectral differences observed between the purified monomer and the non-heat stressed and heat-stressed aggregates. These observed changes in the spectra could be suggestive of differences in β-sheet structure as noted by the reduced positive molar ellipticity near 200 nm (*142*). It also could be ascribed to an increase in disordered structure (random coil-like). Furthermore, the heat-treated aggregate sample was more conformationally altered than the non-heat stressed sample. Thus, this is a case where temperature-stressed dimer aggregate shows irreversibly altered structure by far-UV CD that may suggest an unfolding-mediated aggregation pathway (see Figure 1). The altered structure also can be a consequence of solution condition where the study was not conducted at a structurally stable pH. For example, the far-UV CD of an IgG2 dimer showed no evidence of structural alteration until the pH was adjusted from 6 (more structurally stable) to 4, where more negative molar ellipticity was observed near 218 nm (*143*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

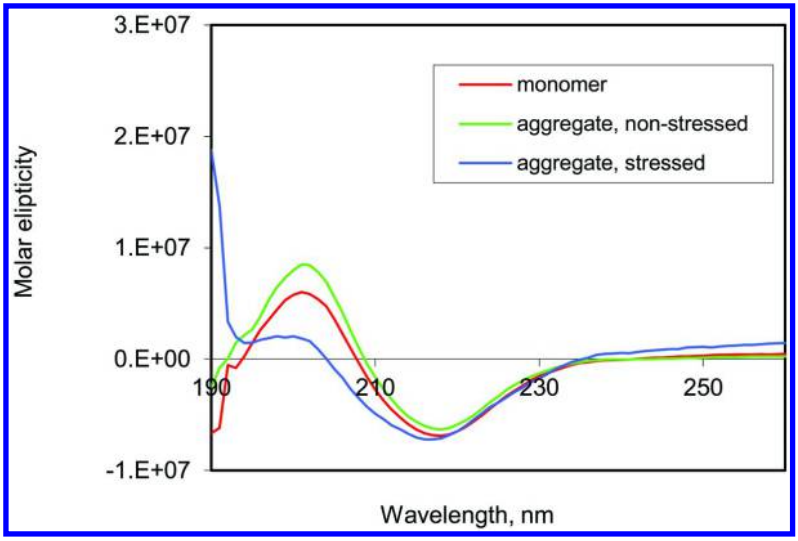

Figure 4. Far-UV of NISTmAb monomer (red line), isolated aggregates (nonstressed) from NISTmAb (green line), and isolated aggregates from NISTmAb heat stressed at 40 °C for 12 months (blue line).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Fourier Transform Infrared Spectroscopy (FTIR)

FTIR is also a commonly used and valuable measure of antibody structure. See the Biophysical chapter/Volume 2, Chapter 6 for more details about this technique. Resolution enhancement of the amide I band of the infrared has been used to characterize the secondary structure of proteins (*144*). It has been reported that spectral features at ~1689 cm^{-1} and ~1616 cm^{-1} can be related to intermolecular β-sheet ascribed to aggregates (*145*). Looking at the second derivative of the isolated NISTmAb aggregate in comparison to the NISTmAb monomer (Figure 5), there are secondary structural alterations in the β-sheet (intramolecular in the vicinity of ~1638 cm^{-1} and ~1689 cm^{-1}) and turn (from ~1660 cm^{-1} to ~1680 cm^{-1}) structures. The identification of distinguishing intermolecular β-sheet, however, is not that obvious. Other FTIR studies of mAbs have shown a characteristic band around 1616 cm^{-1} that remains intact and native-like even when there are distortions in the intramolecular β-sheet frequencies (*146*). Thus assignment to a particular aggregate β-sheet structure is difficult to ascertain with mAbs. What can be ascertained is that structural alteration is observed between the NISTmAb aggregate and native monomer involving β-sheet and turn structures.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

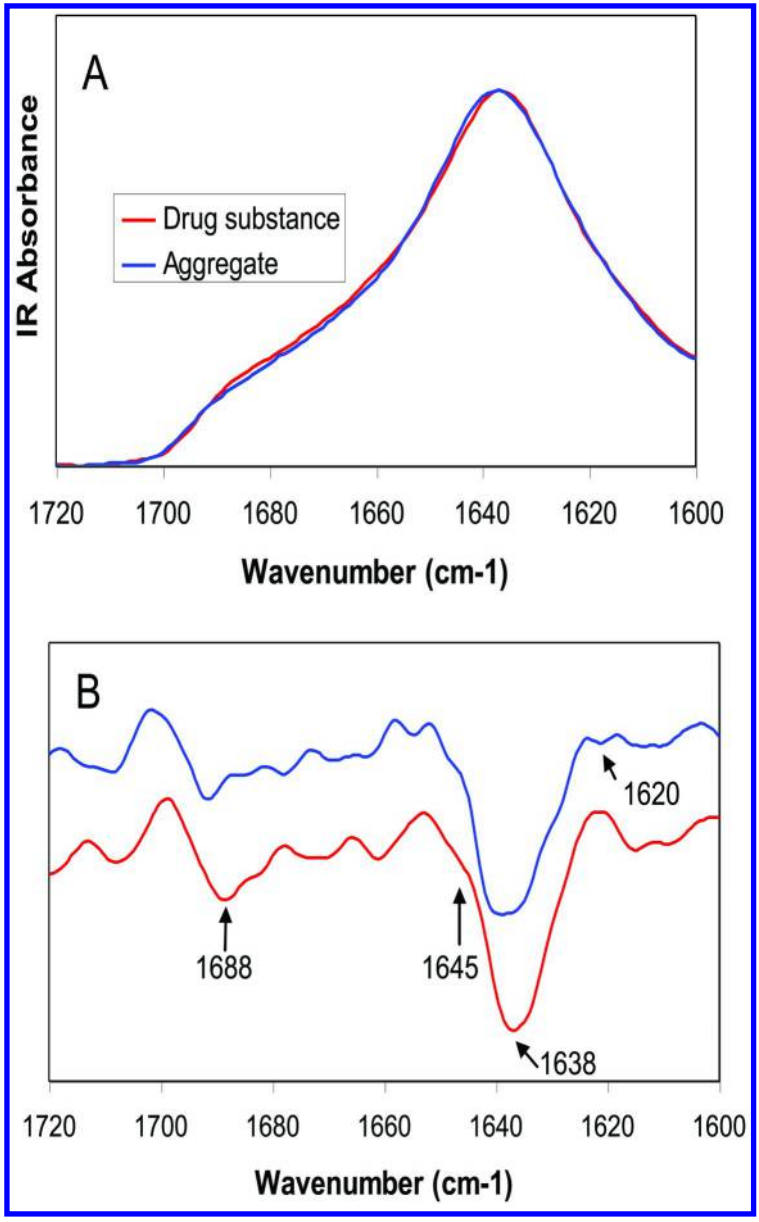

Figure 5. Fourier transform infrared spectroscopy (FTIR) spectra of NISTmAb monomer (red line) and isolated aggregate (blue line). (A) IR absorbance after water bending mode subtraction; (B) second derivative resolution enhancement of the IR absorbance.

Fluorescence

Tryptophan intrinsic fluorescence can be sensitive to conformation and structure alterations involving secondary and tertiary structure. See the Biophysical chapter/Volume 2, Chapter 6 for more details about this technique. Changes in fluorescence intensity depend on the microenvironments in which the Trp residue finds itself. For example, a low dielectric environment (buried in a non-polar core) will be less effected by quenching and thus have a greater fluorescence intensity than would be the case in a high dielectric environment, like water (*147*). Furthermore, a red shift is more indicative of a non-polar (low dielectric) environment moving into a more polar (high dielectric) environment as might be seen when the protein unfolds (*148*, *149*). The fluorescence data of the isolated heat-stressed aggregate (dimer) in Figure 6 suggest a small if any intensity increase, but do exhibit a very small blue shift that could be solvent-exposed Trp becoming buried (in a low dielectric environment) as a consequence of forming aggregate. In contrast, the non-heat stressed isolated aggregate does not show this blue shift behavior. It is noted above that conformational change accompanies the heat-stressed dimer sample described by the infrared and CD data reported above.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

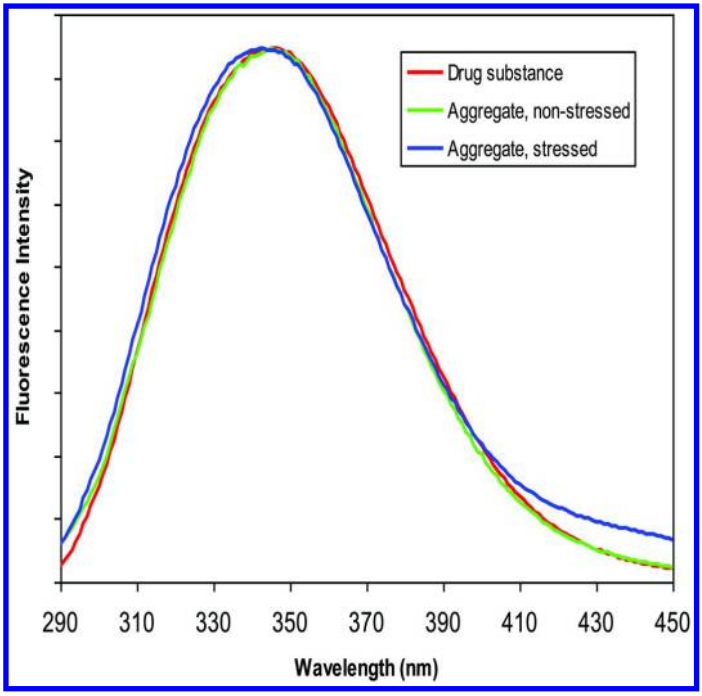

Figure 6. Fluorescence intensity (excitation at 279 nm) comparison of NISTmAb monomer drug substance (red line), isolated aggregate stored at −70 °C (green line), and isolated aggregate from heat-stressed NISTmAb at 40 °C for 12 months (blue line).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Emerging Technologies

Imaging Technologies

Transmission Electron Microscopy (TEM)

TEM is a technique whereby a beam of electrons is transmitted through an ultra-thin specimen and interacts with the sample to form an image that is magnified via focusing onto a fluorescent screen on a layer of photographic film or by a detection sensor like a charge-coupled device (CCD) camera. For biological samples, either fixation using a negative staining material such as uranyl acetate or plastic embedding procedures may be used.

TEM has been widely used for the characterization of mAbs, however, its value for the characterization of mAb aggregation is only recently emerging. Immunoglobulin molecules have hinge regions that connect two Fabs and one Fc together. The hinge regions afford flexibility to bind to antigen partners with the required specificity located in the Fab domain CDRs. The hinge is comprised of upper (DKTHT: IgG1), core (or CPPC: IgG1), and lower amino acid segments (PAPELLGG: IgG1) (*150*). The best imaging methods to date have utilized TEM approaches to resolve intermolecular interactions at the domain level (*151*, *152*). However, the flexibility of the structure can impose a variety of different aggregate shapes and sizes that do not necessarily correspond to rigid orderly structures that pack densely to form regular geometric shapes (see Figure 7). Thus, aggregates appear at early oligomeric stages to be somewhat irregular extended chain and closed ring structures that pack together yielding globular and extended arrays at the dimer, trimer, and tetramer stages (*152*). If rigid structures are used to assess size, closed ring shape structures based on Fab-Fab interactions can generate complexes on the order of 30 nm (dimer), 31 nm (trimer), and 31-43 nm (for tetramers) (*152*). Furthermore, the physical properties of the Fab variable regions (V_H and V_L) can play a role in the formation of aggregate assembly based on hydrophobicity and weak interfacial stabilization (*153*).

If the dimerized version of an intact IgG antibody is considered the simplest aggregate form and the potential building block for higher order forms of aggregates, then possible contrasts and similarities could be extracted from studies already carried out. In fact, studies suggest that considerable heterogeneity exists in just the dimerized version of a mAb (*86*). All domain-domain permutations have been cited as possible in isolated preparations of Epratuzumab dimer (with inter-molecular Fab-Fab, Fab-Fc, Fc-Fc interactions), but they have not always been observed. For example, a more recent evaluation of isolated dimer from Palivizumab suggests that the dimer was composed of Fab-Fc and Fab-Fab (*154*). Results of these findings tend to agree that minor if any conformational alterations of mAb dimers are observed by CD (far- and near-UV), second derivative UV, and intrinsic fluorescence structural methods. Inclusion of higher order aggregates with dimers have shown similar far-UV CD spectra that make it difficult to distinguish aggregates from monomer using this biophysical technique (*155*). This could be due to the method's limited sensitivity to small changes in structure in such cases.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

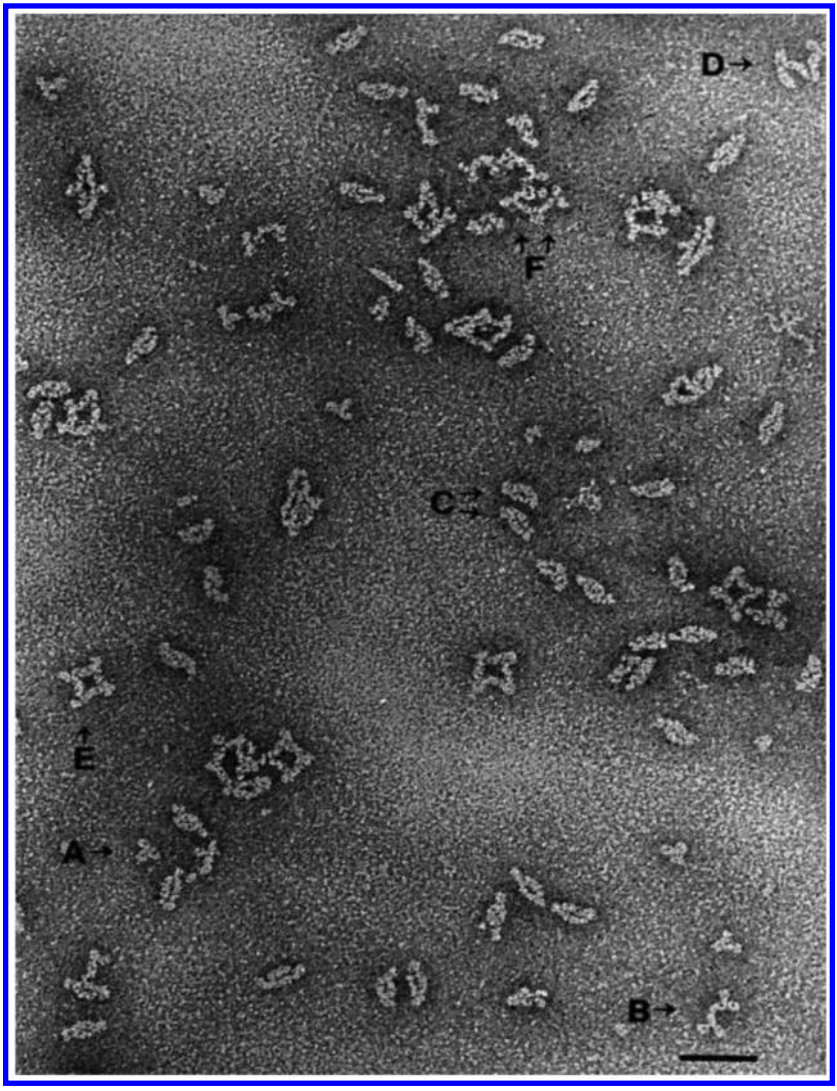

Figure 7. Negative stain transmission electron microscopy (TEM) of a field of mAb complexes that form a variety of initial stage aggregate variations. A = monomer; B = monovalent dimer; C = divalent (closed ring) dimers; D = chain trimer; E = ring tetramer; F = extended chain multimer. Bar = 40 nm. From reference (158).

Investigation of structure and function for different mAb dimers originating from different process stresses suggested the promotion of non-covalent interactions between two Fab domains with reduced potency and antigen binding affinity (*156*). Moreover, low pH stress generated more stable but also non-covalently associated dimers without chemical alterations in the typical "closed" conformations observed by TEM.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Recently, the cryo-TEM method has been used to determine the structure of large molecular entities like the *Thermoplasma acidophilum* 20 S proteasome at 2.8 Å resolution (*157*). Application to relatively large protein molecules makes this technique particularly promising in understanding the macrostructures of higher order aggregates beyond the dimer for antibodies. For example, antibody aggregates in the submicron size range have shown distinct multimer structures surrounded by additional complex aggregates of assorted size and shape (*158*). It may be that in the not too distant future this technology could be applied to elucidate the pathway of aggregate formation at a level of detail not previously possible.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Atomic Force Microscopy (AFM)

AFM can be used to image, measure, and manipulate matter at the nanoscale range. Information is obtained by touching the surface with a highly sensitive mechanical sharp tip probe where the tip radius of curvature is on the order of nanometers (resolution of the technique depends on the sharpness of the probe). Signals from the surface are relayed through a conducting cantilever that can measure mechanical contact forces (e.g., chemical bonding, van der Waals, electrostatic, solvation). Cantilever deflection is accurately measured by laser spot movements from the top surface of the cantilever into a set of photodiodes that electrically send a data signal to a detector and feedback electronics loop assimilated by a computer.

Although AFM is a technique that has been widely used for several decades, the resolution and protein softness of mAbs (as it affects the cantilever tip resolving capability) has so far generated images that crudely reveal particulate structures. Perhaps the best applied methods to image immunoglobulins has been by cryo-AFM imaging at low temperatures (~85 K). Although not as detailed as TEM, the individual domains can be made out for some individual IgGs, however delineation of Fab and Fc domains was not easily resolved in aggregate forms (*159*). Another point of limitation of this technique is that cryo-freezing on a substrate (e.g., mica) may in and of itself contribute to disparities that are not particularly congruent with those observed in solution phases. Nevertheless, useful information of mAb aggregate structural morphology and elucidation of aggregate pathway intermediates has been reported to be feasible by AFM (*160*).

Macroscopic Technologies

Native and Ion Mobility Mass Spectrometry

It is well-established that non-covalent protein assemblies can be preserved through gentle transition into the gas phase—achieved primarily through a soft ionization and desolvation process, shallow pressure gradients, and a reduced ion acceleration—relative to non-native MS approaches (*161–164*). As such, so-called "native mass spectrometry" can yield information on the mass and conformation of non-covalent protein assemblies, including protein aggregates.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Another MS tool that can discriminate aggregates of different topology and shape yet equivalent mass is ion mobility. Ion mobility is a shape-selective tool for modulating the velocities of species in the gas phase, somewhat analogous to electrophoresis in solution. This was originally the preserve of hybrid instruments containing a drift cell, but it is now widely available in a variety of formats, each of which implements a version of separating molecular ions in the gas phase by properties relating to their shape. Thus, the nature of higher order assemblies of protein have been determined with greater resolution using these techniques, complementing other well-established biophysical techniques—such as SEC, light scattering, and gel electrophoresis (*165*). Ion mobility is achieved either by retardation in a drift cell or by gas-induced rollover from the travelling wave in an ion guide. The technique can discern between conformers that have identical mass yet different molecular shape. Together these structural MS tools can be used to characterize the distribution of oligomeric states and even provide insight into the aggregation pathway (*166*, *167*).

By inspection of the mass spectrometric charge state distribution, topologically distinct forms can be observed (*168*, *169*). Further, once a non-covalent complex has been successfully retained into the gas phase, it can be deconstructed to identify the constituent polypeptides, for example, by breaking non-covalent interactions through low energy, collision-induced dissociation (CID) (*170*) or surface-induced dissociation (SID) (*171*). By step-wise inspection of the released species as the dissociation energy is ramped up, the topology of their quaternary structure may even be determined. New developments, such as top-down electron transfer/capture dissociation, provide the possibility of cutting the polypeptide chains of a complex where they are exposed in the gas phase and thus mapping the protein-protein interfaces (*172*). As hydrophobically driven aggregation is a process that is thermodynamically driven by the bulk solvent, it may not be possible to preserve certain protein complexes or aggregates as they desolvate into the gas phase. Although this precludes analysis by MS, it can yield insight into the hydrophobic nature of the interaction observed in solution phase studies.

Thus by native-MS, the dispersivity, composition, topology, and to some extent, structure of protein higher order assemblies have been determined with orthogonality to other well-established biophysical techniques. Although it is mostly restricted to nanoelectrospray ionization, which will significantly limit throughput from a more routine viewpoint, it has advantages in terms of mass accuracy, resolution of oligomeric species, and the ability to potentially deconstruct individual non-covalent complexes from within a complex mixture. Refer to the Ion Mobility chapter/Volume 3, Chapter 4 for more information about this technique.

X-ray and Neutron Scattering

Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) are techniques that can provide structural information about proteins, particularly in disordered systems, which can be difficult to analyze by other

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

techniques. The intensity as a function of scattering angle is measured after passing an incident beam of X-rays or neutrons through a protein solution. Interference in the scattering due to the interaction of X-rays or neutrons with the electron density of nuclei at different positions within the molecule provides information about its structure. A plot of the scattering intensity versus the scattering vector can then be analyzed by Fourier transformation to obtain structural information such as the molecular weight and radius of gyration (*173*). These techniques are not limited by the requirement of obtaining high diffraction quality crystals for crystallography or the molecular mass limitations of NMR (*174*). Both of these are advantages for the analysis of antibodies, which can be hard to crystallize and exceed the molecular weight limits of current conventional NMR technologies. An additional important advantage of SAXS and SANS is that they can be used to probe highly concentrated solutions of antibodies. The structural dynamics and stability of antibodies at high concentration can potentially alter their aggregation propensity and structure, as well as the physical properties of the solutions.

Information from SAXS and SANS can provide valuable insight into the properties and behavior of antibody solutions. For example, formation of reversible dimers with extended structures in dilute solutions of an antibody was characterized with SAXS (*175*). Modeling of the SAXS and rheology data indicated that anisotropic interactions between complementary surfaces are required to nucleate and propagate protein clusters in solution (*175*). SAXS was used to resolve the simultaneous attractive and repulsive interactions and the impact of excipients on the interactions for an IgG2 antibody at high concentration (*176*). The conformational diversity of antibody molecules in different formulation conditions has been probed using SAXS (*177*). In one study, structures of two antibodies were shown to shift toward populations with a more "open" conformation in arginine chloride solutions (*177*). Similarly, SANS has been used to investigate the solution structure and protein-protein interactions of antibodies at high concentration (*178*). In this study of two antibodies (*178*), the SANS data were modeled with analytical form factor for a three-arm representation of the antibody molecules, allowing interpretation of the solution conformation of the antibodies at high concentration (*178*). The two molecules studied were closely related, having only small sequence differences in their CDR, yet they had very different solution viscosities. The results showed that there was little change in the conformation of both antibodies within the crowded, high-concentration solution (*178*). However, the SANS data suggested that the high solution viscosity of one of the antibodies was caused by anisotropic charge-mediated attraction forces, resulting in the formation of dynamic clusters. In contrast, the second antibody protein-protein interactions were repulsive at high concentrations, leading to a lower viscosity (*178*). In a related technique, the structure of high-concentration antibodies adsorbed to surfaces also has been investigated using neutron reflectometry. Neutron reflectometry was used to probe the adsorbed layer thickness and aggregation of antibodies at hydrophilic and hydrophobic surfaces at high solution concentrations (*62*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Limited Proteolysis and Cross-Linking

MS-based approaches provide a unique ability to study dynamic interactions between proteins that may not be possible using NMR or crystallography techniques. Solution phase hydrogen/deuterium exchange (see the Hydrogen/Deuterium-Exchange Mass Spectrometry [HDX-MS] Section), limited proteolysis, and cross-linking techniques are potentially valuable tools to investigate protein-protein interactions in antibodies, including interfaces of self-association.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Limited Proteolysis

Limited proteolytic digestion under non-denaturing conditions with MS can be used to study the tertiary and quaternary structure of proteins. In the analysis of protein structure, the factors that govern the selectivity of cleavage are the sequence specificity of the enzyme and the accessibility and flexibility of the site to the protease due to higher order protein structure. The distribution of amino acids in a protein guides the choice of protease to be used as a structural probe. Proteases that cleave at hydrophilic sites (e.g., trypsin) often are preferred in structural analysis because amino acids with hydrophilic side chains are found in greater abundance at the solvent interface. Resulting peptide mixtures are analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) or reversed phase-high-performance liquid chromatography (RP-HPLC) coupled to electrospray ionization mass spectrometry (ESI-MS) for peptide identification (*179*). Limited proteolysis also can be used to probe the quaternary structure of protein assemblies; here, the protease is used to provide contrast between the associated and unassociated states of the system. The formation of an interface between two proteins will exclude both solvent molecules and macromolecules, such as proteases, and will protect otherwise accessible sites from proteolytic digestion. Application of limited proteolysis to characterize interfaces in protein assemblies is well established (*180*, *181*); however, studies applying this approach to intermolecular interactions in antibody aggregates are less prevalent. Recently, reported limited proteolysis experiments on isolated dimers of Palivizumab (human IgG1) localized dimer interfaces to both Fab-Fc and Fab-Fab but not Fc-Fc interactions (*154*).

Cross-Linking Approaches To Study Antibody Aggregates

Recently, the characterization of therapeutic protein self-association and aggregation has begun to see success through cross-linking studies (*182*). The antibody oligomers and aggregates are typically non-covalent in nature, although they may be covalently linked in some cases through disulfide bonds. Therefore, they are not amenable to some of the typical and robust assays that yield information on molecular weight (MS, gel electrophoresis) or particle size (light scattering, SEC). By covalently stabilizing the non-covalently associated states,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

these standard assays are able to yield this information without risk of "false negative" data at the higher order state level (i.e., missing data on the aggregate species). Well-established cross-linking methods are based on free amine availability (e.g., lysine) and require two such groups to be available within close proximity (< 20 Å) on the surface of the aggregate. A bifunctional cross-linker, such as bis(sulfosuccinimidyl)suberate (BS3) with N-hydroxysuccinimide (NHS) groups separated by a spacer (typically an alkyl chain), can be incubated with the protein aggregate to cross-link these reactive amines. Where the amines are on separate protein chains, the oligomer or aggregate will be covalently stabilized and thus amenable to analytical techniques such as MS and reducing SDS-PAGE that do not preserve non-covalently associated species. However, a serious limitation to applying this approach as it stands in the study of minor population aggregate species is that the stability of the cross-linker reagents in aqueous buffer is low (half-life on the order of minutes at 20 °C). This naturally limits the detection of minority species and limits the ability to probe concentration dependence of antibody oligomerization. One possible solution to this is the emergence of a cross-linking chemistry whose reagents are more stable under typical aqueous formulation buffer conditions (*183–185*). Thus, it allows for a more complete cross-linking of non-covalently associated aggregates. This has been shown to reliably indicate dissociation constants of non-covalent oligomeric species (*186*, *187*). Together, these chemistries represent a way of covalently trapping oligomers, allowing for slow or energetic analysis.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

High Resolution Technologies

Hydrogen/Deuterium-Exchange Mass Spectrometry (HDX-MS)

Recently, HDX-MS has proved itself a sensitive probe of protein conformation and dynamics (*188–194*). There is particular merit in this approach when applied to the study of large and dynamic proteins, such as monoclonal IgG antibodies (cf. 146 kDa for a human IgG1), which are intractable by other techniques (e.g., NMR, X-ray crystallography). See the Higher Order Structure chapter/Volume 3, Chapter 2 for more information on HDX and NMR. This is even more acute when considering protein assemblies and aggregates, the masses of which reach into the megadalton range.

By incubating a protein in a deuterated buffer, the labile hydrogen atoms exchange with deuterium over time. This can be accurately measured as a mass increase of 1.006 Da. The kinetics of exchange, including back-exchange when in a mixture of protonated and deuterated solvent, are now well defined and relate to the hydrogen bonding environment of the exchanging site, its solvent-accessibility, and the temperature and pH of the system (*195*, *196*). The latter two factors are tightly controlled experimentally, and very fast-exchanging labile sites, such as hydroxyls and carboxylates, are exchanged back for hydrogen upon quenching, prior to analysis. Thus, the MS measurement of the protein backbone amide exchange rate kinetics is a sensitive probe of structure and conformation.

HDX-MS can yield submolecular-level information on the sites within a protein that are exposed/protected in a protein aggregate (*197–202*). It

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

also can provide information as to which sites have undergone a structural or conformational rearrangement from the native form (*203–205*). Note that as achieving submolecular localization relies on either proteolytic digestion or top-down fragmentation, the nature of the aggregate will define whether this approach can yield insightful data. Without an effective release of aggregate-derived peptides, signal resulting from the relevant form will not be apparent. Equally, a disordered aggregate will yield data with a degree of randomness that will complicate interpretation. As such, one may choose to infer the aggregated state through comparison of intrinsic (e.g., sequence) and extrinsic (e.g., formulation) factors in a panel of low-order assemblies (monomer, dimer, etc.), bridged with orthogonal data that relate to the impact on rate and propensity of aggregation (*206*, *207*).

Nonetheless, because HDX-MS data can be generated on a solvated mAb form in most formulation buffers, it has the potential to yield relevant information on the aggregates at a submolecular resolution (peptide level or better) that is informative for lead optimization, characterization, and drug formulation development.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Nuclear Magnetic Resonance (NMR) Spectroscopy

It has been noted that NMR spectroscopy of protein aggregates results in low dispersion and severe line-broadening of NMR signals (*208*, *209*). See the Higher Order Structure chapter/Volume 3, Chapter 2 for more about this technique. For detailed structural perturbations at the residue level, ^{1}H, ^{13}C, and ^{15}N resonance assignments have been made for *E. coli*-expressed isotope enriched IgG1:Fc, a 51 kDa protein (*210*). Likewise, the NMR resonance assignments for the reduced form of human IgG1 C_H3 domain, a 26 kDa dimer in solution (amino acid residues 341–447), has been elucidated from ^{1}H, ^{13}C, and ^{15}N resonances (*211*). NMR examination of such isotope-enriched materials at conditions that promote aggregates can provide insights with regard to mechanism. For example, a detailed understanding of conformational changes arising from oxidation of methionines 33 (Met252 in complete heavy chain sequence) and 209 (Met428 in complete heavy chain sequence) on *E. coli*-expressed IgG1:Fc provided insights about aggregation relationships for this complex domain (*212*). The ^{1}H-^{15}N HSQC (heteronuclear single quantum correlation) spectra were able to show residue-specific differences between oxidized and non-oxidized samples that suggested a relaxation of the C_H2 and a compression of the C_H3 domains (see Figure 8). Residues that disappeared or that have shifted portray significant perturbations in amide backbone structure, likely attributed to different chemical and electronic environments. This was further confirmed from DSC measurements, where a slight high temperature shift in the C_H3 domain T_m was observed with a prominent low temperature (destabilizing) shift in the C_H2 domain T_m. The outcome of the oxidation clearly showed a propensity to aggregate as a consequence of methionine oxidation. It should be noted that there remains the possibility that isolated Fab and Fc domains may not behave the same as in the intact antibody molecule.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

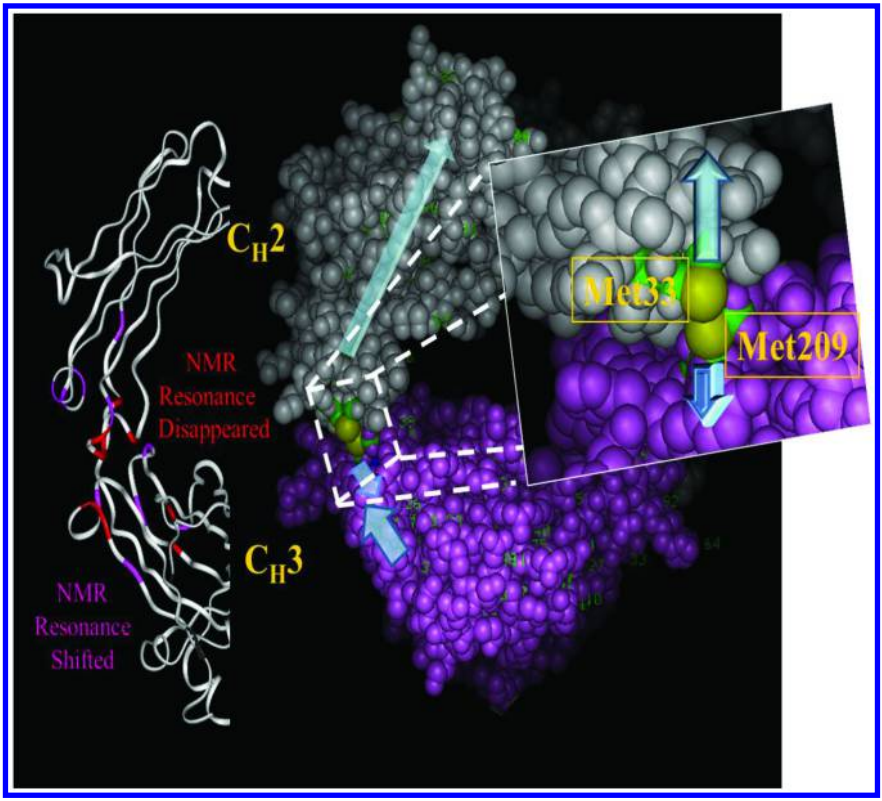

Figure 8. Split out of part of the IgG1 Fc showing the C_H2 and C_H3 domains with a ribbon structure (left) that is color coded to show amino acid regions that exhibit nuclear magnetic resonance (NMR) resonances that disappeared (red) and shifted (magenta) from oxidation of Met33 (Met252 intact heavy chain) and Met 209 (Met428 intact heavy chain) based on ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectra (212). The electron density structure of the same region is shown in the middle with the methionines colored in yellow (protein data bank [PDB] entry IHZH). The stress that occurs presumably by repulsion when both methionines are oxidized is illustrated by the light blue arrows. The C_H3 domain pushes against itself in the form of a compression, while the C_H2 domain expands and becomes less rigid. This influence contributes to a greater susceptibility to an altered conformation that is prone to form aggregates.

Another point to consider is the stabilizing effect of glycosylation pertaining to the C_H2 domain of the Fc. Aglycosylated C_H2 domains of IgG1:Fc have a greater propensity to form aggregates than glycosylated variants at acidic pH unfolding conditions (*114*). Studies of non-glycosylated IgG1 Fc (less the Fab domains) suggest that the C_H2 domain of the Fc is most susceptible to conformational alterations and can lead to non-native aggregate structures as a consequence of methionine oxidation within the C_H2-C_H3 interface (*212*). As for the role of glycosylation on the behavior of the C_H2 domain, it remains unclear as

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

to whether methionine oxidation is the primary route of destabilization or whether a more elaborate mechanism involving glycans is involved (*213*). Of great interest would be to perform similar NMR experiments on isolated Fab domains in future studies to grasp a deeper understanding of intermolecular interactions and their relationship to protein engineered changes to mitigate them.

In Silico Aggregation Prediction

Several algorithms exist for the prediction of aggregation-prone sequence motifs in polypeptides, such as Zyggregator (*214*), TANGO (*215*), PASTA (*216*), and AMYLPRED2 (*217*). These are mostly aimed at the identification of amyloid formation rather than amorphous or residually structured aggregation mechanisms. Recently, however, there has been some success in computationally predicting mAb self-association and aggregation. These and other aggregation-scoring algorithms have been applied to identify potential aggregation hotspots in a diverse set of antibodies, both in pre-clinical and clinical development, with some success (*218–220*).

The SAP algorithm analyzes any given mAb as input and attempts to determine exposed surface patches at high risk of being associated with elevated aggregation (*108*, *221*, *222*). The SAP score for any given amino acid within the protein of interest is determined as a function of solvent-accessible surface area (over the course of an atomistic molecular dynamics simulation, if available) and weighted by the intrinsic hydrophobicity of the residue. The hydrophobicity scale is normalized with glycine at zero; thus, residues that are more hydrophobic have a positive value, and those that are less hydrophobic have a negative value. The sum of these terms for a sphere with radius of 5–10 Å centered over each atom in a residue is then averaged to yield a per residue SAP score. The output for the NISTmAb Fab domain is shown in Figure 9. The NISTmAb Fab region SAP analysis shows a predominance of negative (blue) or neutral (white) scores at the protein surface. There are some areas with positive scores (red), notably amino acids 97 and 98 in CDR3 of the heavy chain, visible in the center of the right panel of the figure. However, the surface area affected is small, and the score never reaches above 0.4. Therefore, these sites are not expected to result in a significant risk of aggregation. In fact, there are only four amino acids in all of the NISTmAb that achieve a score by SAP of > 0.25. By these in silico measures, the NISTmAb could be expected to be resistant to aggregation.

The SAP method requires a three-dimensional structure of the query mAb. This can be experimentally determined or a high-quality homology model can be used, and it need not be a full IgG. Indeed, a particular strength is the flexibility of the algorithm to be applied to any protein construct. This is a critical aspect of the work flow in determining the accuracy of the predictions. Ideally, the algorithm is run on conformations generated using atomistic molecular dynamics simulations. Additionally, the program has been distributed in commercial structural biology software, allowing for local use. Thus, its general applicability to different antibody-derived formats and distribution for local execution means that it is tractable to commercial biotechnology research and development (R&D)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

applications. As these predictive tools become more refined, not least with better training datasets from experimentally validated panels of antibodies, they may have an impact on biotherapeutic lead isolation and lead optimization, as well as on the drug formulation process.

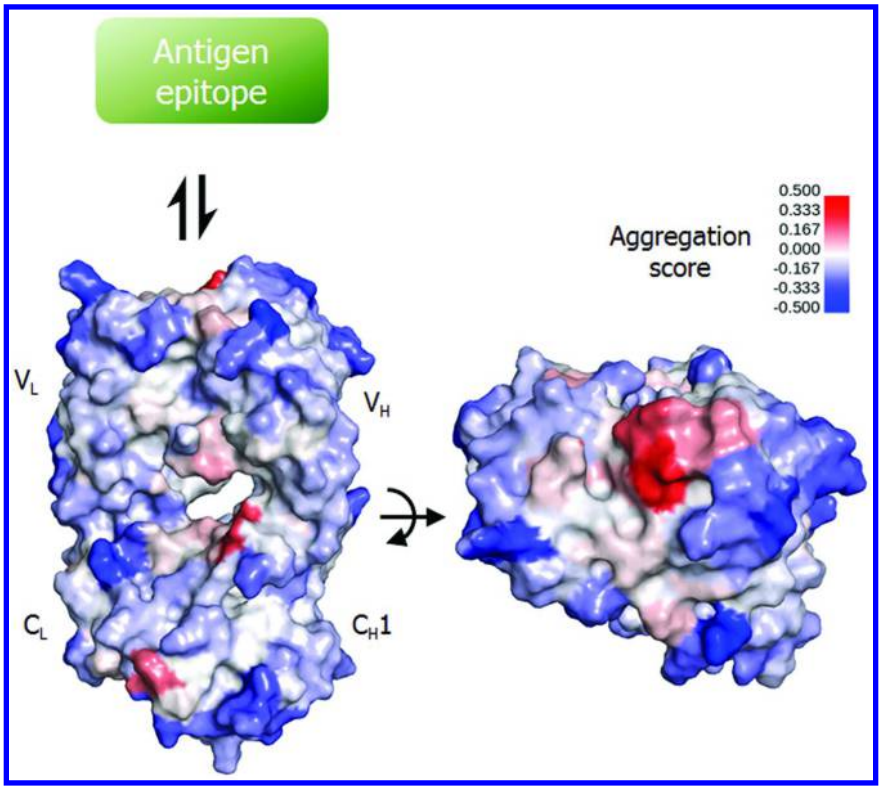

Figure 9. Spatial aggregation propensity score mapped with a 5 Å cutoff for the NISTmAb Fab domain X-ray diffraction co-crystal structure. Each location on the protein surface is assigned an aggregation score that indicates its predicted propensity to result in protein self-association/aggregation events (blue: −0.500 indicates low propensity, red: +0.500 indicates high propensity).

Conclusions and Future Directions

Apparent from the preceding information, it can be concluded that at the very initial stages of aggregate formation for mAbs, there is significant complexity in formation, patterns of assembly, domain interactions that form them, and the physico-chemical properties that govern propensities in different solution environments. Conventional biophysical methods are inadequate to shine the spotlight on the details of interactions that would illuminate the mechanism of the reactions that take place. Thus, the need to find more sophisticated technologies to address this area will become increasingly important in the future. Emerging areas still being explored will include NMR, HDX, and improved imaging

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

technologies. The hope is that applications of these technologies will yield a more accurate picture of the mechanisms for antibody aggregation and promote a better understanding that enables predictive modeling of the phenomena.

References

1. Manning, M. C.; Chou, D. K.; Murphy, B. M.; Payne, R. W.; Katayama, D. S. *Pharm. Res.* **2010**, *27*, 544–575.
2. Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F. *Pharm. Res.* **2003**, *20*, 1325–1336.
3. Kendrick, B. S.; Carpenter, J. F.; Cleland, J. L.; Randolph, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14142–14146.
4. Kim, N.; Remmele, R. L., Jr.; Liu, D.; Razinkov, V. I.; Fernandez, E. J.; Roberts, C. J. *Biophys. Chem.* **2013**, *172*, 26–36.
5. Philo, J. S.; Arakawa, T. *Curr. Pharm. Biotechnol.* **2009**, *10*, 348–351.
6. Ahamed, T.; Esteban, B. N.; Ottens, M.; van Dedem, G. W.; van der Wielen, L. A.; Bisschops, M. A.; Lee, A.; Pham, C.; Thommes, J. *Biophys. J.* **2007**, *93*, 610–619.
7. Mason, B. D.; Zhang-van Enk, J.; Zhang, L.; Remmele, R. L., Jr.; Zhang, J. *Biophys. J.* **2010**, *99*, 3792–3800.
8. Wang, Y.; Lomakin, A.; Latypov, R. F.; Laubach, J. P.; Hideshima, T.; Richardson, P. G.; Munshi, N. C.; Anderson, K. C.; Benedek, G. B. *J. Chem. Phys.* **2013**, *139*, 121904.
9. Sharma, V.; Jaishankar, A.; Wang, Y.-C.; McKinley, G. H. *Soft Matter* **2011**, *7*, 5150–5160.
10. Zhang-van Enk, J.; Mason, B. D.; Yu, L.; Zhang, L.; Hamouda, W.; Huang, G.; Liu, D.; Remmele, R. L., Jr.; Zhang, J. *Mol. Pharm.* **2013**, *10*, 619–630.
11. Ionescu, R. M.; Vlasak, J.; Price, C.; Kirchmeier, M. *J. Pharm. Sci.* **2008**, *97*, 1414–1426.
12. Vermeer, A. W.; Norde, W. *J. Colloid Interface Sci.* **2000**, *225*, 394–397.
13. Jenkins, N.; Murphy, L.; Tyther, R. *Mol. Biotechnol.* **2008**, *39*, 113–118.
14. Ho, S. C.; Koh, E. Y.; van Beers, M.; Mueller, M.; Wan, C.; Teo, G.; Song, Z.; Tong, Y. W.; Bardor, M.; Yang, Y. *J. Biotechnol.* **2013**, *165*, 157–166.
15. Liu, H.; Gaza-Bulseco, G.; Faldu, D.; Chumsae, C.; Sun, J. *J. Pharm. Sci.* **2008**, *97*, 2426–2447.
16. Cromwell, M. E.; Hilario, E.; Jacobson, F. *AAPS J.* **2006**, *8*, E572–579.
17. Buchanan, A.; Clementel, V.; Woods, R.; Harn, N.; Bowen, M. A.; Mo, W.; Popovic, B.; Bishop, S. M.; Dall'Acqua, W.; Minter, R.; Jermutus, L.; Bedian, V. *mAbs* **2013**, *5*, 255–262.
18. Tous, G. I.; Wei, Z.; Feng, J.; Bilbulian, S.; Bowen, S.; Smith, J.; Strouse, R.; McGeehan, P.; Casas-Finet, J.; Schenerman, M. A. *Anal. Chem.* **2005**, *77*, 2675–2682.
19. Hutterer, K. M.; Hong, R. W.; Lull, J.; Zhao, X.; Wang, T.; Pei, R.; Le, M. E.; Borisov, O.; Piper, R.; Liu, Y. D.; Petty, K.; Apostol, I.; Flynn, G. C. *mAbs* **2013**, *5*, 608–613.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

20. Kao, Y. H.; Hewitt, D. P.; Trexler-Schmidt, M.; Laird, M. W. *Biotechnol. Bioeng.* **2010**, *107*, 622–632.
21. Koterba, K. L.; Borgschulte, T.; Laird, M. W. *J. Biotechnol.* **2012**, *157*, 261–267.
22. Trexler-Schmidt, M.; Sargis, S.; Chiu, J.; Sze-Khoo, S.; Mun, M.; Kao, Y. H.; Laird, M. W. *Biotechnol. Bioeng.* **2010**, *106*, 452–461.
23. Gomez, N.; Subramanian, J.; Ouyang, J.; Nguyen, M. D.; Hutchinson, M.; Sharma, V. K.; Lin, A. A.; Yuk, I. H. *Biotechnol. Bioeng.* **2012**, *109*, 125–136.
24. Roberts, G. D.; Johnson, W. P.; Burman, S.; Anumula, K. R.; Carr, S. A. *Anal. Chem.* **1995**, *67*, 3613–3625.
25. Schoneich, C. *Biochim. Biophys. Acta* **2005**, *1703*, 111–119.
26. Bienvenut, W. V.; Deon, C.; Pasquarello, C.; Campbell, J. M.; Sanchez, J. C.; Vestal, M. L.; Hochstrasser, D. F. *Proteomics* **2002**, *2*, 868–876.
27. Grewal, P.; Mallaney, M.; Lau, K.; Sreedhara, A. *Mol. Pharm.* **2014**, *11*, 1259–1272.
28. Matamoros Fernandez, L. E.; Kalume, D. E.; Calvo, L.; Fernandez Mallo, M.; Vallin, A.; Roepstorff, P. *J. Chromatogr. B: Biomed. Sci. Appl.* **2001**, *752*, 247–261.
29. Taylor, S. W.; Fahy, E.; Murray, J.; Capaldi, R. A.; Ghosh, S. S. *J. Biol. Chem.* **2003**, *278*, 19587–19590.
30. Balasubramanian, D.; Kanwar, R. *Mol. Cell. Biochem.* **2002**, *234-235*, 27–38.
31. Correia, M.; Neves-Petersen, M. T.; Jeppesen, P. B.; Gregersen, S.; Petersen, S. B. *PLoS One* **2012**, *7*, e50733.
32. Yang, Y.; Stella, C.; Wang, W.; Schoneich, C.; Gennaro, L. *Anal. Chem.* **2014**, *86*, 4799–4806.
33. Shacter, E. *Drug Metab. Rev.* **2000**, *32*, 307–326.
34. Bee, J. S.; Davis, M.; Freund, E.; Carpenter, J. F.; Randolph, T. W. *Biotechnol. Bioeng.* **2010**, *105*, 121–129.
35. Stadtman, E. R. *Free Radical Biol. Med.* **1990**, *9*, 315–325.
36. Wang, W.; Ignatius, A. A.; Thakkar, S. V. *J. Pharm. Sci.* **2014**, *103*, 1315–1330.
37. Vlasak, J.; Ionescu, R. *mAbs* **2011**, *3*, 253–263.
38. Berasain, P.; Carmona, C.; Frangione, B.; Dalton, J. P.; Goni, F. *Exp. Parasitol.* **2000**, *94*, 99–110.
39. Cordoba, A. J.; Shyong, B. J.; Breen, D.; Harris, R. J. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *818*, 115–121.
40. Perico, N.; Purtell, J.; Dillon, T. M.; Ricci, M. S. *J. Pharm. Sci.* **2009**, *98*, 3031–3042.
41. Chelius, D.; Rehder, D. S.; Bondarenko, P. V. *Anal. Chem.* **2005**, *77*, 6004–6011.
42. Pace, A. L.; Wong, R. L.; Zhang, Y. T.; Kao, Y. H.; Wang, Y. J. *J. Pharm. Sci.* **2013**, *102*, 1712–1723.
43. Wakankar, A. A.; Borchardt, R. T. *J. Pharm. Sci.* **2006**, *95*, 2321–2336.
44. Geiger, T.; Clarke, S. *J. Biol. Chem.* **1987**, *262*, 785–794.
45. Manning, M. C.; Patel, K.; Borchardt, R. T. *Pharm. Res.* **1989**, *6*, 903–918.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

46. Zhang, B.; Yang, Y.; Yuk, I.; Pai, R.; McKay, P.; Eigenbrot, C.; Dennis, M.; Katta, V.; Francissen, K. C. *Anal. Chem.* **2008**, *80*, 2379–2390.
47. Bee, J. S.; Randolph, T. W.; Carpenter, J. F.; Bishop, S. M.; Dimitrova, M. N. *J. Pharm. Sci.* **2011**, *100*, 4158–4170.
48. Thomas, C. R.; Geer, D. *Biotechnol. Lett.* **2011**, *33*, 443–456.
49. Vazquez-Rey, M.; Lang, D. A. *Biotechnol. Bioeng.* **2011**, *108*, 1494–1508.
50. Biddlecombe, J. G.; Smith, G.; Uddin, S.; Mulot, S.; Spencer, D.; Gee, C.; Fish, B. C.; Bracewell, D. G. *Biotechnol. Prog.* **2009**, *25*, 1499–1507.
51. Tyagi, A. K.; Randolph, T. W.; Dong, A.; Maloney, K. M.; Hitscherich, C., Jr.; Carpenter, J. F. *J. Pharm. Sci.* **2009**, *98*, 94–104.
52. van Reis, R.; Zydney, A. *J. Membr. Sci.* **2007**, *297*, 16–50.
53. Nakanishi, K.; Sakiyama, T.; Imamura, K. *J. Biosci. Bioeng.* **2001**, *91*, 233–244.
54. Pinholt, C.; Hartvig, R. A.; Medlicott, N. J.; Jorgensen, L. *Expert Opin. Drug Delivery* **2011**, *8*, 949–964.
55. Norde, W. *Colloids Surf., B* **2008**, *61*, 1–9.
56. Rabe, M.; Verdes, D.; Seeger, S. *Adv. Colloid Interface Sci.* **2011**, *162*, 87–106.
57. Kastantin, M.; Langdon, B. B.; Schwartz, D. K. *Adv. Colloid Interface Sci.* **2014**, *207*, 240–252.
58. Norde, W.; Lyklema, J. *Adv. Colloid Interface Sci.* **2012**, *179-182*, 5–13.
59. Norde, W.; Favier, J. P. *Colloids Surf.* **1992**, *64*, 87–93.
60. Singh, S. K.; Kolhe, P.; Mehta, A. P.; Chico, S. C.; Lary, A. L.; Huang, M. *Pharm. Res.* **2011**, *28*, 873–885.
61. Buijs, J.; Norde, W.; Lichtenbelt, J. W. T. *Langmuir* **1996**, *12*, 1605–1613.
62. Couston, R. G.; Skoda, M. W.; Uddin, S.; van der Walle, C. F. *mAbs* **2013**, *5*, 126–139.
63. Kueltzo, L. A.; Wang, W.; Randolph, T. W.; Carpenter, J. F. *J. Pharm. Sci.* **2008**, *97*, 1801–1812.
64. Majumdar, S.; Ford, B. M.; Mar, K. D.; Sullivan, V. J.; Ulrich, R. G.; D'Souza, A J. *J. Pharm. Sci.* **2011**, *100*, 2563–2573.
65. Gerhardt, A.; Bonam, K.; Bee, J. S.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2013**, *102*, 429–440.
66. Kiese, S.; Papppenberger, A.; Friess, W.; Mahler, H. C. *J. Pharm. Sci.* **2008**, *97*, 4347–4366.
67. Mahler, H. C.; Muller, R.; Friess, W.; Delille, A.; Matheus, S. *Eur. J. Pharm. Biopharm.* **2005**, *59*, 407–417.
68. Bee, J. S.; Schwartz, D. K.; Trabelsi, S.; Freund, E.; Stevenson, J. L.; Carpenter, J. F.; Randolph, T. W. *Soft Matter* **2012**, *8*, 10329–10335.
69. de Jongh, H. H.; Wierenga, P. A. *Biopolymers* **2006**, *82*, 384–389.
70. Rudiuk, S.; Cohen-Tannoudji, L.; Huille, S.; Tribet, C. *Soft Matter* **2012**, *8*, 2651–2661.
71. Bee, J. S.; Stevenson, J. L.; Mehta, B.; Svitel, J.; Pollastrini, J.; Platz, R.; Freund, E.; Carpenter, J. F.; Randolph, T. W. *Biotechnol. Bioeng.* **2009**, *103*, 936–943.
72. Patapoff, T. W.; Esue, O. *Pharm. Dev. Technol.* **2009**, *14*, 659–664.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

73. Lee, H. J.; McAuley, A.; Schilke, K. F.; McGuire, J. *Adv. Drug. Delivery Rev.* **2011**, *63*, 1160–1171.
74. Mahler, H. C.; Senner, F.; Maeder, K.; Mueller, R. *J. Pharm. Sci.* **2009**, *98*, 4525–4533.
75. Zhou, R.; Schlam, R. F.; Yin, S.; Gandhi, R. B.; Adams, M. L. *J. Pharm. Sci.* **2011**, *100*, 1605–1606.
76. Lam, X. M.; Yang, J. Y.; Cleland, J. L. *J. Pharm. Sci.* **1997**, *86*, 1250–1255.
77. Chen, B.; Bautista, R.; Yu, K.; Zapata, G. A.; Mulkerrin, M. G.; Chamow, S. M. *Pharm. Res.* **2003**, *20*, 1952–1960.
78. Ouellette, D.; Alessandri, L.; Piparia, R.; Aikhoje, A.; Chin, A.; Radziejewski, C.; Correia, I. *Anal. Biochem.* **2009**, *389*, 107–117.
79. Salinas, B. A.; Sathish, H. A.; Shah, A. U.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2010**, *99*, 2962–2974.
80. Zhou, S.; Evans, B.; Schoneich, C.; Singh, S. K. *AAPS PharmSciTech* **2012**, *13*, 284–294.
81. Zhou, S.; Schoneich, C.; Singh, S. K. *AAPS PharmSciTech* **2011**, *12*, 411–421.
82. Liu, W.; Swift, R.; Torraca, G.; Nashed-Samuel, Y.; Wen, Z. Q.; Jiang, Y.; Vance, A.; Mire-Sluis, A.; Freund, E.; Davis, J.; Narhi, L. *PDA J. Pharm. Sci. Technol.* **2010**, *64*, 11–19.
83. Bee, J. S.; Nelson, S. A.; Freund, E.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2009**, *98*, 3290–3301.
84. Jiang, Y.; Nashed-Samuel, Y.; Li, C.; Liu, W.; Pollastrini, J.; Mallard, D.; Wen, Z. Q.; Fujimori, K.; Pallitto, M.; Donahue, L.; Chu, G.; Torraca, G.; Vance, A.; Mire-Sluis, T.; Freund, E.; Davis, J.; Narhi, L. *J. Pharm. Sci.* **2009**, *98*, 4695–4710.
85. Wang, W.; Roberts, C. J. *AAPS J.* **2013**, *15*, 840–851.
86. Remmele, R. L., Jr.; Callahan, W. J.; Krishnan, S.; Zhou, L.; Bondarenko, P. V.; Nichols, A. C.; Kleemann, G. R.; Pipes, G. D.; Park, S.; Fodor, S.; Kras, E.; Brems, D. N. *J. Pharm. Sci.* **2006**, *95*, 126–145.
87. Remmele, R. L., Jr.; Zhang-van Enk, J.; Dharmavaram, V.; Balaban, D.; Durst, M.; Shoshitaishvili, A.; Rand, H. *J. Am. Chem. Soc.* **2005**, *127*, 8328–8339.
88. Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859–1877.
89. Privalov, P. L. *Methods Mol. Biol.* **2009**, *490*, 1–39.
90. Kayser, V.; Chennamsetty, N.; Voynov, V.; Helk, B.; Forrer, K.; Trout, B. L. *J. Pharm. Sci.* **2011**, *100*, 2526–2542.
91. Brummitt, R. K.; Nesta, D. P.; Chang, L.; Chase, S. F.; Laue, T. M.; Roberts, C. J. *J. Pharm. Sci.* **2011**, *100*, 2087–2103.
92. Bhatnagar, B. S.; Bogner, R. H.; Pikal, M. J. *Pharm. Dev. Technol.* **2007**, *12*, 505–523.
93. Gomez, G.; Pikal, M. J.; Rodriguez-Hornedo, N. *Pharm. Res.* **2001**, *18*, 90–97.
94. Lee, J. C.; Timasheff, S. N. *J. Biol. Chem.* **1981**, *256*, 7193–7201.
95. Piedmonte, D. M.; Summers, C.; McAuley, A.; Karamujic, L.; Ratnaswamy, G. *Pharm. Res.* **2007**, *24*, 136–146.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

96. Chang, B. S.; Kendrick, B. S.; Carpenter, J. F. *J. Pharm. Sci.* **1996**, *85*, 1325–1330.
97. Strambini, G. B.; Gabellieri, E. *Biophys. J.* **1996**, *70*, 971–976.
98. Strambini, G. B.; Gonnelli, M. *Biophys. J.* **2007**, *92*, 2131–2138.
99. Collins, K. D.; Neilson, G. W.; Enderby, J. E. *Biophys. Chem.* **2007**, *128*, 95–104.
100. Zhang, Y.; Cremer, P. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 658–663.
101. Zhang, Y.; Cremer, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15249–15253.
102. Schwierz, N.; Horinek, D.; Netz, R. R. *Langmuir* **2013**, *29*, 2602–2614.
103. Young, L.; Jernigan, R. L.; Covell, D. G. *Protein Sci.* **1994**, *3*, 717–729.
104. Remmele, R. L., Jr.; Zhang-van Enk, J.; Phan, D.; Yu, L. *J. Phys. Chem. B* **2012**, *116*, 7240–7251.
105. Fowkes, F. M. *Ind. Eng. Chem.* **1964**, *56*, 40–52.
106. Nieba, L.; Honegger, A.; Krebber, C.; Pluckthun, A. *Protein Eng.* **1997**, *10*, 435–444.
107. Donaldson, S. H., Jr.; Lee, C. T., Jr.; Chmelka, B. F.; Israelachvili, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15699–15704.
108. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11937–11942.
109. Voynov, V.; Chennamsetty, N.; Kayser, V.; Helk, B.; Trout, B. L. *mAbs* **2009**, *1*, 580–582.
110. Vermeer, A. W.; Giacomelli, C. E.; Norde, W. *Biochim. Biophys. Acta* **2001**, *1526*, 61–69.
111. Spiegelberg, H. L.; Abel, C. A.; Fishkin, B. G.; Grey, H. M. *Biochemistry* **1970**, *9*, 4217–4223.
112. Kayser, V.; Chennamsetty, N.; Voynov, V.; Forrer, K.; Helk, B.; Trout, B. L. *Biotechnol. J.* **2011**, *6*, 38–44.
113. Li, C. H.; Narhi, L. O.; Wen, J.; Dimitrova, M.; Wen, Z. Q.; Li, J.; Pollastrini, J.; Nguyen, X.; Tsuruda, T.; Jiang, Y. *Biochemistry* **2012**, *51*, 10056–10065.
114. Latypov, R. F.; Hogan, S.; Lau, H.; Gadgil, H.; Liu, D. *J. Biol. Chem.* **2012**, *287*, 1381–1396.
115. Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. *Annu. Rev. Immunol.* **2007**, *25*, 21–50.
116. Wang, X.; Kumar, S.; Buck, P. M.; Singh, S. K. *Proteins* **2013**, *81*, 443–460.
117. Sinclair, A. M.; Elliott, S. *J. Pharm. Sci.* **2005**, *94*, 1626–1635.
118. Hristodorov, D.; Fischer, R.; Joerissen, H.; Muller-Tiemann, B.; Apeler, H.; Linden, L. *Mol. Biotechnol.* **2013**, *53*, 326–335.
119. Schaefer, J. V.; Pluckthun, A. *J. Mol. Biol.* **2012**, *417*, 309–335.
120. Thakkar, S. V.; Kim, J. H.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Joshi, S. B.; Volkin, D. B.; Middaugh, C. R. *J. Pharm. Sci.* **2012**, *101*, 4444–4457.
121. Lowe, D.; Dudgeon, K.; Rouet, R.; Schofield, P.; Jermutus, L.; Christ, D. *Adv. Protein Chem. Struct. Biol.* **2011**, *84*, 41–61.
122. Ryazantsev, S.; Tischenko, V.; Nguyen, C.; Abramov, V.; Zav'yalov, V. *PLoS One* **2013**, *8*, e64076.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

123. Padlan, E. A. *Mol. Immunol.* **1994**, *31*, 169–217.
124. Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. *J. Pharm. Sci.* **2007**, *96*, 1–26.
125. Czajkowsky, D. M.; Hu, J.; Shao, Z.; Pleass, R. J. *EMBO Mol. Med.* **2012**, *4*, 1015–1028.
126. Franey, H.; Brych, S. R.; Kolvenbach, C. G.; Rajan, R. S. *Protein Sci.* **2010**, *19*, 1601–1615.
127. Pepinsky, R. B.; Silvian, L.; Berkowitz, S. A.; Farrington, G.; Lugovskoy, A.; Walus, L.; Eldredge, J.; Capili, A.; Mi, S.; Graff, C.; Garber, E. *Protein Sci.* **2010**, *19*, 954–966.
128. van der Neut Kolfschoten, M.; Schuurman, J.; Losen, M.; Bleeker, W. K.; Martinez-Martinez, P.; Vermeulen, E.; den Bleker, T. H.; Wiegman, L.; Vink, T.; Aarden, L. A.; De Baets, M. H.; van de Winkel, J. G.; Aalberse, R. C.; Parren, P. W. *Science* **2007**, *317*, 1554–1557.
129. Rispens, T.; Davies, A. M.; Ooijevaar-de Heer, P.; Absalah, S.; Bende, O.; Sutton, B. J.; Vidarsson, G.; Aalberse, R. C. *J. Biol. Chem.* **2014**, *289*, 6098–6109.
130. Teerinen, T.; Valjakka, J.; Rouvinen, J.; Takkinen, K. *J. Mol. Biol.* **2006**, *361*, 687–697.
131. Perchiacca, J. M.; Bhattacharya, M.; Tessier, P. M. *Proteins* **2011**, *79*, 2637–2647.
132. Dudgeon, K.; Rouet, R.; Kokmeijer, I.; Schofield, P.; Stolp, J.; Langley, D.; Stock, D.; Christ, D. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10879–10884.
133. Wu, S. J.; Luo, J.; O'Neil, K. T.; Kang, J.; Lacy, E. R.; Canziani, G.; Baker, A.; Huang, M.; Tang, Q. M.; Raju, T. S.; Jacobs, S. A.; Teplyakov, A.; Gilliland, G. L.; Feng, Y. *Protein Eng. Des. Sel.* **2010**, *23*, 643–651.
134. Rosenberg, A. S. *AAPS J.* **2006**, *8*, E501–507.
135. Barth, H. G.; Boyes, B. E.; Jackson, C. *Anal. Chem.* **1994**, *66*, 595R–620R.
136. Sahin, E.; Roberts, C. J. *Methods Mol. Biol.* **2012**, *899*, 403–423.
137. Kok, C. M.; Rudin, A. *Die Makromolekulare Chemie, Rapid Communications* **1981**, *2*, 655–659.
138. Krayukhina, E.; Uchiyama, S.; Nojima, K.; Okada, Y.; Hamaguchi, I.; Fukui, K. *J. Biosci. Bioeng.* **2013**, *115*, 104–110.
139. Pekar, A.; Sukumar, M. *Anal. Biochem.* **2007**, *367*, 225–237.
140. Philo, J. S. *Curr. Pharm. Biotechnol.* **2009**, *10*, 359–372.
141. Schuck, P.; Perugini, M. A.; Gonzales, N. R.; Howlett, G. J.; Schubert, D. *Biophys. J.* **2002**, *82*, 1096–1111.
142. Kelly, S. M.; Price, N. C. *Curr. Protein Pept. Sci.* **2000**, *1*, 349–384.
143. Van Buren, N.; Rehder, D.; Gadgil, H.; Matsumura, M.; Jacob, J. *J. Pharm. Sci.* **2009**, *98*, 3013–3030.
144. Pelton, J. T.; McLean, L. R. *Anal. Biochem.* **2000**, *277*, 167–176.
145. Dong, A.; Prestrelski, S. J.; Allison, S. D.; Carpenter, J. F. *J. Pharm. Sci.* **1995**, *84*, 415–424.
146. Andya, J. D.; Hsu, C. C.; Shire, S. J. *AAPS PharmSci* **2003**, *5*, E10.
147. Muino, P. L.; Callis, P. R. *J. Phys. Chem. B* **2009**, *113*, 2572–2577.
148. Kozachkov, L.; Padan, E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15769–15774.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

149. Weichel, M.; Bassarab, S.; Garidel, P. *Bioprocess Int.* **2008**, 42–52.
150. Correia, I.; Sung, J.; Burton, R.; Jakob, C. G.; Carragher, B.; Ghayur, T.; Radziejewski, C. *mAbs* **2013**, *5*, 364–372.
151. Roux, K. H. *Int. Arch. Allergy Immunol.* **1999**, *120*, 85–99.
152. Yang, J.; Mayer, M.; Kriebel, J. K.; Garstecki, P.; Whitesides, G. M. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 1555–1558.
153. Ewert, S.; Huber, T.; Honegger, A.; Pluckthun, A. *J. Mol. Biol.* **2003**, *325*, 531–553.
154. Iwura, T.; Fukuda, J.; Yamazaki, K.; Kanamaru, S.; Arisaka, F. *J. Biochem.* **2014**, *155*, 63–71.
155. Luo, Y.; Lu, Z.; Raso, S. W.; Entrican, C.; Tangarone, B. *mAbs* **2009**, *1*, 491–504.
156. Paul, R.; Graff-Meyer, A.; Stahlberg, H.; Lauer, M. E.; Rufer, A. C.; Beck, H.; Briguet, A.; Schnaible, V.; Buckel, T.; Boeckle, S. *Pharm. Res.* **2012**, *29*, 2047–2059.
157. Campbell, M. G.; Veesler, D.; Cheng, A.; Potter, C. S.; Carragher, B., *Elife* **2015** *11*, 4.
158. Sung, J. J.; Pardeshi, N. N.; Mulder, A. M.; Mulligan, S. K.; Quispe, J.; On, K.; Carragher, B.; Potter, C. S.; Carpenter, J. F.; Schneemann, A. *J. Pharm. Sci.* **2015**, *104*, 750–759.
159. Zhang, Y.; Sheng, S.; Shao, Z. *Biophys. J.* **1996**, *71*, 2168–2176.
160. Lee, H.; Kirchmeier, M.; Mach, H. *J. Pharm. Sci.* **2011**, *100*, 416–423.
161. Benesch, J. L.; Robinson, C. V. *Curr. Opin. Struct. Biol.* **2006**, *16*, 245–251.
162. Hanson, C. L.; Videler, H.; Santos, C.; Ballesta, J. P.; Robinson, C. V. *J. Biol. Chem.* **2004**, *279*, 42750–42757.
163. Marcoux, J.; Robinson, C. V. *Structure* **2013**, *21*, 1541–1550.
164. Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J. *Nat. Chem.* **2011**, *3*, 126–132.
165. Rose, R. J.; van Berkel, P. H.; van den Bremer, E. T.; Labrijn, A. F.; Vink, T.; Schuurman, J.; Heck, A. J.; Parren, P. W. *mAbs* **2013**, *5*, 219–228.
166. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T. *Nat. Chem.* **2011**, *3*, 172–177.
167. Young, L. M.; Cao, P.; Raleigh, D. P.; Ashcroft, A. E.; Radford, S. E. *J. Am. Chem. Soc.* **2014**, *136*, 660–670.
168. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134*, 3429–3438.
169. Hilton, G. R.; Thalassinos, K.; Grabenauer, M.; Sanghera, N.; Slade, S. E.; Wyttenbach, T.; Robinson, P. J.; Pinheiro, T. J.; Bowers, M. T.; Scrivens, J. H. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 845–854.
170. Hall, Z.; Hernandez, H.; Marsh, J. A.; Teichmann, S. A.; Robinson, C. V. *Structure* **2013**, *21*, 1325–1337.
171. Wysocki, V. H.; Jones, C. M.; Galhena, A. S.; Blackwell, A. E. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 903–913.
172. Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 343–350.
173. Perkins, S. J.; Okemefuna, A. I.; Fernando, A. N.; Bonner, A.; Gilbert, H. E.; Furtado, P. B. *Methods Cell Biol.* **2008**, *84*, 375–423.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

174. Svergun, D. I.; Koch, M. H. J. *Rep. Prog. Phys.* **2003**, *66*, 1735–1782.
175. Lilyestrom, W. G.; Yadav, S.; Shire, S. J.; Scherer, T. M. *J. Phys. Chem. B* **2013**, *117*, 6373–6384.
176. Mosbaek, C. R.; Konarev, P. V.; Svergun, D. I.; Rischel, C.; Vestergaard, B. *Pharm. Res.* **2012**, *29*, 2225–2235.
177. Lilyestrom, W. G.; Shire, S. J.; Scherer, T. M. *J. Phys. Chem. B* **2012**, *116*, 9611–9618.
178. Yearley, E. J.; Zarraga, I. E.; Shire, S. J.; Scherer, T. M.; Gokarn, Y.; Wagner, N. J.; Liu, Y. *Biophys. J.* **2013**, *105*, 720–731.
179. Kriwacki, R. W.; Siuzdak, G. *Methods Mol. Biol.* **2000**, *146*, 223–238.
180. Monti, M.; Amoresano, A.; Giorgetti, S.; Bellotti, V.; Pucci, P. *Biochim. Biophys. Acta* **2005**, *1753*, 44–50.
181. Serpa, J. J.; Parker, C. E.; Petrotchenko, E. V.; Han, J.; Pan, J.; Borchers, C. H. *Eur. J. Mass Spectrom.(Chichester, England)* **2012**, *18*, 251–267.
182. Zhao, A.; Hao, G.; Gu, J. *J. Pharm. Biomed. Anal.* **2013**, *73*, 99–102.
183. Bich, C.; Maedler, S.; Chiesa, K.; DeGiacomo, F.; Bogliotti, N.; Zenobi, R. *Anal. Chem.* **2010**, *82*, 172–179.
184. Chen, F.; Nielsen, S.; Zenobi, R. *J. Mass Spectrom.* **2013**, *48*, 807–812.
185. Madler, S.; Seitz, M.; Robinson, J.; Zenobi, R. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1775–1783.
186. Cubrilovic, D.; Haap, W.; Barylyuk, K.; Ruf, A.; Badertscher, M.; Gubler, M.; Tetaz, T.; Joseph, C.; Benz, J.; Zenobi, R. *ACS Chem. Biol.* **2014**, *9*, 218–226.
187. Gasilova, N.; Nazabal, A. *Methods Mol. Biol.* **2012**, *803*, 219–229.
188. Houde, D.; Peng, Y.; Berkowitz, S. A.; Engen, J. R. *Mol. Cell. Proteomics* **2010**, *9*, 1716–1728.
189. Manikwar, P.; Majumdar, R.; Hickey, J. M.; Thakkar, S. V.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. *J. Pharm. Sci.* **2013**, *102*, 2136–2151.
190. Nasr, M. L.; Shi, X.; Bowman, A. L.; Johnson, M.; Zvonok, N.; Janero, D. R.; Vemuri, V. K.; Wales, T. E.; Engen, J. R.; Makriyannis, A. *Protein Sci.* **2013**, *22*, 774–787.
191. Phillips, J. J.; Yao, Z. P.; Zhang, W.; McLaughlin, S.; Laue, E. D.; Robinson, C. V.; Jackson, S. E. *J. Mol. Biol.* **2007**, *372*, 1189–1203.
192. Snijder, J.; Uetrecht, C.; Rose, R. J.; Sanchez-Eugenia, R.; Marti, G. A.; Agirre, J.; Guerin, D. M.; Wuite, G. J.; Heck, A. J.; Roos, W. H. *Nat. Chem.* **2013**, *5*, 502–509.
193. Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158–170.
194. Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R. *Drug Discovery Today* **2014**, *19*, 95–102.
195. Clarke, J.; Itzhaki, L. S. *Curr. Opin. Struct. Biol.* **1998**, *8*, 112–118.
196. Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. *Protein Sci.* **1997**, *6*, 1101–1109.
197. Iacob, R. E.; Bou-Assaf, G. M.; Makowski, L.; Engen, J. R.; Berkowitz, S. A.; Houde, D. *J. Pharm. Sci.* **2013**, *102*, 4315–4329.
198. Kheterpal, I.; Lashuel, H. A.; Hartley, D. M.; Walz, T.; Lansbury, P. T., Jr.; Wetzel, R. *Biochemistry* **2003**, *42*, 14092–14098.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

199. Qi, W.; Zhang, A.; Patel, D.; Lee, S.; Harrington, J. L.; Zhao, L.; Schaefer, D.; Good, T. A.; Fernandez, E. J. *Biotechnol. Bioeng.* **2008**, *100*, 1214–1227.
200. Singh, J.; Udgaonkar, J. B. *J. Mol. Biol.* **2013**, *425*, 3510–3521.
201. Wang, S. S.; Tobler, S. A.; Good, T. A.; Fernandez, E. J. *Biochemistry* **2003**, *42*, 9507–9514.
202. Zhang, A.; Singh, S. K.; Shirts, M. R.; Kumar, S.; Fernandez, E. J. *Pharm. Res.* **2012**, *29*, 236–250.
203. Nazabal, A.; Dos Reis, S.; Bonneu, M.; Saupe, S. J.; Schmitter, J. M. *Biochemistry* **2003**, *42*, 8852–8861.
204. Zhang, A.; Jordan, J. L.; Ivanova, M. I.; Weiss, W. F. t.; Roberts, C. J.; Fernandez, E. J. *Biochemistry* **2010**, *49*, 10553–10564.
205. Zhang, A.; Qi, W.; Good, T. A.; Fernandez, E. J. *Biophys. J.* **2009**, *96*, 1091–1104.
206. Majumdar, R.; Manikwar, P.; Hickey, J. M.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. *Biochemistry* **2013**, *52*, 3376–3389.
207. Ramm, K.; Gehrig, P.; Pluckthun, A. *J. Mol. Biol.* **1999**, *290*, 535–546.
208. Page, R.; Peti, W.; Wilson, I. A.; Stevens, R. C.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1901–1905.
209. Sugiki, T.; Yoshiura, C.; Kofuku, Y.; Ueda, T.; Shimada, I.; Takahashi, H. *Protein Sci* **2009**, *18*, 1115–1120.
210. Liu, D.; Cocco, M. J.; Rosenfied, R.; Lewis, J. K.; Ren, D.; Li, L.; Remmele, R. L., Jr.; Brems, D. N. *Biomol. NMR Assignments* **2007**, *1*, 233–235.
211. Liu, D.; Cocco, M.; Matsumura, M.; Ren, D.; Becker, B.; Remmele, R. L., Jr.; Brems, D. N. *Biomol. NMR Assignments* **2007**, *1*, 93–94.
212. Liu, D.; Ren, D.; Huang, H.; Dankberg, J.; Rosenfeld, R.; Cocco, M. J.; Li, L.; Brems, D. N.; Remmele, R. L., Jr. *Biochemistry* **2008**, *47*, 5088–5100.
213. Burkitt, W.; Domann, P.; O'Connor, G. *Protein Sci.* **2010**, *19*, 826–835.
214. Tartaglia, G. G.; Vendruscolo, M. *Chem. Soc. Rev.* **2008**, *37*, 1395–1401.
215. Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat. Biotechnol.* **2004**, *22*, 1302–1306.
216. Trovato, A.; Chiti, F.; Maritan, A.; Seno, F. *PLoS Comput Biol* **2006**, *2*, e170.
217. Tsolis, A. C.; Papandreou, N. C.; Iconomidou, V. A.; Hamodrakas, S. J. *PLoS One* **2013**, *8*, e54175.
218. Buck, P. M.; Kumar, S.; Wang, X.; Agrawal, N. J.; Trout, B. L.; Singh, S. K. *Methods Mol. Biol.* **2012**, *899*, 425–451.
219. Wang, X.; Das, T. K.; Singh, S. K.; Kumar, S. *mAbs* **2009**, *1*, 254–267.
220. Wang, X.; Singh, S. K.; Kumar, S. *Pharm. Res.* **2010**, *27*, 1512–1529.
221. Chennamsetty, N.; Helk, B.; Voynov, V.; Kayser, V.; Trout, B. L. *J. Mol. Biol.* **2009**, *391*, 404–413.
222. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *J. Phys. Chem. B* **2010**, *114*, 6614–6624.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch005

Chapter 6

Simultaneous Multiple Sample Light Scattering (SMSLS) for Continuous Monitoring of Protein Aggregation

Michael F. Drenski,[1] Mark L. Brader,[2] and Wayne F. Reed[*,3]

[1]Advanced Polymer Monitoring Technologies, Inc., New Orleans, Louisiana 70125, United States
[2]Biogen Idec, Protein Pharmaceutical Development, 14 Cambridge Center, Cambridge, Massachusetts 02142, United States
[3]Tulane University, New Orleans, Louisiana 70115, United States
[*]E-mail: wreed@tulane.edu

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Simultaneous multiple sample light scattering (SMSLS) is a high-throughput technology that is essentially many total intensity light scattering devices in a single instrument. As such, it is aimed at making both absolute light scattering measurements of protein molecular weight and virial coefficients in solution under equilibrium conditions, as well as continuously monitoring molecular weight changes, in real time, during aggregation, degradation, or other time-dependent processes. The current prototype has 16 cells in which measurements on different samples can be made simultaneously and completely independently of each other, allowing samples to be changed in and out without disturbing other ongoing experiments. Thus, samples can be subjected independently to different stressors, such as temperature steps and ramps and controlled stepper motor-based stirring, and titrations can be performed during experiments. Depolarized scattering, which signals shape anisotropy, also is monitored. Results from the NISTmAb and three industrial monoclonal antibodies are presented below in terms of equilibrium properties, and then thermal instability; instability against stirring, including effects of the air/liquid interface; and detection of the onset and evolution of particulates are investigated.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Background

The intensity of light scattered from proteins in solution is exquisitely sensitive to the molecular weight of the protein. With sensitive, low-noise light scattering instrumentation built from readily available commercial materials it is possible to detect variations of molecular mass of less than 1% (e.g., a single protein dimer amidst 100 monomers) (*1*).

Although such good quality commercial instrumentation for single samples exists from several sources, SMSLS expands this capacity to many simultaneous, independent samples in an integrated instrument under the control of a single computer. Hence, SMSLS is the equivalent of numerous single-sample light scattering instruments packaged into a single unit, with the additional ability to manipulate samples during experiments and provide a wide variety of stressors and perturbations independently to each sample.

One of the central tenets of the authors in developing SMSLS is that protein aggregation is a kinetic process. As such, the direct, quantitative monitoring of the process itself provides the most compelling means of interrogating the process and its sensitivity to factors such as formulation and protein sequence variants and stressors such as temperature, stirring, shear stress, air/liquid interfaces, different surfaces, ions, and so forth. Because almost all aspects of commercial therapeutic protein production, distribution, and use are fraught with a variety of processes that induce protein aggregation, new versatile methodologies capable of probing multiple stressors are needed.

There is precedent in the literature for use of time-dependent total intensity light scattering (LS) measurements (*2*). Sokolowski et al. measured continuous equivalent weight average molecular weight (M_w) versus time for the aggregation of prion proteins on a custom-built LS system (*3*). Intermittent measurement of both LS and dynamic light scattering (DLS) were made for the aggregation of tobacco mosaic virus coat protein (*4*). A similar intermittent measurement was made for aggregation of insulin and alcohol dehydrogenase (*5*). Use of SMSLS, which is a total intensity LS method, for monitoring protein aggregation was recently published (*1*).

Total intensity LS is sometimes also referred to as "static LS," or SLS, and often termed MALS (multi-angle light scattering) when multiple detection angles are used. It is most frequently used for determining molecular weight, virial coefficients, and radius of gyration of polymers in solution, and also can be attached to a size exclusion chromatography (SEC) system. In contrast, DLS autocorrelates fluctuations in scattered light intensity (e.g., those due to Brownian motion of analytes in solution) in order to find particle diffusion coefficients and the first hydrodynamic interaction term K_D. Models, such as the Stokes-Einstein model for spheres, then frequently are used to interpret the diffusion coefficients in terms of equivalent spherical hydrodynamic diameters. DLS is inherently less sensitive for characterizing spheroidal entities, such as globular proteins, since the diffusion coefficient varies as $1/(\text{molar mass})^{1/3}$, whereas scattering intensity is directly proportional to molar mass.

Methods that do not use chromatographic columns for assessing aggregation include analytical ultracentrifugation, DLS, fluorescence, and field flow

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

fractionation (*6*, *7*). SEC remains the routine method for quantitative aggregation analysis in the biopharmaceutical industry, and it often is coupled with a SLS detector (*8*) to provide an in-line molecular weight characterization of the individual chromatographic peaks. Other methods also have been reviewed (*9*, *10*).

SMSLS Instrumentation

SMSLS Hardware

Several different SMSLS prototypes have been built and reported on by this group, including versions with batch and flow cells (*1*, *11–13*). The latest prototype system is built by Advanced Polymer Monitoring Technologies, Inc. (New Orleans, a Tulane University spin-off company), shown in Figure 1. It consists of 16 fully independent sample holders, each with the ability to fit one sample cell ranging in path length from 3 mm to 10 mm. The variation in path length allows the user to use more or less sample when necessary for either conserving product with small volumes or for stirring samples with larger volumes. Incident on each sample is a vertically polarized 660 nm, 35 mW miniature diode laser for SLS measurements. The light scattered from the solutions within each sample cell is continuously monitored in the scattering plane at 90° to the laser beam polarization direction and fiber-optically channeled to an allocated series of pixels on a highly sensitive, linear array charge-coupled device (CCD) camera. The power of each laser incident on the sample can be specifically tuned by automatically rotating an absorptive neutral density filter (NDF) into the laser beam path. Miniature servo motors are used to achieve the specific positioning of the appropriate NDF on each laser independently. In addition to tuning the laser power, the exposure time of the CCD camera can be adjusted as a whole to optimize the sensitivity of the system appropriate to the sample scattered light being monitored.

All 16 sample holders have temperature control from ambient up to 180 °C and potentially higher. Thermoelectric Peltier devices have been affixed to 8 of the 16 sample cells that allow temperature control below ambient to about 12 °C, before condensation sets in on the sample cells walls. Lower temperatures can be achieved by purging the system with moisture-free instrument air. Samples can cycle through freeze/thaw cycles in the Peltier-equipped sample cells, and their behavior monitored in the thawed state on each cycle, without removing samples from the instrument during cycling.

Of the 16 sample holders, 12 are equipped with fully controllable stepper motors. When a small, polytetrafluoroethylene (PTFE)-encapsulated magnetic stir bar is inserted into the sample cell, the magnetic field couples with that of another magnet attached to the rotating drive shaft of the stepper motor, thus allowing controlled mixing from 1 revolution per minute (RPM) to 2000 RPM. Rather than a stepper motor, the other 4 sample holders are fitted with a second fiber optic that is 90° to the scattering plane to monitor depolarized scattering, including its onset and evolution in systems where particle anisotropy changes in time (e.g., the

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

fibrillar amyloid aggregation associated with Alzheimer's and related diseases) (*14*).

Modularity and expansion is key in the design of the SMSLS system. Each sample holder has been developed to fit the same footprint within the system so that different versions of the sample holder can be selected per the user's most common needs. In addition to the already-implemented vertical and depolarized scattering, as well as controlled temperature and stirring, developments are underway for low-angle and multi-angle scattering, DLS, and fluorescence detection.

Sensitivity and Minimum Volume

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Instrumental sensitivity is an important aspect for applicability to the analysis of aggregates and particles in biopharmaceutical products. The sensitivity of the SMSLS for detecting aggregation is purely a function of the signal/noise ratio of the LS intensity baseline when unaggregated protein is monitored. For a given signal/noise ratio, detection of minimum aggregation depends on the molar mass and concentration of the native protein before aggregation begins. For example, at 0.005 g/cm^3 for a 200,000 g/mol protein, Drenski et al. compared scattering from aggregate-free protein solution and reported the detection of aggregates via light scattering changes when 3 dimers were present among 1,000 unaggregated proteins at the 3σ (three standard deviation) confidence level, and when 6 dimers were present among 1,000 unaggregated proteins at the 6σ level (*1*).

Although subvisible particles usually do not constitute a significant enough mass fraction to be quantified by mass (e.g., directly by a secondary aggregate peak in SEC or by loss of SEC main peak area), SMSLS is capable of counting them. Trace levels (< 0.01% by mass) of protein aggregates/particles may be sufficient to adversely impact product quality.

Furthermore, in the early stages of biopharmaceutical development, material quantity is limited as efforts focus on protein sequence, candidate selection, and developability screening. For these applications, highly sensitive techniques that consume small protein quantities become especially important. The 3 mm SMSLS cell can use as little as 26 µl of sample. For a 0.1 mg/ml solution, for which aggregation kinetics are easily measurable, 2.6 µg of protein suffices for quantitative determination of aggregation kinetics. The controlled stir feature can be used with a 4 mm cell and a minimum volume of 65 µl.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

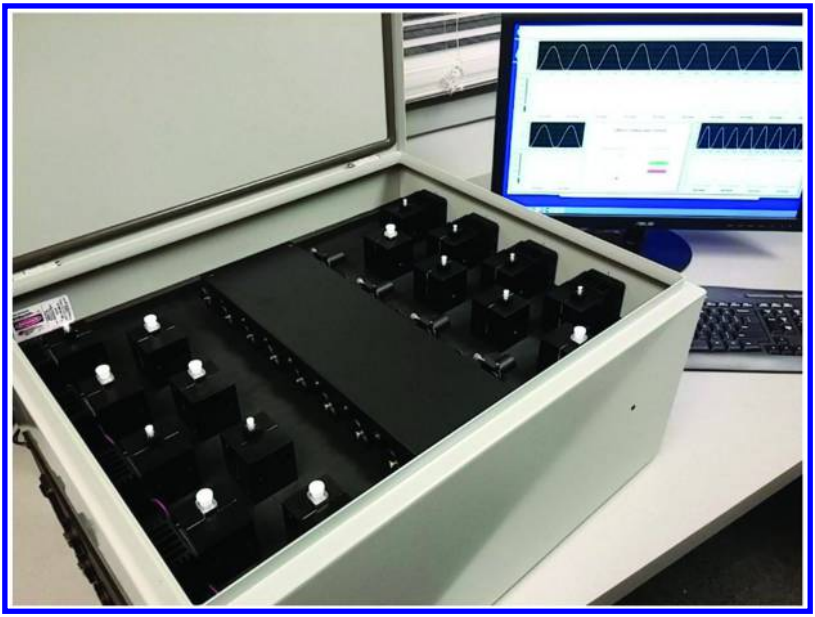

Figure 1. Picture of 16-cell simultaneous multiple sample light scattering (SMSLS) system (Advanced Polymer Monitoring Technologies, Inc., New Orleans).

SMSLS Software

As with the SMSLS hardware, the software has been developed with modularity and expansion in mind. Each function of the SMSLS system is controlled through independent modules that report their status to the data logging module for each sample. Therefore, any change to a particular cell—be it temperature, stirring speed, or laser intensity—does not have any effect on the data logging for any other sample cell within the system. This independence of each sample cell allows the user to remove or change some samples when necessary while leaving others within the system to continue monitoring, undisturbed.

The logging module allows the user to digitally track, record, and recover all pertinent information about the sample in each cell that is used for analysis calculations. Additionally, the logging software has a section for notes to be input to the header of each file so that the user can input comments that are important to the knowledge capture of each sample. The logging capabilities of this software have been carefully conceived to allow for the development of a back-end databasing and analysis software platform. This database and analysis software allows the users to search archives and automatically generate meaningful analysis of the raw data.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Modes of Operation

Equilibrium Measurements

The SMSLS instrument can be used for both equilibrium and non-equilibrium measurements. Similar to single-sample SLS systems with absolute calibration, equilibrium measurements by SMSLS provide M_w, and virial coefficients A_2, A_3, A_4, for which results are discussed below. For equilibrium measurements, it is most advantageous to use a single cell and make serial dilutions off of a concentrated monoclonal antibody (mAb) stock, a technique that also can be used in single-sample SLS instruments. A solvent with known Rayleigh scattering ratio, such as toluene, or a molecular weight standard can be used to calibrate the system. Typically, 50 to 80 µl of concentrated mAb can be used in a 4 mm cell, and then diluted multiple times. Care must be taken to filter "dust" and other large scatterers from the solution, usually with a 0.22 µm, narrow diameter membrane filter, before the experiments, and dilutions should be made by filtering clean buffer directly into the scattering cell. A complete equilibrium determination, starting with a high concentration of 0.100 g/cm^3 mAb stock, can be made with 10 mg or less of protein. For a rapid M_w determination, a single measurement at low concentration (e.g., 1 mg/ml), where virial coefficient effects for most mAb are negligible, can suffice to obtain a good estimate of M_w. This can be advantageous when very little sample is available, and such a measurement could be done with as little as 0.020 mg of mAb.

Monitoring Kinetic Processes

The SMSLS instrument can be used to monitor changes in polymer or colloid solutions, such as aggregation, micellization, phase changes, dissolution, micro-crystallization, and so forth. Each cell is fully independent of the others so that samples can be introduced into and removed from the instrument at any time without affecting measurements ongoing for other samples. Selected stressors, such as temperature ramps, controlled stirring, titration of samples during measurements, addition of gases, use of different surfaces, and so forth, can be made individually to each cell. Users of the current SMSLS leave it running for months at a time, allowing continuous monitoring of many frequently changing samples.

Materials

Four proteins were used in this study: the mAb supplied by NIST (NISTmAb); and three mAbs supplied from pharmaceutical companies, mAb A (mAbA), mAb B (mAbB), and mAb C (mAbC). For the NISTmAb, a histidine buffer adjusted to pH 6.0 was used. Highly purified mAbs, mAbA and mAbB, were provided by Biogen Idec in formulation buffers specific to each molecule. These were used as supplied with dilution with corresponding placebo to adjust concentration where necessary.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Equilibrium Characterization

Protein solutions in equilibrium can be characterized by the usual approaches used for total intensity light scattering to yield M_w, and the second virial coefficient A_2, as well as possibly the higher virial coefficients (i.e., A_3, A_4, and so forth). For particles of size larger than about 10 nm, the z-averaged mean square radius of gyration $\langle S^2 \rangle_z$ also can be measured. These measurements can be made on traditional single-sample light scattering instruments, but the SMSLS instrument offers the advantage of very low sample consumption, as outlined above.

Most common among the approaches is the Zimm equation (equation 2, below) within the Rayleigh-Debye approximation that assumes the optical path length accrued between a scatterer and the detector is due only to the geometric path length. For a particle of characteristic dimension D and index of refraction n_p immersed in a liquid of index n_o, the Rayleigh-Debye approximation is met for incident light of vacuum wavelength λ when the following criterion holds:

$$\left| n_p - n_0 \right| D << \frac{\lambda}{n_0} \tag{1}$$

The well-known Zimm approximation (*15*) allows determination of M_w, second and third virial coefficients A_2 and A_3, respectively, and particle shape factor P(q). Zimm has shown that to second order in concentration, c (g/cm^3), the quantity $Kc/I_R(q,c)$, where $I_R(q,c)$ is the excess Rayleigh scattering ratio, can be approximated for a monodisperse polymer of mass M by:

$$\frac{Kc}{I_R(q,c)} = \frac{1}{MP(q)} + 2A_2 c + \left[3A_3 Q(q) - 4A_2^{\,2} MP(q)(1 - P(q)) \right] c^2 \tag{2}$$

This equation forms the basis of the well-known Zimm equation, which at low concentrations and for $q^2 \langle S^2 \rangle_z << 1$ can be written for a polydisperse polymer population as:

$$\frac{Kc}{I_R(q,c)} = \frac{1}{M_w} \left(1 + \frac{q^2 < S^2 >_z}{3} \right) + 2A_2 c \tag{3}$$

which directly permits determination of M_w, A_2 (a double z-average), and $\langle S^2 \rangle_z$ by measuring $I_R(q,c)$ for a series of protein concentrations, c (g/cm^3), and scattering angles. In this approximation, the measured $\langle S^2 \rangle_z$ is independent of particle morphology (e.g., spheroid, random coil, rod). K is an optical constant, given for vertically polarized incident light by:

$$K = \frac{4\pi^2 n_0^{\,2} (\partial n / \partial c)^2}{N_A \lambda^4} \tag{4}$$

where n_0 is the solvent index of refraction, λ is the vacuum wavelength of the incident light, $\partial n/\partial c$ is the differential refractive index for the polymer in

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

the chosen solvent, and q is the amplitude of the scattering wave-vector. $q = (4\pi n_o/\lambda)\sin(\theta/2)$, where θ is the scattering angle, measured in the scattering plane perpendicular to the vertically polarized incident light.

In the limit of q = 0, or for small particles such that $q^2 < S^2 >_z << 1$, third and fourth virial coefficients, A_3 and A_4, can additionally be determined by:

$$\frac{Kc}{I_R(0,c)} = \frac{1}{M_w} + 2A_2c + 3A_3c^2 + 4A_4c^3 + \ldots \quad (5)$$

Values of I_R from samples were obtained by using toluene as the reference, for which at λ = 660 nm and T = 25 °C, $I_{R,toluene} = 1.183\times10^{-5}$ cm^{-1}.

Although equation 5 directly reveals the effects of A_2 and the higher virial coefficients on the concentration-dependent intensity of scattered light, it is worth recalling the physical significance of these coefficients. A_2 is a measure of the excluded volume between two particles. A hard sphere of radius R is easiest to visualize, because in this case the centers of two spheres cannot approach each other any closer than a distance 2R (the interaction potential is positive and infinite when the two hard spheres touch; i.e., they cannot interpenetrate), so the excluded volume is $4(2R)^3\pi/3$. In the case of a real mAb, the surface interactions between particles can be quite complex and involve areas of repulsive (e.g., electrostatic) and attractive potentials (e.g., "sticky" hydrophobic surface domains). A_2 averages over all possible spatial orientations of the surface and volume (e.g., long-range electrostatic) interactions to produce a single, scalar value. The result is that even if the mAb resembles a sphere, the excluded volume can be less than (i.e., there are attractive potential areas), equal to (the mAb acts as a net hard sphere), or greater than (e.g., long-range electrostatic repulsive interactions from net mAb charge) the actual excluded volume of the equivalent hard mAb spheres. A_2 is very sensitive to factors such as pH (affecting net mAb charge) and ionic strength (controlling interparticle electrostatic shielding) so that any values of A_2 given will be relevant only to the specific solution the mAb is in. A_3 is based on the same excluded volume as A_2 but is the effect of three particles interacting at a time, A_4 involves four-body interactions, and so forth.

Dilutions of the 0.100 g/ml NISTmAb stock were made and measured in the SMSLS device to obtain the Rayleigh ratio data versus concentration in Figure 2a. The downwards concave nonlinearity up to 0.048 g/ml is due chiefly to the strong, positive second virial coefficient A_2. The maximum at approximately 0.048 g/ml shows the combined effect of A_2 and A_3; a positive A_2 alone will cause a plateau to be reached, but a positive A_3 is required to decrease LS intensity with increasing concentration.

Figure 2a also contains Rayleigh ratio data for mAbA and mAbB, as well as equilibrium data, for these are shown along with the NIST protein in Table 1. The comparative A_2 effects are apparent. All three proteins have similar masses, yet widely varying A_2. A_2 for the NISTmAb is 20 times larger than for mAbB, but the NISTmAb, which has the largest A_2, yields LS intensity that falls substantially below the other two proteins and reaches a maximum before mAbA. A_2 for mAbB is so small that I_R versus concentration is almost a straight line. The maximum

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

concentrations of the NIST protein, mAbA, and mAbB are 0.100, 0.190, and 0.040 g/cm^3, respectively.

Figure 2b shows the Debye plot, $Kc/I_R(c)$ versus c, for the NISTmAb data from Figure 2a. The values of M_w, A_2, and A_3 are shown in Figure 2b and in Table 1. M_w, which is listed as "M_0" in Figure 2b to indicate unaggregated protein molar mass, is within 1.2% of the NISTmAb value of 147,000 Da. The value $\partial n/\partial c$ = 0.190 cm^3/g was used, as per the average value from the following National Institutes of Health (NIH) article (*16*) on protein $\partial n/\partial c$ values. The pronounced positive second derivative (upward curvature) shows the strong effect of A_3; if only an A_2 effect were present, the plot would be linear.

The inset to Figure 2b shows the z-averaged inverse apparent hydrodynamic diameter, D_H, versus concentration from independent DLS measurements (90Plus DLS from Brookhaven Instruments Corp., Holtsville, NY). The intercept at c = 0 corresponds to the self-diffusion coefficient of the protein and hence to its D_H. That is, D_H is given by the Stokes-Einstein equation (*17*):

$$D_H = \lim_{c \to 0} \frac{k_B T}{3\pi\eta D(c)} \tag{6}$$

where D(c) is the concentration-dependent mutual diffusion coefficient, k_B is Boltzmann's constant, and η is the solution viscosity.

A_3 also can be determined via best fit to equation 5 or from the concentration at which $I_{R,max}$ is reached, c_m, according to:

$$A_3 = \frac{1}{3M_w c_m^{\,2}} \tag{7}$$

Another approximation is the relationship between A_2 and A_3 found by Boltzmann for hard spheres according to equation 8:

$$A_3 = \frac{5MA_2^{\,2}}{8} \tag{8}$$

A_3 determined by the three methods are seen to be in good agreement for the NISTmAb, seen in Table 1, and, separately for mAbA. mAbB had such a small A_2 that its predicted A_3 in Table 1 is 100 times lower than that of mAbA and 300 times lower than that of the NISTmAb. This extremely low predicted value for A_3 and the fact that mAbB was available only up to 0.040 g/cm^3 made A_3 unmeasurable. With this low predicted A_3, it would require an mAbB concentration of 0.716 g/cm^3 to reach the LS intensity maximum, according to equation 5. On the other hand, with the strong A_3 and high available concentration of mAbA, it was possible to measure up to A_4, shown in Table 1. A_4 was close to the value predicted by Boltzmann for a hard sphere (*1*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

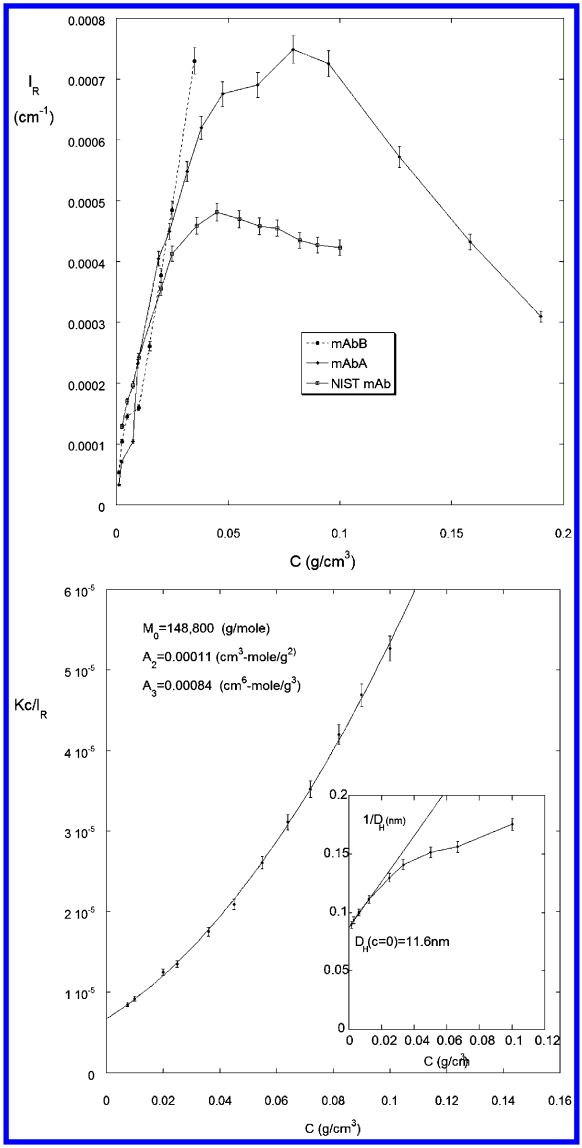

Figure 2. (a) Rayleigh scattering ratio for NISTmAb, monoclonal antibody A (mAbA), and monoclonal antibody B (mAbB). The lines between data points are guides for the eye. (b) Debye plot for the NISTmAb. The inset is the reciprocal hydrodynamic diameter, D_H, versus concentration (c) from dynamic light scattering (DLS) measurements.

Table 1. Equilibrium Properties of Three Proteins Used in This Work

Property	*NISTmAb*	*Monoclonal antibody A (mAbA)*	*mAbB*
M_w (g/mol)**	148,800 ± 2%	153,340 ± 3%	190,900 ± 3%
A_2 (cm^3-mol/g^2)	1.0×10^{-4} ± 4%	7.33×10^{-5} ± 6%	5.33×10^{-6} ± 34%
A_3, quadratic fit (cm^6-mol/g^3)	8.4×10^{-4} ± 8%	3.37×10^{-4} ± 18%	NM*
A_3, from maximum of I_R	9.6×10^{-4} ± 14%	4.46×10^{-4} ± 15%	NM*
A_3, Boltzmann	1.1×10^{-3} ± 10%	3.37×10^{-4} ± 14%	3.4×10^{-6} ± 71%
A_4 (cm^9-mol/g^4)	NM*	2.9×10^{-3} ± 24%	NM*
d (nm) from A_2 (quad fit)	12.4 ± 3%	11.1 ± 4%	5.4 ± 13%
D_H (nm) from dynamic light scattering (DLS)	11.6 ± 2%	11.9 ± 2%	17.4 ± 3%

Error bars are random and fitting errors, not systematic errors in concentration (c) or dn/dc. * NM = not measurable. ** This value of weight average molecular weight, M_w, is that of the unaggregated protein and hence forms the "M_0" in the denominator of M_w/M_0.

The hard sphere equivalent diameters (d) shown in Table 1 were determined according to:

$$A_2 = \frac{2\pi d^3 N_A}{3M^2} \tag{9}$$

where M_w from Table 1 was used for M in equation 9.

A_2 is sometimes used in the biotechnology community as an indication of protein stability because a large value, such as for the NISTmAb and mAbA, ≈ 10^{-4} cm^3-mol/g^2, indicates strong repulsion among proteins, whereas small values, < 10^{-5} (mAbB) indicate weak repulsion and possible "sticky" spots (e.g., hydrophobic or dipolar patches) on protein surfaces. By this informal criterion, the A_2 values for the NISTmAb and mAbA suggest good protein stability in solution.

From this point of view, the relationship of the equivalent hard sphere, d, by A_2 in equation 9 and the equivalent hydrodynamic diameter, D_H, from the diffusion constant determined by DLS is interesting. D_H, which is from the extrapolation of the diffusion coefficient to zero concentration, is independent of interprotein interaction potentials, whereas A_2 is directly proportional to the excluded volume resulting from interprotein potentials. If the proteins act as hard spheres, then $d = D_H$. If there is an extra repulsive term (e.g., from electrostatic repulsion between proteins), then $d > D_H$. If there is attractive interprotein potential, which can enhance aggregation, then $d < D_H$. Table 1 shows that $d \approx D_H$ for the NISTmAb and mAbA, suggesting good stability, whereas $d < D_H$ for mAbB, suggesting poorer stability. These trends are corroborated by the kinetics below.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

The use of experimentally determined A_3 (and A_4) for assessment of protein stability was recently introduced (*1*). It has been further found by the same authors that the most stable proteins have a positive A_3, which is immediately detectable from raw scattering data by the existence of a maximum in Rayleigh Scattering ratio $I_{R,max}$, such as seen in Figure 3a for the NISTmAb and mAbA.

In terms of the of using only a single angle at 90° in the current SMSLS, assuming that protein aggregation proceeds in a spheroidal fashion at least early on, then using Zimm's approximation for $q^2\langle S^2\rangle_z << 1$—where $\langle S^2\rangle_z$ is the z-average mean square radius of gyration, $q = (4\pi n/\lambda)\sin(\theta/2)$ is the amplitude of the scattering vector, and θ the scattering angle—allows estimation of the error from using just θ = 90°. For a sphere, $\langle S^2\rangle = 3D_H^2/20$ so that for the NISTmAb, $\langle S^2\rangle_Z^{1/2} = 4.5$nm, and the error in M_w is only 0.26%. In other words, the error in using θ = 90° instead of extrapolating to θ = 0° is only 0.26% according to the term $\left(1+\frac{q^2<S^2>_z}{3}\right)$ from equation 3. Using just θ = 90° detection, globular proteins of this density could be measured up to $M_w/M_0 = 240$ or 3.5×10^7 g/mol with only 10% error. Of course, the SMSLS platform accommodates multi-angle detection for cases where extrapolations to q = 0 are required.

Determination of M_w and A_2 also allow assessing the "diluteness" of a protein solution before any aggregation occurs. This is gauged by verifying that the following dimensionless product is much less than unity:

$$2A_2cM_w + 3A_3c^2M_w << 1 \tag{10}$$

This is the ratio of the second plus third virial coefficient terms to the leading term in equation 5. For example, for the NISTmAb, using the parameters from Table 1, the product will be 0.32 at 0.010 g/cm^3, the concentration used for a number of experiments presented below (i.e., the second and third virial terms will have a 32% effect on the LS).

The issue of protein concentration and virial coefficient effects on LS is of particular importance, as it relates directly to the major current trend in biotechnology innovation focusing on development of stable, very high concentration protein therapeutics (c > 0.1 g/cm^3) (*18*). This relatively new direction for biotechnology product development has suddenly focused much more attention on proximity energies and colloidal aspects of protein behavior (*19*).

Kinetics of Protein Aggregation

Time-Dependent Signatures of Protein Aggregation

There are many mechanisms of protein aggregation; some are reversible, whereas many are irreversible. Some involve a small number of proteins to form oligomers, such as dimers, trimers, and so forth, whereas others can continue to aggregate until colloidal particles form, followed by precipitation (*20–25*).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Work is underway to understand the different time-dependent LS signatures associated with specific aggregation processes, which will allow for rapid interpretation of protein aggregation kinetics. Some characteristic signatures are seen in Figure 3a, in which the aggregation is induced either by temperature or stirring; (i) "self-limiting" aggregation in which the LS reaches a plateau and proceeds no further; (ii) "unlimited" aggregation in which aggregation can increase with a positive second derivative; and (iii) an increase in aggregation until particulates form, seen in the increase in "noise" level of the LS signal that indicates the appearance and increase in particulate levels. After particles reach a certain size they begin to precipitate, which causes a decrease in signal. Data have been cropped for (iii) in this figure to avoid overshadowing the other data.

The data in Figure 3a is expressed in terms of M_w/M_0. The representation of aggregation in terms of the ratio of M_w to unaggregated, native protein mass M_0, M_w/M_0, is especially convenient because it represents directly the average number of proteins per aggregate.

If aggregates are already present when SMSLS monitoring begins, then additional aggregation is monitored, but caution must be used in the M_w/M_0 representation. Because it is always possible to measure the starting M_w by the above equilibrium procedure, it can be determined quickly if there are pre-existing aggregates in solution. If $M_w(t = 0) > M_0$, then the known M_0 for the native mAb can be used to form the ratio M_w/M_0.

Despite the variety of signatures seen in Figure 3a, the earliest phase of aggregation is normally linear, just as virtually any continuous, monotonic function of time is linear over a short enough interval of time, given by the first linear term of a power series of the function.

Figure 3b shows the early linear regimes from the three data sets shown in Figure 3a. M_w/M_0, is used again, and the slope of this linear portion makes a good choice for aggregation rate. The aggregation rate, defined as $d(M_w/M_0)/dt$, is the fractional increase in aggregate mass per second. Formally:

$$\text{Aggregation Rate} = \text{AR}(\text{s}^{-1}) \equiv \frac{d(M_w / M_0)}{dt}\Big|_{t=0} \tag{11a}$$

With this, AR becomes a genuine physical quantity with units of s^{-1}. Although these are the units of hertz, the latter is reserved for periodic phenomena, which aggregation is not. Hence, the unit of AR is here simply termed "AR" and its unit definition is:

$$1\ \text{AR} \equiv 1\,\text{s}^{-1} \tag{11b}$$

An advantage of this definition is that it is compatible with the International System of Units (SI; MKSA [meter, kilogram, second, ampere]), CGS (centimeter, gram, second), and most other systems of units. Defining AR this way allows one to speak of "milli-AR" (10^{-3} s^{-1}), "micro-AR" (10^{-6} s^{-1}), and so forth.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

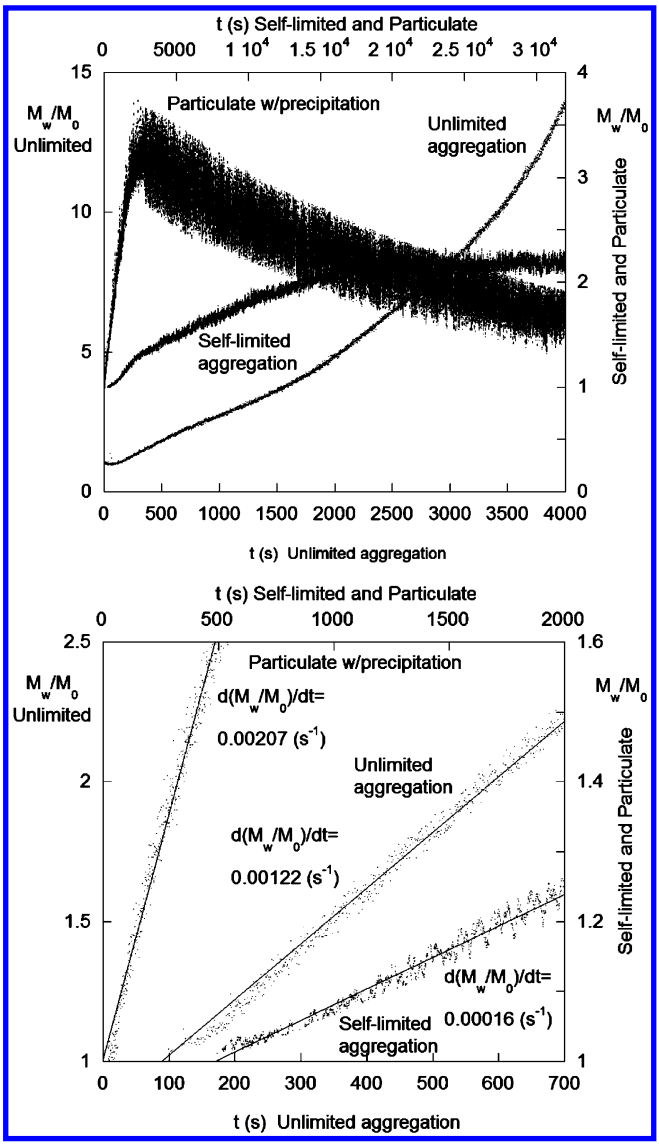

Figure 3. (a) Different time-dependent aggregation signatures for cases (i), (ii), and (iii) in the text. (i) monoclonal antibody B (mAbB) at T = 35 °C and stirring at 50 revolutions per minute (RPM); (ii) monoclonal antibody A (mAbA) at T = 60 °C, no stirring; and (iii) mAbB at T = 35 °C and stirring at 1,000 RPM. Note the different x- and y-scales, as labeled. (b) Early linear phase of the data in (a). Note the different x-and y-scales, as labeled.

Use of the early linear regime gives acceptable levels of reproducibility. For example, six measurements of aggregation on the NISTmAb at 75 °C yielded an AR of $0.00011 s^{-1} \pm 31\%$. While this error bar may seem large, it is actually quite acceptable because, as seen below, the AR changes by orders of magnitude as

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

stressors such as temperature and stirring are varied. On the other hand, beyond the linear regime, the time-dependent signatures of nominally identical reactions can diverge significantly, indicating that stochastic variables come into play that affect the reproducibility of long-term aggregation. A further advantage of using the initial linear regime for AR is that it allows the earliest possible detection of aggregation and quantification of kinetics. Hence, AR based on the linear regime is preferred at this time.

Determining AR at High Protein Concentration

For dilute solutions, as defined with reference to the dimensionless quantity in equation 10, discussed above, it is particularly easy to form M_w/M_0: one subtracts the pure protein buffer scattering from the scattering of the initial unaggregated protein solution. This latter is then the excess Rayleigh ratio due solely to the protein. Subsequently, the baseline subtracted data divided by this excess Rayleigh ratio is M_w/M_0.

In the case where more concentrated solutions are being investigated the AR is:

$$AR \equiv \frac{d(M_w / M_0)}{dt}\Big|_{t=0} = \frac{\left(1 + 2 < A_2 > cM_w + 3 < A_3 > c^2 M_w\right)^2}{I_{R,0}\left(1 + 2A_2 cM_0 + 3A_3 c^2 M_0\right)} \frac{dI_R}{dt}\Big|_{t=0} \quad (12)$$

where $I_{R,0} = I_R(t=0)$, and $<A_2>$ and $<A_3>$ are the complex, double z-averages of A_2 and A_3, respectively, as the aggregation process makes the scattering population increasingly polydisperse. Because these averages cannot be computed without knowledge of the aggregate population and of M_w itself, a further simplification is needed.

In the early phase of aggregation, when $M_w/M_0 << 1$, the approximations can be used that $<A_2> \approx A_2$, $<A_3> \approx A_3$, and $M_w \approx M_0$, so that:

$$AR \approx \frac{1}{I_{R,0}}\left(1 + 2A_2 cM_0 + 3A_3 c^2 M_0\right)\frac{dI_R}{dt}\Big|_{t=0} \quad (13)$$

Or, if M_w/M_0 is computed uncorrected, as described above, the corrected value is:

$$AR \approx \left(1 + 2A_2 cM_0 + 3A_3 c^2 M_0\right)\frac{d\left(M_w / M_0\right)_{\text{uncorrected}}}{dt}\Big|_{t=0} \quad (14)$$

It is emphasized that this approximation is only valid early enough in aggregation, when M_w/M_0 is not much greater than 1, so that the LS is still dominated by the proteins still in unaggregated state. The detailed nature of how the virial coefficients for the aggregates change with aggregation state is a very complex issue in itself. Use of the above correction at low M_w/M_0 avoids this problem and also adheres to the guiding principle in this work, namely that the best means of obtaining quantitative, reproducible AR, in terms of $d(M_w/M_0)/dt$,

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

is to measure this quantity very early in the aggregation. This has the further advantage of yielding AR rapidly during an assay, instead of having to wait extended periods for an aggregate population to build up that would be detectable by less sensitive means.

Figure 4 shows M_w/M_0 for NISTmAb aggregation at T = 75 °C for three different concentrations: 0.010, 0.060, and 0.090 g/cm^3 (run simultaneously in the SMSLS instrument). Although the data are shown up to M_w/M_0 substantially > 1 in order to illustrate virial coefficient effects on the overall scattering, the AR were determined in the low M_w/M_0 range. The inset shows the low M_w/M_0 AR for the NISTmAb at T = 75 °C with and without the virial coefficient corrections for $d(M_w/M_0)/dt$, using the values of M_w and A_2 from Table 1, and the average of the two experimentally determined values from Table 1 of $A_3 = 0.0009$. The corrected rates show that AR increases with concentration for the NISTmAb at T = 75 °C. In a two-step aggregation process involving a first step of unfolding and a second step of proteins sticking together in a diffusion-controlled collision, the concentration dependence of AR in the inset shows that diffusion is the rate-limiting step in this case.

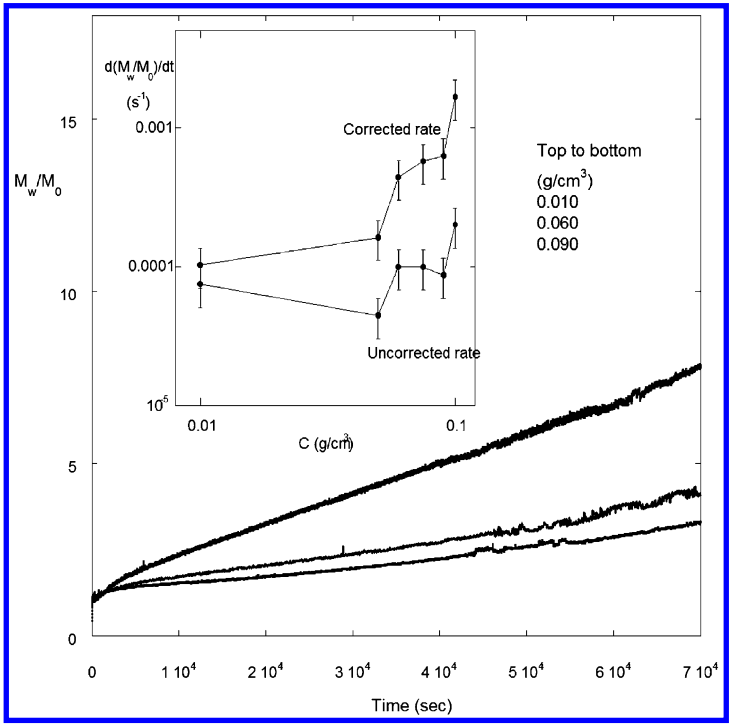

Figure 4. Equivalent weight average molecular weight ratio to unaggregated protein molar mass (M_w/M_0) for NISTmAb aggregation at T=75 °C for three different concentrations: 0.010, 0.060, and 0.090 g/cm^3. The inset shows corrected and uncorrected aggregation rate (AR) for NISTmAb at T = 75 °C, by equation 13 and 14 with M_w, A_2, and A_3 from Table 1.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Aggregation Due to Temperature

Most proteins will aggregate as temperature rises sufficiently. This is generally thought to be due to partial unfolding of the protein, or "denaturation" as Pauling originally conjectured, seeing that the energies of activation corresponded to the breaking of many cooperative low-energy bonds, such as hydrogen bonds (*26*).

Figure 5a shows examples of thermally induced aggregation of the NISTmAb run simultaneously at various temperatures. At 68 °C, the experiment lasted nearly 3 days. NISTmAb has excellent temperature stability compared to several other proteins, but nonetheless has an "unlimited aggregation" signature once aggregation begins (i.e., M_w/M_0 is concave upwards). Figure 5b zooms in on the early linear regime for some of the Figure 5a curves from which the AR were subsequently computed and used below in the Arrhenius plot of Figure 6. Such initial linear regimes have been found in most experiments and form the basis for quantifying AR in terms of $d(M_w/M_0)/dt$.

For experiments on all mAbs that report a fixed temperature, the temperature of the aluminum block sample cell holder is first brought up to the desired temperature and allowed to stabilize prior to inserting the sample cell containing the mAb solution. With 100 µl of solution, a typical volume used in many experiment, the time to stabilization of the sample was approximately 10 seconds. With 26 to 50 µl of solution, the time to stabilization was well under 10 seconds.

Arrhenius Plots

Figure 6 is an Arrhenius plot of AR in the early linear regime for the NISTmAb protein, mAbA, mAbB, and mAbC. The NISTmAb protein is the most thermally stable of the four proteins shown. Measured AR for the proteins over the temperature range shown vary by more than 9 orders of magnitude. Error bars are based on standard deviations from multiple measurements at selected temperatures.

Table 2 shows some of the important features of the Arrhenius analysis: the activation energy, ΔE_{act}, for protein unfolding, which for mAbA and mAbB breaks into two distinct Arrhenius regimes with different ΔE_{act} in each, separated by a breakpoint at the melting temperature, T_m. Also shown as a benchmark is AR at T = 65 °C, which contrasts the widely different stabilities.

Of high interest is the fact that for the NISTmAb protein, mAbA, and mAbB, there is no correlation between thermal stability of proteins and activation energy, which is proportional to the slopes of the lines in Figure 6. Whereas ΔE_{act} varies by only a factor of 20% among these three proteins and with no particular relation to stability, the AR vary by over 10^6. There is also no correlation of protein stability with T_m, as is apparent for mAbA and mAbB in Figure 6.

ΔE_{act} for mAbC is fourfold lower than the other proteins, but at 65 °C, it aggregates 5×10^6 faster than the NISTmAb. Although this may indicate some slight correlation with ΔE_{act} in this case, this research group has generally found no correlation between ΔE_{act} and AR among about 10 different proteins; ΔE_{act} can be the same for proteins widely separated by AR at any given temperature.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

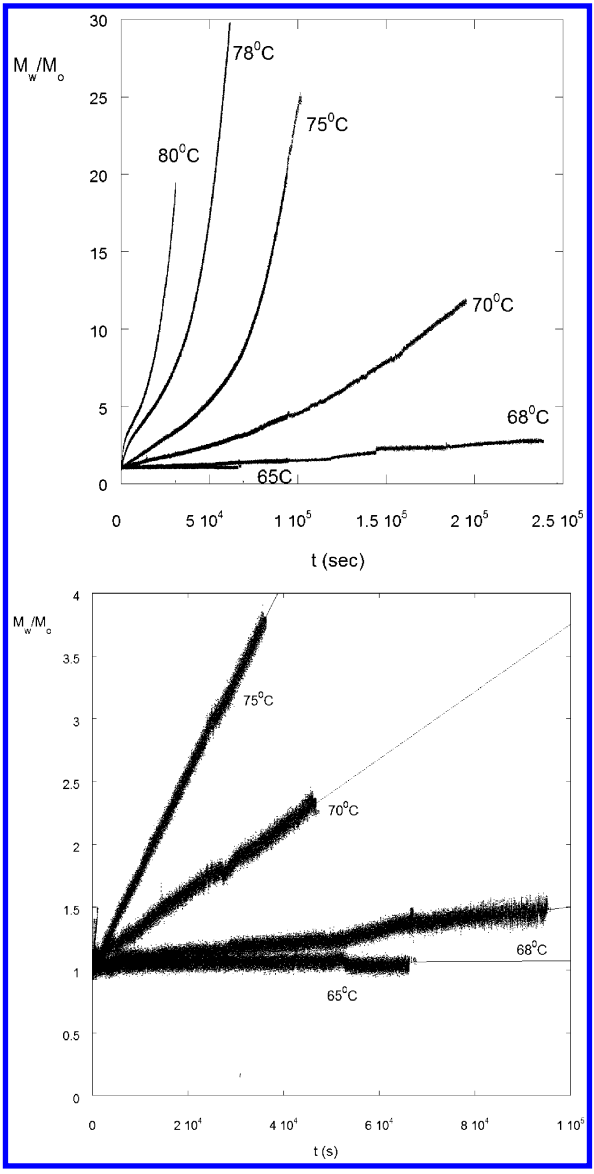

Figure 5. (a) Aggregation of NISTmAb at concentration (c) = 0.010 g/cm³ at several temperatures. (b) Early linear regime of aggregation from Figure 6a at various temperatures from which aggregation rate (AR) are determined by equation 11a.

Figure 7 shows an easy way to understand the AR. Namely, the reciprocal of the AR in Figure 6 gives the time it takes for the protein to aggregate to the point where the average M_w is that of a dimer (i.e., when $M_w/M_0=2$). This leads to a convenient unit, τ_D, the number of days to average dimerization:

$$\tau_D(\text{days}) = \frac{1.157\text{x}10^{-5}}{\text{AR}(\text{s}^{-1})} \tag{15}$$

For example, AR ≈ 0.1 μAR (= 10^{-7}s) leads to τ_D = 100 days. It is noted that although this is strictly the time for M_w/M_0 to become equal to 2, it does not reveal information on the nature of the population (i.e., any limitless forms of aggregate populations can yield M_w/M_0, such as those containing a very small number of large aggregates or a large number of small aggregates).

The time to dimerization for the NISTmAb, mAbA, and mAbB is expressed in days in Figure 7. A direct extrapolation of the Arrhenius behavior, for which data start at T=68 °C and up, is shown. It predicts that at pharmaceutically relevant temperatures, the NISTmAb is extraordinarily stable. At T = 37 °C it is extrapolated to be stable for 1,000 years, whereas at a refrigerated T = 4 °C it is extrapolated to be stable for 14 billion years, the age of the known universe.

Clearly, such extrapolations can be misleading. Drenski et al. (*1*) recently found that Arrhenius extrapolations into the T < 40 °C range dramatically overestimated the stability of the mAbs in that study. In fact, for T < 40 °C, thermal stress was no longer relevant to the stability of the protein solutions, and other stochastic stressors, as yet not clearly identified, are responsible for instability for T < 40 °C.

In spite of the failure of the Arrhenius extrapolations to predict AR in the T < 40 °C regime, the more unstable mAbB has 100-fold higher AR than the much more stable mAbA for T > 40 °C, and the NISTmAb is more stable than both.

The use of accelerated and forced degradation temperatures (~40–70 °C) to speed therapeutic protein stability experiments plays a major role in the rapid screening of protein formulations and candidate molecules. However, many aspects of the (purported) connections between high-temperature protein degradation, conformational stability, and pharmaceutical stability remain ambiguous. A recent study representing the first report in which thermally induced unfolding profiles of commercially viable therapeutic mAbs are shown together with actual long-term pharmaceutical storage stability data has been presented (*27*). In accord with the SMSLS findings described here, that study established little correlation between the first thermal unfolding transition and long-term pharmaceutical stability, indicating that mAbs capable of commercially viable stability profiles can exhibit a relatively broad range of thermal unfolding behaviors.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Table 2. The Various Proteins Studied by Simultaneous Multiple Sample Light Scattering (SMSLS), Their Sources, and Some Properties Found by SMSLS

Name	*$\Delta E_{act,lowT}$ (Kcal/M)*	*$\Delta E_{act,highT}$ (Kcal/M)*	*AR (s^{-1}) At T = 65 °C*	*T_m (°C)*
NISTmAb	114	NF	9.1×10^{-7}	NF
Monoclonal antibody B (mAbB)	130	32	1.8×10^{-2}	68
mAbA	136	23	6.5×10^{-4}	69
mAbC	37	NF	4.72	NF

AR = aggregation rate; ΔE_{act} = activation energy; NF = not found by this method up to the highest T measured (84 °C); T_m = melting temperature.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

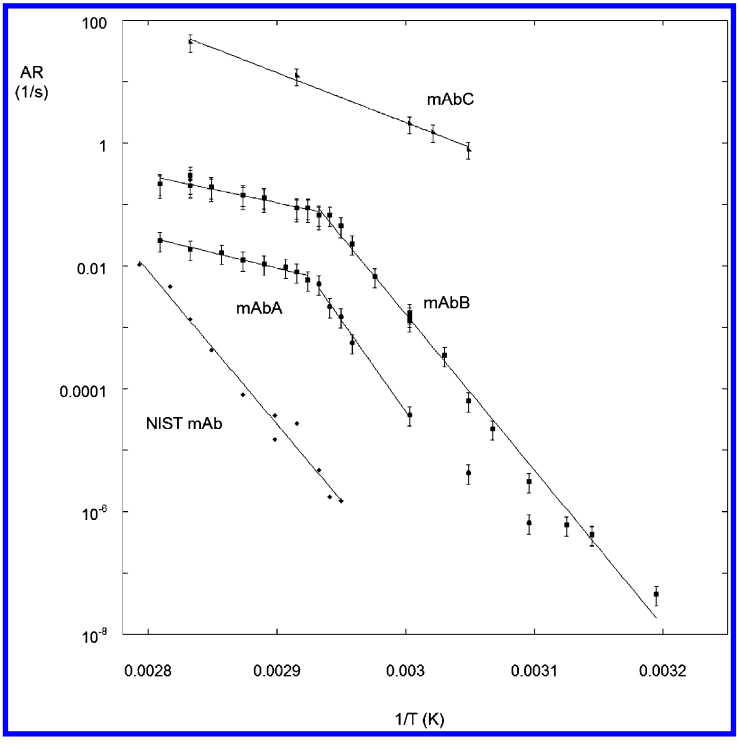

Figure 6. Arrhenius plot of NISTmAb, monoclonal antibody A (mAbA), and mAbB; and mAbC formulation #4.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

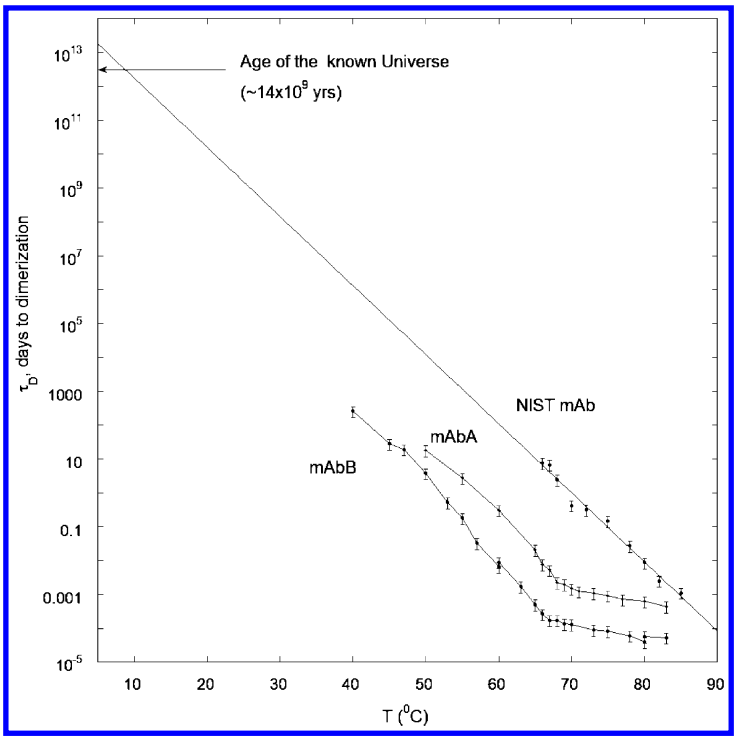

Figure 7. Time to dimerization, τ_D, in days, by equation 15, for NISTmAb, monoclonal antibody A (mAbA), and mAbB.

Stirring Effects on Protein Aggregation

Whereas the NISTmAb is the most thermally stable of several proteins recently investigated by SMSLS, it is nonetheless subject to aggregation under stirring, as are the other proteins in Table 2.

Figure 8 shows the kinetic data for stir-induced aggregation of the NISTmAb and how AR increases as stir rate increases. The stirring was at T = 35 °C, where the NISTmAb is thermally stable. The inset to Figure 8 shows the AR for both the NISTmAb and mAbB. Remarkably, there is no significant difference between the ARs of these mAbs, despite the NISTmAb being orders of magnitude more stable against thermal stress than mAbB. This suggests that the damage mechanism of the mAbs via stirring is quite different than thermal stress. This supports the notion that there exist different aggregation kinetic pathways and associated ARs for different stressors.

Figure 9 shows photos of the NISTmAb after stirring many hours under different stirring rates, from 100 to 1000 RPM, and how the turbidity increases with stir rate.

Bee et al. (*28*) and others (*29*, *30*) have found that protein aggregation due to stirring can sometimes be traced to increased exposure of the proteins to the air interface, rather than the mechanical shear stress of stirring. SMSLS

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

allows testing this by filling the vials up to the cap where there is no longer a liquid/gas interface. Figure 10 shows the difference in aggregation for the NISTmAb at 35 °C stirred at 500 RPM for a sample with the air/liquid interface ("uncapped") and with the air/liquid interface ("capped"). The AR of the latter is 2.5 times slower than the former. Hence, although the air/liquid interface leads to more rapid aggregation kinetics, removing it slows down, but does not stop, aggregation. This suggests that other aggregation stressors are operative. These could include mechanical shear and/or the liquid/solid interface interactions between the solution and sample cell walls and stir bar, enhanced by stirring. Although different proteins were found to have orders of magnitude difference in thermally induced aggregation, the same proteins have very similar AR under stirring, suggesting that a different damage mechanism is in effect under stirring, which is different from thermally induced unfolding. The capability to make such distinctions between different interfacial stresses may represent a convenient approach applicable to evaluating protein candidate molecules and trial formulations early in development for relative susceptibility to processing stresses.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

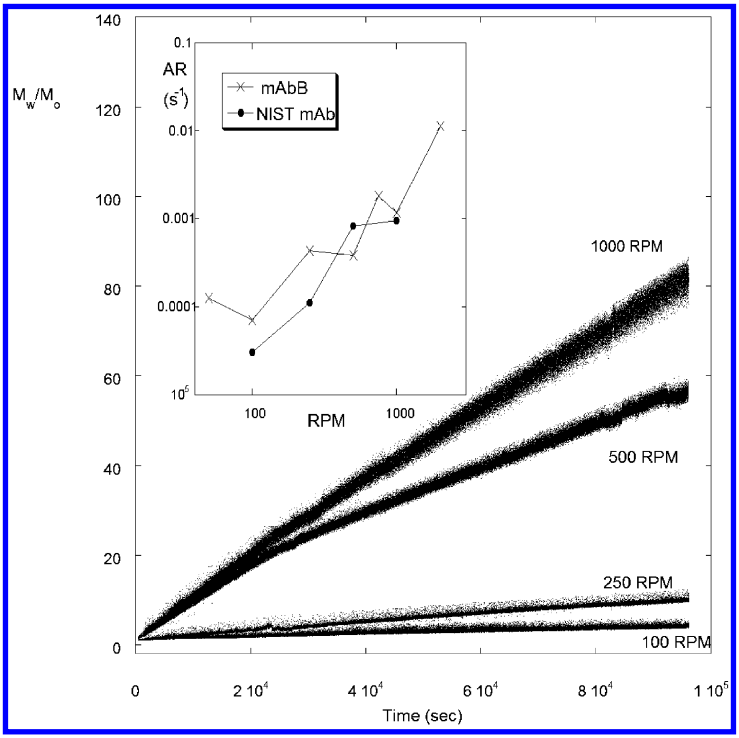

Figure 8. Aggregation of NISTmAb for increasing stirring from 100 revolutions per minute (RPM) to 1000 RPM at T = 35 °C. The inset shows aggregation rate (AR) (from low equivalent weight average molecular weight to unaggregated protein molar mass ratio [M_w/M_0] up to 5), versus RPM.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

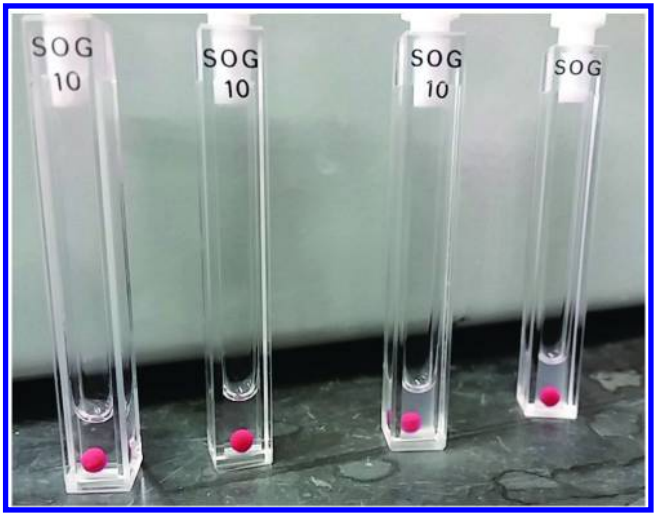

Figure 9. Photographs showing the increasing cloudiness of NISTmAb when stirred at increasingly high revolutions per minute (RPM) over the same period of time.

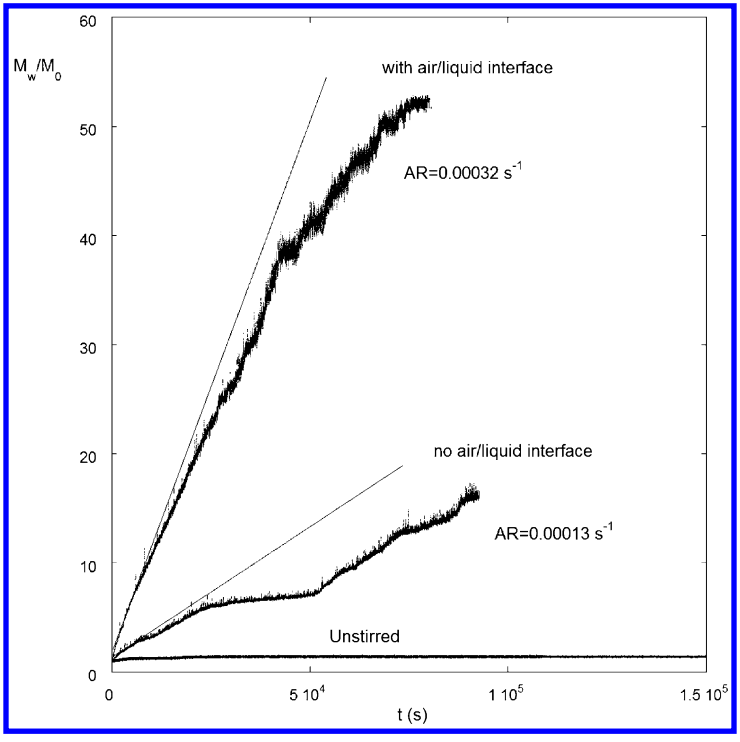

Figure 10. The effect of air/liquid interface on stir-induced aggregation of the NISTmAb at 500 revolutions per minute (RPM), T = 35 °C and concentration (c) = 0.010 g/cm³. Also shown is the stability of the protein without stirring.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Varying Formulation Conditions

An important application of SMSLS will be in the area of designing formulations, where optimizing factors such as excipients, pH, ionic strength, concentrations, and so forth must usually be worked through empirically to find the best choice. The ability of SMSLS to simultaneously handle multiple samples, as well as the ability to titrate samples with different agents while monitoring in real time, will increase formulation screening throughput enormously.

Figure 11 shows the aggregation behavior of mAbC at T = 60 °C, without stirring, at a concentration of 0.001 g/cm^3, for four different formulations which vary in pH and ionic strength. As seen, the differences in stability due to formulation conditions are dramatic, so much so that a logarithmic time scale is needed to appreciate the different ARs. The inset to Figure 11 shows AR versus formulation number. There is over a 600-fold difference in rate between the most stable and least stable formulations.

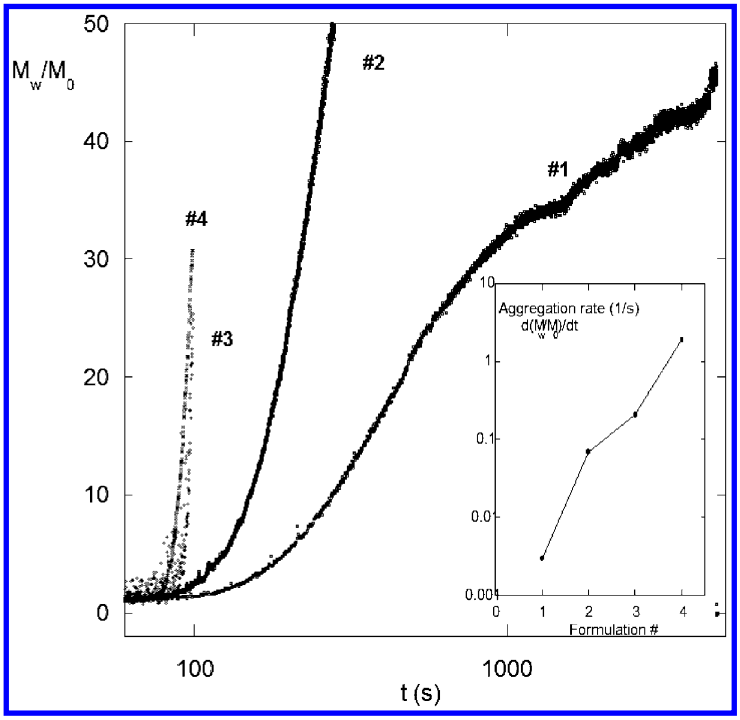

Figure 11. Aggregation of monoclonal antibody C (mAbC) at 60 °C, four formulations. The inset shows the aggregation rate, AR (s^{-1}), calculated for the linear regime up to equivalent weight average molecular weight ratio to unaggregated protein molar mass (M_w/M_0) = 5 for formulations #1 through #4.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Particulates That Form During Protein Aggregation

Particulates in therapeutic protein formulations can arise from a number of sources: highly aggregated protein and silicone oil, as well as adventitious particles from syringes, "dust," and processing equipment. Protein aggregates can reduce drug availability and, worse, provoke allergic and immune responses, whereas metal and oil particles may create heterogeneous particles possessing even greater immunogenicity as well as other adverse physiological consequences (*31*, *32*).

The U.S. Food and Drug Administration (FDA) has an interest in regulating these, and the National Institute of Science and Technology (NIST) is seeking means of standardizing their characterization. Major efforts are underway to better characterize particulates in protein solutions (*33–36*) The issue of "subvisible" particles and the need for more rigorous quantification, monitoring, and control has received much attention in recent years (*37*).

Historically, the control of aggregates and particles in biotechnology products has relied almost exclusively on SEC for soluble aggregate quantification and on light obscuration methods for particle counting. However, particles within the size range 0.1 μm to 10 μm have been largely overlooked, despite awareness that particles within this range are capable of provoking immunogenicity (*31*). The biotechnology industry is striving to identify new and improved methods capable of detecting particles within this range, as well as methodologies to more effectively probe their origins and ways in which bioprocessing methods and stresses influence this. The potential connection between subvisible and submicron particles has also been noted, and recent studies have begun to explore these interrelationships with a goal of better understanding product robustness and predicting stability (*38*). Emerging LS-based methods are likely to play an important role in providing more sensitive approaches to probing the earliest stages of protein aggregation and how the progression to submicron, subvisible, and visible particle formation occurs. SMSLS will aid in the characterization of particulates via its ability to resolve individual large particles that cause light scattering spikes (LSS). Case (iii) in Figure 3a shows a massive number of unresolved LSS in the form of a "noise" band that begins at about 2000 seconds.

The term Heterogeneous, Time-Dependent Static Light Scattering (HTDSLS) was introduced by Schimanowski et al. in connection with their instrument that could resolve and count LSS from individual large particles and simultaneously measure the background scattering from a population of homogeneous scatterers (*39*). The "heterogeneous" in the acronym refers to the fact that the solution contains both particulates and a background population of much weaker scatterers. The authors of that work were able to determine particle number density in solutions while recovering the background scattering. A demonstration was made by growing *Escherichia coli* (*E. coli*) bacteria in a broth in which poly(vinylpyrrolidone) (PVP) was dissolved. HTDSLS furnished the increase in time of the bacterial population and characterization of PVP M_w and R_g.

One of the conditions for performing HTDSLS is that there be relative motion between the incident beam and the particles. This ensures that particles pass

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

swiftly through the scattering volume, yielding well-defined LSS. Schimanowski et al. provided this by using a light scattering flow cell.

In the SMSLS system, the stirring feature can be used to provide the relative motion. Figure 12 shows light scattering peaks from 2 μm latex spheres in water, collected from the SMSLS system for the case where (i) there is no relative motion and the LSS are diffusion controlled and (ii) when 80 RPM stirring was used. Sampling was at 10 Hz. The diffusion-controlled peaks are of irregular shape and duration, whereas the LSS at 80 RPM are well defined, have narrower and more tightly controlled widths, and occur more frequently. Because the beam intensity over the scattering volume is not uniform, even monodisperse particles such as these produce a distribution of LSS peak heights. Taken over a long enough sampling period, the integral of the LS intensity over time is the same for both cases (i) and (ii), as expected. Each LSS peak corresponds to a single particle. In some cases, an LSS can have more than one peak, as seen in figure 12 for the case of stirred 2 μm latex spheres, indicating more than one particle in the scattering volume during the duration of the LSS.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

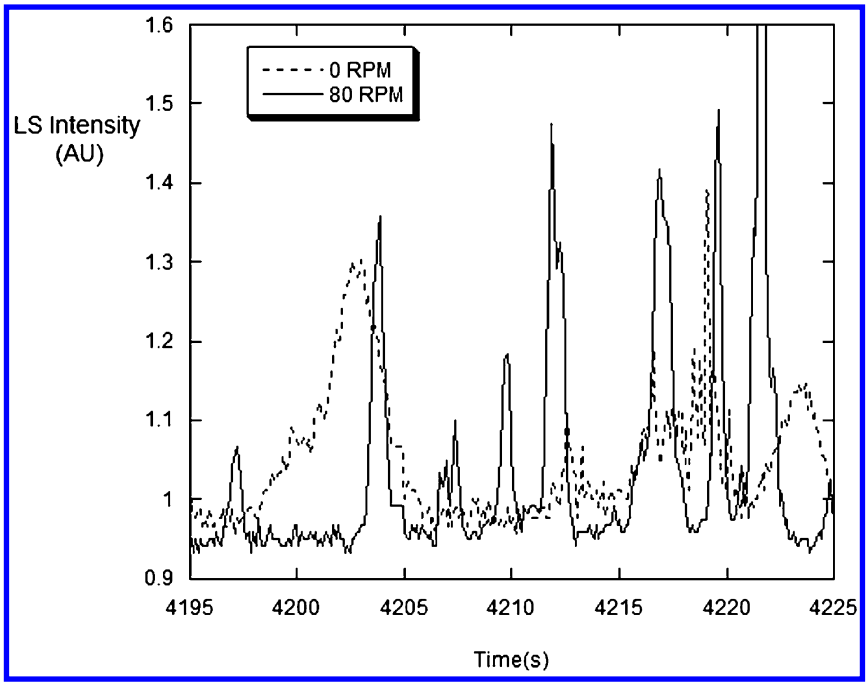

Figure 12. Latex spheres of 2 μm diameter stirred (solid line) and unstirred (dashed line) at 0 and 80 revolutions per minute (RPM), respectively. The LSS (light scattering spikes) in the unstirred case are due to diffusion and are irregular. In the stirred case, the LSS are well defined, narrower, and more frequent. LSS arise from individual particles passing through the scattering volume in the sample cell.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

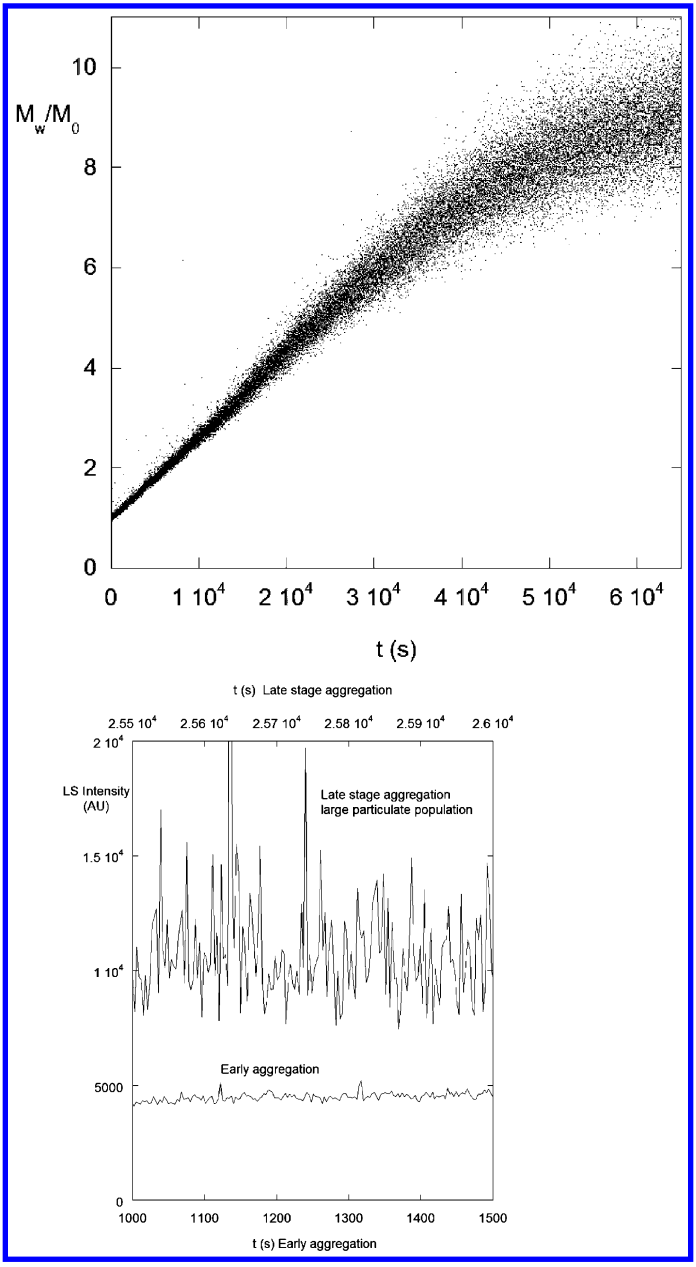

Figure 13. (a) Aggregation and increase in particulate population for mAbB stirred at 100RPM with no air/liquid interface. (b) Swaths from Figure 13a data.

The increasing width of the light scattering data in Figure 13a shows particulates forming for mAbB when stirred with no air/liquid interface at 100RPM at 0.010 g/cm^3. Figure 13b shows two 500 s swaths from the Figure 13a data. In the early swath, there are few particulates, and these are small, as seen by the low amplitudes of the LSS. Later in the aggregation process, for a 500 s swath starting at 25,500 s, the particulate population has a higher number density and the particles are much larger, as can be seen by the higher density and amplitudes of the LSS.

With further development, the LSS spectra potentially should yield particle density and how it changes in time, and a measure of the molar weight distribution for the particles.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Conclusions

SMSLS is a versatile platform in prototype stage that allows quantitative, real-time, high-throughput monitoring of aggregation in protein formulations. The terms "AR" (aggregation rate, equations 11a and 11b) and τ_D (average time to dimerization, equation 15) have been introduced to provide convenient, characteristic measures of aggregation.

SMSLS also is suited for obtaining absolute equilibrium characteristics: M_w, A_2, A_3, and, in some cases, A_4. The current version is essentially 16 independent light scattering instruments in one, with individually controllable sample stressors and data collection. Stressors such as fixed temperatures, temperature quenches or jumps, controlled stirring, and exposure to air/liquid interface and to other materials (e.g., metals, oils) can be applied and the effects on aggregation monitored. Samples can be titrated during monitoring experiments. Depolarized scattering from anisotropic structures, such as fibrillar aggregates, is currently being investigated. Multi-angle sample cell holders also have been made.

Virtually all proteins studied to date using SMSLS aggregate under thermal stress. All such proteins have temperature-dependent ARs that yield well-defined Arrhenius plots and activation energies. So far, no correlation has been found between activation energies and the propensity to aggregate for any given protein formulation, nor has any correlation with T_m been found.

All proteins so far tested with SMSLS aggregate under stirring stress. AR is generally, but not always, less when the air/liquid interface is eliminated during stirring. The SMSLS instrument creates the possibility to apply the stirring stress in the presence of different materials, thereby evaluating susceptibility to different contact surfaces.

Although there are correlations between the equilibrium properties, such as higher stability being found for proteins with high A_2 and A_3 under equilibrium conditions, as well as the hard sphere diameter from A_2 being close to the D_H from DLS, protein aggregation is ultimately a kinetic phenomenon. As such, quantitative monitoring of the aggregation process is one of the most direct means of assessing protein formulation stability.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

SMSLS was used here to monitor instabilities that varied in temperature-stressed AR by almost three orders of magnitude among formulations at different pH and ionic strength for the same protein.

The ability of SMSLS detection to measure light scattering peaks from individual particles passing through a sample cell's scattering volume offers the opportunity to monitor and characterize the onset and evolution of sub-micron- and micron-scale particulates in protein solutions at the same time AR and other features of the soluble protein and aggregate population are characterized.

Acknowledgments

We thank Dr. John Schiel of NIST for providing the NISTmAb and buffer materials. APMT acknowledges support from the National Science Foundation (NSF) Small Business Innovation Research (SBIR) program for some of the SMSLS developments. Biogen Idec is acknowledged for partial support of this work and for providing protein samples.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

References

1. Drenski, M. F.; Brader, M. L.; Alston, R. W.; Reed, W. F. *Anal. Biochem.* **2013**, *437*, 185–197.
2. Rarity, J. G.; Seabrook, R. N.; Carr, R. J. G.; Weitz, D. A. *Proc. R. Soc. London, Ser. A* **1989**, *423*, 89–102.
3. Sokolowski, F.; Modler, A. J.; Masuch, R.; Zirwer, D.; Baier, M.; Lutsch, G.; Moss, D. A.; Gast, K.; Naumann, D. *J. Biol. Chem.* **2003**, *278*, 40481–40492.
4. Panyukov, Y.; Yudin, I.; Drachev, V.; Dobrov, E.; Kurganov, B. *Biophys. Chem.* **2007**, *127*, 9–18.
5. Lyutova, E. M.; Kasakov, A. S.; Gurvits, B. Y. *Biotechnol. Prog.* **2007**, *23*, 1411–1416.
6. Arakawa, T.; Philo, J.; Ejima, D.; Tsumoto, K.; Arisaka, F. *Bioprocess Int.* **2007**, 36–47.
7. Rambaldi, D. C.; Reschiglian, P.; Zattoni, A. *Anal. Bioanal. Chem.* **2011**, *399*, 1439–1447.
8. Wen, J.; Arakawa, T.; Philo, J. S. *Anal. Biochem.* **1996**, *240*, 155–166.
9. Li, Y.; Weiss, W. F. t.; Roberts, C. J. *J. Pharm. Sci.* **2009**, *98*, 3997–4016.
10. Philo, J. S. *Curr. Pharm. Biotechnol.* **2009**, *10*, 359–372.
11. Drenski, M. F.; Mignard, E.; Alb, A. M.; Reed, W. F. *J. Comb. Chem.* **2004**, *6*, 710–716.
12. Drenski, M. F.; Reed, W. F. *J. Appl. Polym. Sci.* **2004**, *92*, 2724–2732.
13. Reed, W. F. *Device and Method of Simultaneously Measuring the Light Scattering from Multiple Liquid Samples Containing Polymers and/or Colloids*; Google Patents, 2003.
14. Liu, R.; Barkhordarian, H.; Emadi, S.; Park, C. B.; Sierks, M. R. *Neurobiol. Dis.* **2005**, *20*, 74–81.
15. Zimm, B. H. *J. Phys. Colloid Chem.* **1948**, *52*, 260–267.
16. Zhao, H.; Brown, P. H.; Schuck, P. *Biophys. J.* **2011**, *100*, 2309–2317.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

17. Berne, B.; Pecora, R. *Dynamic Light Scattering*; Dover: Malabar, FL, 1990.
18. Narasimhan, C.; Mach, H.; Shameem, M. *Ther. Delivery* **2012**, *3*, 889–900.
19. Laue, T.; Demeler, B. *Nat. Chem. Biol.* **2011**, *7*, 331–334.
20. Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F. *Pharm. Res.* **2003**, *20*, 1325–1336.
21. Cromwell, M. E.; Hilario, E.; Jacobson, F. *AAPS J.* **2006**, *8*, E572–579.
22. Dobson, C. M. *Semin. Cell Dev. Biol.* **2004**, *15*, 3–16.
23. Li, Y.; Roberts, C. J. *J. Phys. Chem. B* **2009**, *113*, 7020–7032.
24. Mahler, H. C.; Friess, W.; Grauschopf, U.; Kiese, S. *J. Pharm. Sci.* **2009**, *98*, 2909–2934.
25. Roberts, C. J. *Biotechnol. Bioeng.* **2007**, *98*, 927–938.
26. Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1936**, *22*, 439–447.
27. Brader, M.; Estey, T.; Bai, S.; Alston, R. W.; Lucas, K.; Lantz, S.; Landsman, P.; Maloney, K. *Mol. Pharm.* **2015**, *12*, 1005–10017, DOI: 10.1021/mp400666b.
28. Bee, J. S.; Stevenson, J. L.; Mehta, B.; Svitel, J.; Pollastrini, J.; Platz, R.; Freund, E.; Carpenter, J. F.; Randolph, T. W. *Biotechnol. Bioeng.* **2009**, *103*, 936–943.
29. Maa, Y. F.; Hsu, C. C. *Biotechnol. Bioeng.* **1997**, *54*, 503–512.
30. Treuheit, M. J.; Kosky, A. A.; Brems, D. N. *Pharm. Res.* **2002**, *19*, 511–516.
31. Rosenberg, A. S. *AAPS J.* **2006**, *8*, E501–507.
32. Schellekens, H. *Discovery Med.* **2010**, *9*, 560–564.
33. Barnard, J. G.; Singh, S.; Randolph, T. W.; Carpenter, J. F. *J. Pharm. Sci.* **2011**, *100*, 492–503.
34. Filipe, V.; Hawe, A.; Jiskoot, W. *Pharm. Res.* **2010**, *27*, 796–810.
35. Huang, C. T.; Sharma, D.; Oma, P.; Krishnamurthy, R. *J. Pharm. Sci.* **2009**, *98*, 3058–3071.
36. Malloy, A. *Mater. Today* **2011**, *14*, 170–173.
37. Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G.; Fan, Y. X.; Kirshner, S.; Verthelyi, D.; Kozlowski, S.; Clouse, K. A.; Swann, P. G.; Rosenberg, A.; Cherney, B. *J. Pharm. Sci.* **2009**, *98*, 1201–1205.
38. Bai, S.; Murugesan, Y.; Vlasic, M.; Karpes, L. B.; Brader, M. L. *J. Pharm. Sci.* **2013**, *102*, 347–351.
39. Schimanowski, R.; Strelitzki, R.; Mullin, D. A.; Reed, W. F. *Macromolecules* **1999**, *32*, 7055–7063.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch006

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Chapter 7

Informatics for Mass Spectrometry-Based Protein Characterization

Wenzhou Li,[†,1] Hua Xu,[†,2] and Oleg Borisov[*,3]

[1]Amgen Inc., 1 Amgen Center Drive, Thousand Oaks, California 91320, United States
[2]PepsiCo, Inc., 650 Brush Ave., New York, New York 10465, United States
[3]Novavax, Inc., 20 Firstfield Road, Gaithersburg, Maryland 20878, United States
[†]These authors contributed equally to this work.
[*]E-mail: oborisov@novavax.com

Mass spectrometry has become a central technology in protein research, covering a wide spectrum of tasks ranging from straightforward mass analysis to determination of amino acid residue connectivities and their modification status to insights into the higher order structure of proteins. These information-rich modern liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods, however, result in an overwhelming amount of data, which would be unfeasible to process without appropriate informatics software solutions. We argue that even the most powerful and modern mass spectrometer would not be of significant value to protein scientists unless task-appropriate tools were available to process these data files. This chapter starts with a brief introduction to proteomics and the state of bioinformatics applications in proteomics research, emphasizing the concepts and algorithms behind different tools. Informatics solutions employed in therapeutic protein development are then discussed, specifically for peptide mapping.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Introduction

Proteins represent one of the most diverse and complex classes of biomolecules, contributing to enzymatic, transport, signaling, and structural functions within cells. Utilization of modern analytical methodologies to analyze complex biological systems or subsets thereof results in generation of huge amounts of information-rich data that cannot be processed manually without the employment of software tools and "smart" algorithms. In the past decade, bioinformatics has become an integral part of research and development in the biomedical sciences (*1*). Bioinformatics now plays an essential role both in deciphering genomic, transcriptomic, and proteomic data generated by high-throughput experimental technologies and in organizing information gathered from traditional biological methods. Several excellent reviews have been recently published on the topic (*2*, *3*).

Liquid chromatography (LC) coupled with mass spectrometry (MS) or tandem MS (MS/MS) is often a technology of choice when it comes to the analysis of complex protein mixtures. Typical LC-MS/MS methods exploit a reversed-phase (RP) chromatographic separation mode to fractionate peptide mixtures generated by proteolytic digestion of proteins prior to introduction into the mass spectrometer. The method offers excellent sensitivity for detecting peptides, thereby enabling identification and quantification of thousands of proteins from a single run. Mass measurement of each of the eluting peptide precursors (MS) is first performed and then complemented with corresponding fragmentation (MS/MS) of these precursors. As modern instruments achieve higher sensitivity, increased resolution, and faster data acquisition speed, the size of data files collected are growing exponentially. The data analysis can be a true bottleneck if there are no readily available informatics tools. Finding optimum robust solutions that can improve the efficiency and quality of data analysis becomes an important consideration for the choice of LC-MS workflows.

Although the primary focus of this chapter is on the bioinformatics for MS-based protein characterization in biopharmaceutical applications, we note that most, if not all, of the MS software tools currently used in the biopharmaceutical industry originate from MS-based proteomics research. It is now well accepted that the term "proteomics" implies the use of LC-MS/MS methods to identify proteins in complex biological samples in order to gain knowledge on the whole proteome and insights on systems biology. In proteomics research, a protein may be identified through a sampling of only a small subset of peptides sufficient to indicate its presence, whereas protein modification status often is the secondary goal. In biopharmaceutical applications, mass spectrometry now plays a central characterization role in understanding the heterogeneity profile of a specific recombinant protein and, less frequently, a set of proteins, defining the biotherapeutic product in an effort to ensure the purity, safety, and quality of that product. The employment of LC-MS/MS for "protein characterization" is widely referred to in the scientific literature as "peptide mapping with mass spectrometry" (or simply "peptide mapping"). Protein characterization implies understanding sequence of the protein(s) of interest and its alternations and provides information on modifications, which are assessed for their impact on

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the quality of that protein. In this case, the intent is to identify each and every peptide, and, for that matter, each individual amino acid, along with the entirety of heterogeneity due to post-translational modification (PTM).

In principal, the LC-MS/MS instrumentation and methods used in proteomics investigations and in biopharmaceutical applications are similar but not exactly the same. Peptide mapping with LC-MS/MS implies that proteolytic mixture of peptides originating from a biotherapeutic protein are chromatographically separated and subsequently detected by MS (or "mapped" with respect to their "retention time—*m/z*" coordinates). A peptide map can serve as an identity test for a protein or biotherapeutic product and has multiple uses, including primary sequence determination, PTM identification and quantitation, sequence variant analysis, and product comparability evaluations.

Despite *prima facie* similarities of LC-MS/MS methods applied to proteomics and peptide mapping studies, there also are some fundamental differences, with the main difference defined by their application purpose. The scope of applications ranging from the in-breadth analysis of proteolytic digests of multiple complex proteins, as in case of proteomics, to the in-depth peptide mapping analysis, in most cases, of a single protein sample implies differences in the experimental scale, sample handling, sample preparation, chromatographic conditions, separation requirements, and data analysis approaches. For instance, most proteomic studies utilize discovery-type workflows targeting identification of multiple proteins. In the case of peptide mapping, the entire primary structure needs to be confirmed, and the presence of variants, which are often at low levels, is of greater interest than in identification-focused proteomics. Furthermore, proteomics investigations frequently deal with miniscule amounts of sample, preferring nano-flow modes of separation in order to achieve desired sensitivity, whereas peptide mapping applications use standard gas-assisted ionization sources to ensure robustness.

The chapter starts with a brief introduction to proteomics and the state of bioinformatics applications in proteomics research, emphasizing the concepts and algorithms behind different tools, and leading to the discussion of informatics solutions employed in therapeutic protein development, specifically for peptide mapping.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Informatics for MS-Based Proteomics

Proteome and Proteomics

The term "proteome," first introduced by Mark Wilkins in 1994 (*4–6*), defines the entire protein complement in an organism (*7*). Proteomics today encompasses the study of the proteome of an organism or cellular/sub-cellular subsets and represents one of the greatest challenges facing biological scientists in the post-genomics era (*8*). The goal of proteomics is to determine the sequences, structures, functions, amounts, activities, and modifications of all proteins in real time (*7*, *9*).

Proteins directly influence biological functions of cells by fulfilling relevant functions and controlling dynamic processes. For example, it is crucial to

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

determine changes in protein expression to understand cell development during carcinogenesis because proteins eventually determine cell phenotype (*10–15*). The importance of monitoring cell protein content can be explained by the fact that small changes in protein expression may account for the alteration of protein cascades that ultimately results in transformation and uncontrolled growth of cells. Moreover, changes in global protein expression need to be pinpointed to identify protein biomarkers involved in cancer-transforming processes. This identification may afford timely identification of cancer and the development of possible treatments according to the molecular profile of the tumor. Therefore, proteomics plays a crucial role in the detection of abnormal proteins that can be used for disease diagnosis and drug discovery. Enormous efforts are being devoted to the development of new technologies to detect abnormal proteins as they may signal the early formation of cancer (*11*, *14*, *16–21*). Early detection of cancer is by far the single best metric relating to survival, and thus these efforts are of extremely high significance. The classic example is the screening of patients for prostate-specific antigen (PSA). PSA is a significant predictor for determination of initial treatment among men with localized disease and the strongest predictor for success by hormone therapy as the first choice therapy (*22*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

MS-Based Proteomics

MS has been recognized as a powerful method to study complex proteome samples. A typical mass spectrometer consists of an ion source that ionizes analyte molecules, a mass analyzer that separates ions based on *m/z*, and a detector that measures the abundance of ions of each *m/z* value in a unit of time (*23*). Since the 1990s, advancement of MS instruments and techniques has revolutionized protein analysis in proteome research. Currently, MS is the most important proteomics tool and a single core technology that drives the field of proteomics biomarker discovery (*23–26*). Peptide mass fingerprinting, accurate mass-time tags, and MS/MS are the most popular approaches in proteomics studies.

Peptide Mass Fingerprinting (PMF)

PMF is a proteomics methodology that identifies proteins by matching experimental peptide masses based on the calculated masses of proteolytic peptides for all protein sequences in a database. It is based on the principle that proteins of different amino acid sequences produce different sets of peptides with specific masses after enzymatic digestion. The set of peptide masses for each protein in the database constitutes a mass fingerprint unique to the specific protein that can identify it within a protein mixture (*9*). Figure 1 shows the schematic of a standard PMF experiment (*27–29*). In PMF, proteins are first digested into peptides by protease(s). The resulting peptide mixture is analyzed by MS to gain experimental peptide masses. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) is the most commonly used MS system for PMF due to its simplicity, good mass accuracy and resolution, and high sensitivity (*23*). During data analysis, proteins in a sequence database containing all possible

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

proteins in the sample are digested *in silico* to generate theoretical peptides (*30*, *31*). Theoretical peptide masses are then calculated for the *in silico*-generated peptides. Protein identifications are achieved by comparing the experimental and theoretical masses of peptides. In addition to manual interpretation of PMF data, software tools also are available to perform automated PMF data analysis (*32–34*). PMF, in principal, also can be used to characterize PTMs in proteins due to the fact that each PTM introduces a specific mass shift to the peptide mass (*35–38*).

PMF is sometimes insufficient, however, to give unambiguous identifications, especially when PTMs are involved. PMF relies on the masses of peptides for protein identification, and there may be multiple peptides created from the sample that have masses within the mass accuracy of the mass spectrometer. Furthermore, specific sites of modification in a peptide with PTMs cannot be identified by PMF alone. Refinements to PMF include improved separations, such as multidimensional separations (*8*, *15*, *39*), and enhanced mass accuracy of mass spectrometers used for the analysis. Improved separation methods can reduce the complexity of the sample, and enhanced mass accuracy lowers the number of isobaric peptides created from a given protein database and thus reduces the ambiguity of the protein identifications (*40*). For example, Zhang et al. have demonstrated that peptides with acetylation and those with tri-methylation (mass shift 42.0106 Da vs. 42.0470 Da) can be differentiated by the use of PMF on an Fourier transform ion cyclotron resonance (FT-ICR) MS, which provided mass accuracy of $\Delta m/z < 0.001$ Da (*41*). Despite many refinements, PMF still is not well suited for analysis of complex protein samples.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

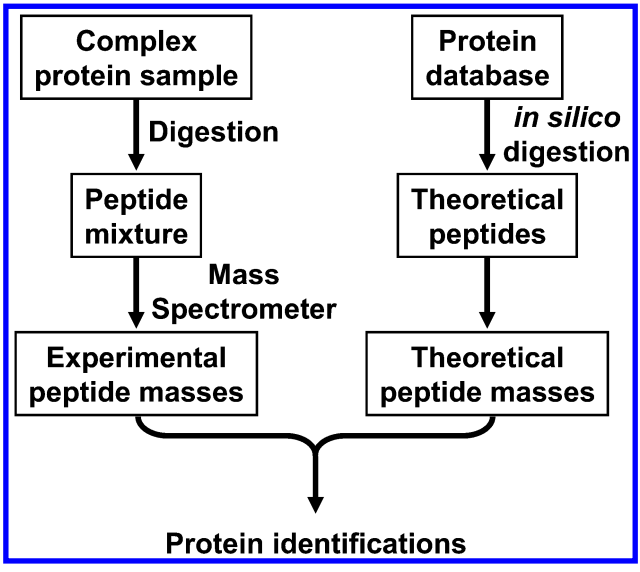

Figure 1. Schematic diagram describing typical peptide mass fingerprinting workflow (29).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Accurate Mass and Time (AMT) Tags

A concept of AMT tags, introduced by Richard Smith's research group at the Pacific Northwest National Laboratory, enables comprehensive high-throughput characterization of proteomes (*42*, *43*). The concept is similar to PMF but with the addition of a time domain as the important characteristic of detected peptides. The AMT methodology is designed to better utilize instrument duty cycle by reducing potential undersampling of coeluting peptide for improved quantitative fidelity compared to a conventional LC-MS/MS experiment (*44*). The main principal of the method is based on a knowledge of the combination of molecular mass and chromatographic elution times of peptides. That information is used as unique marker for the parent proteins. Mass spectrometric high mass accuracy (low ppm or better) measurements and efficient and reproducible chromatographic separation of proteolytic peptides on a RP column are the prerequisites for the successful application of this technology. For example, the achievement of 1 ppm mass accuracy has practical utility for analysis of small proteomes. Further improvement of mass measurement accuracy to sub-ppm levels can have significant impact on defining unique peptide mass tags in both bacterial and eukaryotic systems. The AMT technology consists of a range of analytical techniques and experimental protocols designed to rapidly manage the identification of large number of proteins in highly complex biological mixtures. The approach relies on first establishing an AMT tag database for an organism, tissue, or cell line by performing high-resolution shotgun proteomic analysis, and then retrieving information from this database to obviate the need for subsequent MS/MS analyses. Initial validation of peptides for use as AMT tags is based on a comprehensive analysis of a sample by data-dependent LC-MS/MS, using multiple injections and exclusion lists to increase coverage, followed by database searching and further definition of relative retention times using high resolution LC-MS data. Once AMT tags are generated, routine analysis of samples may proceed in a high-throughput mode using single dimension LC-MS analysis (*43*) because identification of peptides is based on the mass resolution and dynamic range of a single spectrum rather than a rate-limiting speed of MS/MS acquisition.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Tandem MS

As mentioned previously in this chapter, large-scale identification and characterization of proteins in complex proteome samples mainly rely on LC-MS/MS, as shown in the schematic in Figure 2 (*9*). In this approach, peptide precursor *m/z* values and the corresponding signature product ion spectra are used to identify and characterize the sequences and modifications of the precursor peptide ions. MS/MS provides the sequencing information for peptides, relating back to the identity of the peptides (*45–49*). Therefore, MS/MS achieves more reliable and unambiguous peptide and protein identifications than PMF even when instruments with lower mass measurement accuracy are used for analysis of complex samples (*23*). Before discussing informatics tools for analysis of mass

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

spectrometry data, we will provide a brief description of the basics of peptide identification using MS/MS.

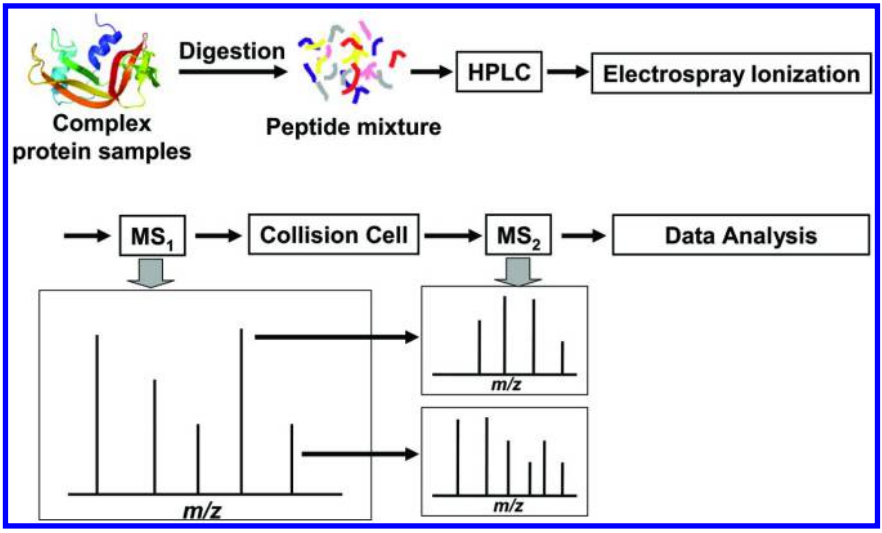

Figure 2. Schematic diagram depicting the principals of the tandem LC-MS/MS experiment.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Peptide Fragmentation

Tandem mass spectra are created by fragmenting precursor peptide ions into small pieces of product ions. The widely accepted nomenclature for fragmentation of protonated peptide ions (Figure 3) was proposed by Roepstorff et al. in 1984 and later improved by Biemann, Johnson, and others (*50–54*). Only charged fragments created during peptide fragmentation can be detected by MS. If the charge of the precursor remains on the N-terminal side following fragmentation, the charged fragments are called *a*, *b*, or *c* ions. If the charge remains on the C-terminal side, the charged fragment are called *x*, *y*, or *z* ions. Therefore, three series or six types of product ions (i.e., *a/x* ions, *b/y* ions, and *c/z* ions) can be created following the fragmentation of protonated peptide ions as shown in Figure 4. The subscript of an ion indicates the number of amino acid residues of the fragment, whereas its superscript represents the charge of the product ion.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

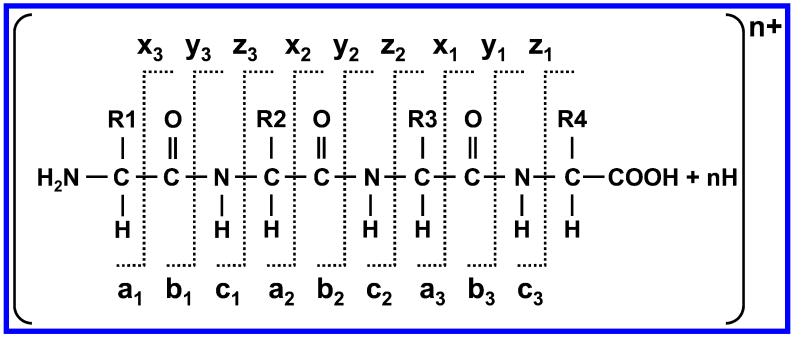

Figure 3. Nomenclature for fragmentation of protonated peptide ions. Adapted from http://www.matrixscience.com/help/fragmentation_help.html.

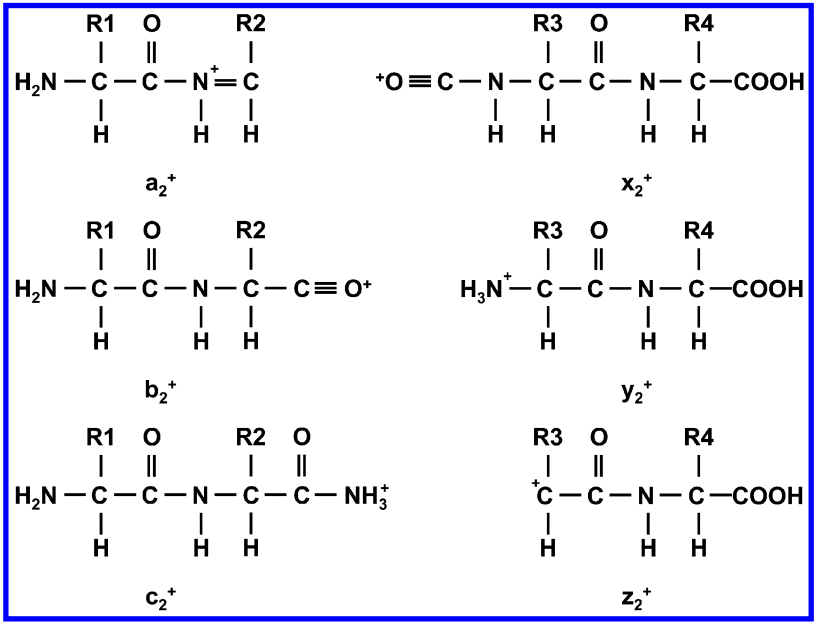

Figure 4. Example product ions created from peptide fragmentation. Adapted from http://www.matrixscience.com/help/fragmentation_help.html.

Several fragmentation methods have been developed for MS/MS. The most common methods are low- and high-energy collision-induced dissociation (CID) (*45*, *46*, *48*, *55*), electron capture dissociation (ECD) (*56*), electron transfer dissociation (ETD) (*57–59*), and surface-induced dissociation (SID) (*60–62*). Among them, low-energy CID (eV range of fragmentation energies) is the most frequently used fragmentation method, generating extensive complementary *b* and *y* ions due to the backbone fragmentation with some well-studied selectivity (*63*). Orbitrap instruments also employ a CID technique that takes place in a quadrupole cell, called higher-energy CID (HCD), although the collision energy

is still in the eV range. In contrast, most TOF/TOF instruments use high-energy CID with several keV energies and shorter activation time to fragment peptide ions, and those spectra usually exhibit increased side-chain fragmentation and abundant immonium ions (*64*). Different from collision-induced dissociation methods, ECD and ETD are radical-based fragmentation methods, primarily generating *c* and *z* ions through the cleavage of N-Cα bonds. These methods have advantages for longer peptides and peptides with labile modifications.

The abundance distribution of product ions in tandem mass spectra is governed by multiple factors, including fragmentation method and conditions, type of a mass spectrometer, and peptide sequence, among others. A mobile proton model has been proposed to describe the fragmentation pattern of protonated peptides in a qualitative manner (*65*, *66*). Several computational algorithms have been developed to quantitatively predict ion abundances in tandem mass spectra for peptides (*67*, *68*). That information assists with confident assignments of peptide sequences, as will be discussed later in the chapter.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Tandem MS Data Analysis

Tandem mass spectrometric data are the basis for identification of peptide sequences in most common applications. Interpreting peptide sequences, however, from the tandem MS data is a complicated job. Noise peaks due to chemical or electronic noise further complicate interpretation. Manual interpretation, although effective, is time-consuming and labor-intensive (*9*). Furthermore, results can be biased and highly dependent on the expertise of the interpreter. Manual interpretation of a large data set from a complex sample is an impossible mission. For example, a typical tandem MS data set contains 10^3 to 10^4 individual spectra. A human protein database can generate ~2×10^6 theoretical peptides. When combined, ~2×10^9 to 2×10^{10} comparisons between an experimental spectrum and a theoretical spectrum would be required to examine all of the possibilities. Therefore, computer programs are normally used to analyze tandem MS data sets to obtain peptide and protein identifications.

Various algorithms have been developed to automate data analysis for modern high-throughput MS/MS experiments. These algorithms fall under two main categories: *de novo* sequence inference and database searching (*69*, *70*). The *de novo* approach identifies peptide sequences directly from the tandem MS data (*71–79*). The most typical way is to match the mass difference between fragment ions to the mass of an amino acid residue, thus implying partial sequences on the peptide backbone. This type of algorithm is usually computationally expensive and limited by the mass accuracy of the tandem MS data and the continuity of fragment ion series (*70*). The database searching algorithms, on the other hand, identify peptides by comparing experimental tandem MS data with that from a protein sequence database (*80–95*). Because of their relatively lower computational expense and higher compatibility with low mass accuracy spectra, database searching programs are currently the most common tools used for analysis of complex data sets (*63*, *69*, *70*, *96*, *97*). However, it also should be noted that database searching programs can only identify those peptides that

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

are present in the provided protein database, which can limit their application when analyzing poorly defined biological systems. Automated tandem MS data analysis via database searching is discussed in detail in the next section.

Automated Database Searching for Tandem Mass Spectrometric Data Analysis

Database searching programs are most widely used to identify and characterize proteins, peptides, and their PTMs from tandem MS data. The schematic for a typical database search of tandem MS data is shown in Figure 5. In this approach, all potential peptides are created from the sequence database via *in silico* digestion with proteases. Theoretical spectra containing product ion series appropriate for the given fragmentation technique are created for the peptides. All tandem mass spectra in the data set are then compared with the theoretical spectra to give potential peptide matches. Scoring models are used to evaluate and score those peptide matches (*63*). Peptide matches with significant scores are assigned to the protein sequences in the database to produce protein matches. Proteins are scored either by summing the scores of their peptide matches or by probabilistic models (*98*, *99*).

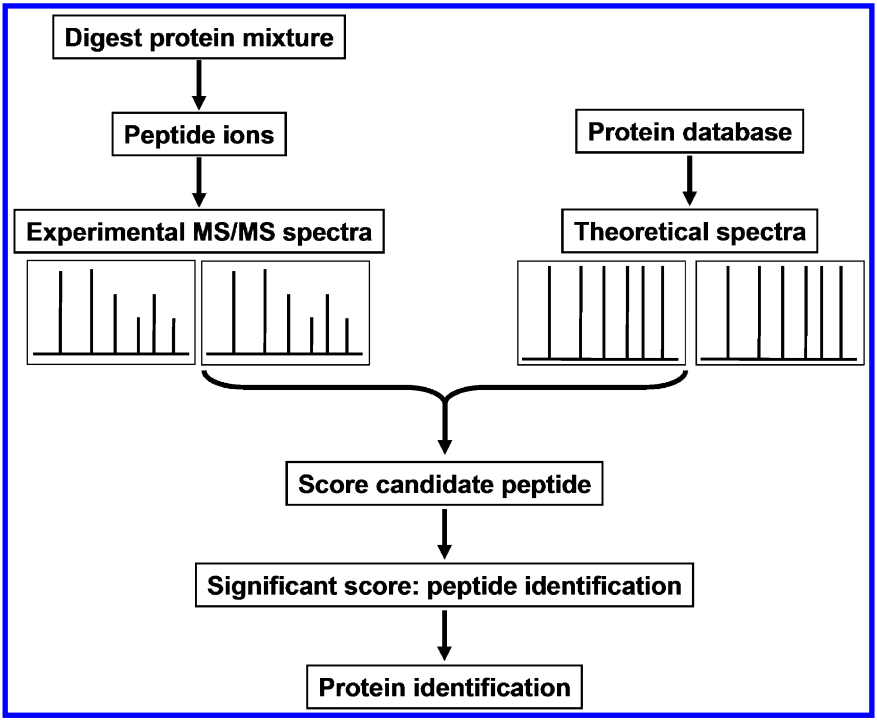

Figure 5. Schematic diagram for a typical database searching process of tandem mass spectrometric data.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

To date, several algorithms have been developed to automate the database searching process. Table 1 provides examples of the database search programs that can be accessed via the World Wide Web. These programs use different scoring models to evaluate potential peptide matches for experimental spectra and to differentiate true peptide identifications from false ones. Because of these differences, it can be beneficial to employ multiple search programs when analyzing tandem mass spectra of poor quality or with many product ions from unexpected fragmentation. The key comparison between these searching programs lies in their scoring algorithms (*63*).

Table 1. Examples of Database Searching Programs That Can Be Accessed via the World Wide Web

Program	*Website*
Mascot	http://www.matrixscience.com/
MassMatrix	http://www.massmatrix.net/; http://sourceforge.net/projects/massmatrix/
MS-Tag/MS-Seq	http://prospector.ucsf.edu/
ProteinLynx	http://www.waters.com/
PepFrag	http://prowl.rockefeller.edu/prowl/pepfrag.html
SEQUEST	http://fields.scripps.edu/sequest/index.html
SpectrumMill	http://www.chem.agilent.com/
X!Tandem	www.thegpm.org/tandem/index.html
Peaks	http://www.bioinfor.com/peaks/features/peaksdb.html
Andromeda	http://www.maxquant.org/
MyriMatch	http://fenchurch.mc.vanderbilt.edu/lab/software.php

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Scoring Algorithms for Database Searching

There are four major types of scoring models, categorized by Sadygov et al. (*63*), which are currently used in database searching programs: descriptive, interpretative, stochastic, and statistical/probabilistic models.

Descriptive Scoring Algorithms

These scoring methods are based on simple models to reconstruct spectra with basic information on relative ion abundances in tandem mass spectra of theoretical peptides. Potential peptide matches then are evaluated by mathematical models that compare experimental spectra with the predicted theoretical ones to give

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

scores. Programs of this type include SEQUEST (*80*), Sonar (*85*), and SALSA (*83*).

Interpretative Scoring Algorithms

Interpretative scoring models interpret partial mass intervals extracted from experimental tandem spectrum, typically corresponding to a mass of certain amino acid residues (sequence tags) (Figure 6). The sequence tags from experimental tandem mass spectra then are searched for against those in the protein database for potential peptide matches. Introduced by Mann and Wilm in 1994, the concept of sequence tags has become a widely used method to filter database entries based on partially interpreted sequence information of a peptide (*100*). The potential peptide matches then are evaluated by either probabilistic or correlational models. PeptideSearch (*81*), GutenTag (*88*), and Popitam (*89*) are examples of database search programs that employ an interpretative scoring model.

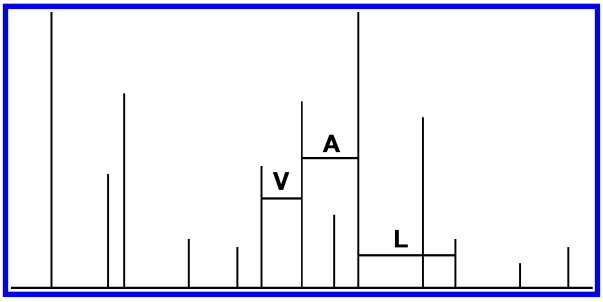

Figure 6. A simplified representation showing that the partial three-amino-acid sequence tag VAL can be inferred from a tandem mass spectrum.

Stochastic Scoring Algorithms

This type of algorithms uses stochastic models to generate tandem mass spectra for theoretical peptides. A stochastic model estimates the probability of fragment ions (e.g., *b*-, *y*-, and *a*-ions) by use of a large database of manually curated spectra with known peptide identities. The probability of observing the fragment ions from one specific cleavage event in peptides can be estimated by the frequency of the cleavage in the database. Database searching programs that use stochastic models include SCOPE (*82*) and OLAV (*90*).

Statistical/Probabilistic Scoring Algorithms

This type of scoring model employs statistical models to correlate peptide sequences to experimental spectra. The probability-based score is a direct measure of the probability that a match is significant. Probability-based scores

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

from different search algorithms can be directly compared, whereas descriptive scores must be converted to probabilities for comparisons (*101*). A group of programs that are based on statistical/probabilistic models have been developed, including Mascot, MassMatrix, X!Tandem, Andromeda (*102*), Peaks (*103*), MyriMatch (*104*), and others (*32*, *84*, *86*, *87*, *91–93*, *105–110*).

It should be noted that different algorithms differ in the selectivities they offer for identification of peptides. Thus, a common strategy suggests using two or more software approaches to provide complementary data and maximize the number of identifications from a dataset. In addition, approaches based on sequence tagging, such as interpretive algorithms, are useful in identifying peptides with unknown modifications because the presence of a specific sequence tag is often sufficient to point to its belonging to the region within the peptide.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Representative Scoring Models

Among all the database searching programs available, we will focus the discussion on scoring algorithms of SEQUEST and Mascot, as the most frequently used programs, and MassMatrix, which offers unique capabilities to identify disulfide- and chemically cross-linked peptides.

Scoring Models in SEQUEST

SEQUEST, developed by Eng et al. in 1994 (*80*), was the first automated approach for database searching based on the descriptive model. SEQUEST now has emerged as a part of the Proteome Discoverer software distributed by Thermo Fisher Scientific, Inc. (Waltham, MA). In SEQUEST, scoring of potential peptide matches is a two-step process, consisting of the determination of the preliminary score, S_p, and the cross-correlation score, XCorr, between the experimental spectrum and the corresponding theoretical one. The preliminary score in SEQUEST is generated by an empirical model ("empirical" means based on experience rather than theory) for evaluating potential peptide matches. The preliminary score is based on the number and total abundance of matched product ions in the experimental spectrum and is calculated by

$$S_p = \left(\sum_k i_k \right) n_i (1+\beta)(1+\rho) / n_t \qquad (1)$$

where $\sum_k i_k$ and n_i are, respectively, the total abundances and the total number of all matched peaks, β is a bonus for consecutive matched ion series (usually 0.075), ρ is a bonus for detecting immonium ions (usually 0.15), and n_t is the total number of ions in the theoretical spectrum. The S_p score in SEQUEST is empirically derived and is a very crude measurement of the quality of a peptide match. It is mainly used to pre-filter potential peptide matches so that only a limited

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

number of candidates are further evaluated by the more sophisticated and time-consuming scoring model based on cross-correlation.

The more sophisticated cross-correlation analysis is then applied to the top 500 potential peptide matches from the preliminary empirical model. The cross-correlation, XCorr, score is calculated based on an empirical model that assigns relative ion abundances in simulated tandem spectra of theoretical peptides. In the tandem mass spectrum for a theoretical peptide, the model assigns an abundance of 50 to all *b* and *y* fragment ions, an abundance of 25 to ions with *m/z* values ± 1 Da of the *b* and *y* ions, and an abundance of 10 to *a* ions and *b* and *y* ions with the neutral loss of a molecule of water or ammonia. The cross-correlation score, XCorr, is then calculated by

$$\mathrm{XCorr} = \sum_{i=0}^{n-1} x[i]y[i+\tau] \qquad (2)$$

where $x[i]$ and $y[i]$ represent ion abundances in experimental and theoretical spectra, respectively, and τ is the displacement value that compensates for systematic mass shifts to maximize the XCorr score. Thus, the XCorr score is a degree of cross correlation between the experimental spectrum and the theoretically predicted one. This score is used in SEQUEST as the main criterion to measure the quality of a peptide match. The XCorr score is biased in that peptides with longer length and higher charge states have higher scores than those with shorter length and smaller charge states. Different score thresholds need to be used for the peptides with different charge states and length in a search. However, there are no generally accepted rule-of-thumb values for the score thresholds in part because thresholds are dependent on the instrument type used to collect the data (e.g., linear ion trap [LTQ] and quadrupole TOF [QTOF] data should have different thresholds) and on other experimental conditions that affect fragmentation efficiency. The thresholds for a given false discovery rate can be determined by receiver operating characteristic (ROC) analysis using the target-decoy search strategy described in section 2.3.3.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Scoring Model in Mascot

Mascot, developed by Perkins et al. in 1999 (*32*), is a popular database searching program that incorporates a statistical/probabilistic scoring algorithm. Although the detailed algorithm of the scoring model has not been published, the model calculates the probability that a match between an experimental spectrum and the corresponding theoretical one is a random event. Therefore, the lower is the probability, the better the match. Mascot reports score as the negative common logarithm of the probability times 10. The probability-based score in Mascot is less biased than the descriptive scores in SEQUEST. Therefore, one score threshold can be used for all peptide matches in a search. The score threshold for significance of peptide matches depends on the size of a database and is calculated by

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

$$S_{\min} = -10\log_{10}\left(\frac{0.05}{n_{\text{entry}}}\right) \quad (3)$$

where $S_{\min}$ is the score threshold and n_{entry} the number of peptides of the database being searched that fall within the mass tolerance window about the precursor mass (*32*). The score threshold is calculated and reported by Mascot for each search.

Scoring Models in MassMatrix

The MassMatrix database searching program, developed by Xu et al., offers unique search functions to identify disulfide bond linkages and chemical crosslinks in proteins and peptides (*106*, *107*). In MassMatrix, proteins and peptides with disulfide bonds can be identified with high confidence without chemical reduction or other derivatization. MassMatrix has three statistical/probabilistic scoring models with no empirical parameters involved, which produce three statistical scores called pp, pp2 and pp_{tag}:

$$\text{pp} = -\log\left(\sum_{x=n_{\text{match}}}^{n} \frac{n!}{x!(n-x)!} p_2^x (1-p_2)^{n-x}\right) \quad (4)$$

$$\text{pp2} = -\log\left\{\sum_{x=0}^{n}\left[P(Y \geq I_{\text{match}} \mid x)\, C_n^x p_2^x (1-p_2)^{n-x}\right]\right\} \quad (5)$$

$$\text{pp}_{\text{tag}} = -\log\left\{\sum_{z=0}^{n}\left[P(t \geq n_{\text{T}} \mid z)\, \frac{C_M^z\, C_{N-M}^{n-z}}{C_N^n}\right]\right\} \quad (6)$$

where n_{match} is the number of matched experimental product ions, n is the total number of ions in the experimental spectrum, p_2 is the probability that any single matched product ion could be random, I_{match} is the total abundance of matched experimental ions, Y is the total abundance of randomly matched experimental ions, n_{T} is the number of sequence tags of matched theoretical ions, t is the number of sequence tags of randomly matched theoretical ions, N is the total number of theoretical product ions, and M defines the number of theoretical b/y non-neutral loss ions (*106*, *107*). Among them, pp and pp2 scores are sensitive to high mass accuracy and explicitly take mass accuracy into account. High mass accuracy not only reduces false positives, but also improves the scores and reliability of true positive matches. The pp_{tag} score is a sequence tag-based score and the pp2 score is an abundance-based score. These two scores take advantage of the information on sequence tags of peptides inferred from tandem mass spectra and the information on peaks' abundances in experimental tandem MS data. All three statistical scores in MassMatrix are unbiased and not dependent on peptide length and charge. Therefore, each score needs only one threshold for all peptide matches in a search. The score thresholds should be dependent on the size of

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the database and the mass accuracy of precursor masses. As a rule of thumb, a peptide match with the maximum of pp and pp2 scores bigger than 6.0 and a pp_{tag} score bigger than 2.0 is considered a good quality match.

Evaluation of Algorithms for Database Searching

Many database searching programs have been developed, and these programs perform differently, especially when dealing with low mass accuracy tandem MS data. Therefore, it is crucial to employ a standard method to compare and evaluate performance of different database searching programs. The performance of database searching methods normally is evaluated by a ROC analysis. In this analysis, tandem MS data sets, collected from complex proteome samples, are searched against a combined database containing both a target protein sequence database and a decoy protein sequence database (*111*). Target protein sequences could be present in a real proteome sample, whereas decoy protein sequences are simulated protein sequences that are not present in that sample. Decoy sequences normally are created by reversing or randomizing the amino acid sequences of proteins in either the target database or a large protein database such as the human protein database. Peptide matches from each database search are categorized into two groups: true positives, when hits represent the actual identity of related tandem mass spectra, and false positives, when matched sequences are the wrong identification of related tandem mass spectra. The ROC analysis is based on two assumptions: (i) true peptide matches only can be present in the target protein sequences; and (ii) false peptide matches are evenly distributed across the whole database (i.e., the target and the decoy databases combined). The first assumption is reasonable because decoy protein sequences are simulated and completely uncorrelated with proteins in a sample. The second assumption also is feasible based on the principle that false peptide matches are due to random events (*111*).

For the case where the size of the target database equals to the size of the decoy database, the total number of false positive (FP) peptide matches is calculated by multiplying the number of peptide matches, *m*, in the decoy database by two. The number of true positives (TPs) then is calculated by subtracting the number of peptide matches in the decoy database from the number of peptide matches in the target database.

$$\begin{cases} \mathrm{TP} = m_{\mathrm{target}} - m_{\mathrm{decoy}} \\ \mathrm{FP} = 2m_{\mathrm{decoy}} \end{cases} \quad (7)$$

For the case where the size of the target database is much smaller than the size of the decoy database, the total number of false positive peptide matches equals the number of peptide matches in the decoy database, and the number of true positives equals the number of peptide matches in the target database.

$$\begin{cases} \mathrm{TP} = m_{\mathrm{target}} \\ \mathrm{FP} = m_{\mathrm{decoy}} \end{cases} \quad (8)$$

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

The TP and FP of a search is based on the score threshold. The higher the threshold is, the smaller TP and FP are. ROC curves that evaluate the performance of a database search program on a data set are created by plotting TP against FP as the score threshold decreases. An ideal database search program should have all true positives with scores higher than all false positives. Therefore, it will have a ROC curve with a right angle, as shown in Figure 7. In ROC analysis, a curve toward the left indicates high specificity and a curve toward the top indicates high sensitivity (Figure 7) (*91*).

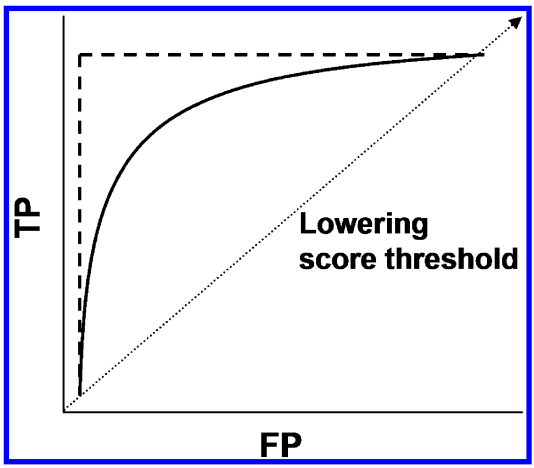

Figure 7. A representation of ROC curves for an ideal (dashed line, both specificity and sensitivity are equal to 1.0) and a real (solid line, both specificity and sensitivity are less than 1.0) database searching program. The curves are created by plotting TP against FP as the score threshold decreases along the dotted line with arrow.

Informatics for Peptide Mapping

Difference between MS-Based Proteomics and Peptide Mapping

The quote "try to learn something about everything and everything about something" of the famous British biologist and educator of the 19th century, Thomas H. Huxley, in our opinion well defines the applications of LC-MS/MS to the proteomics and the therapeutic protein analysis problems. Proteome investigation targets detection, identification, and quantitation of as many proteins as possible in a complex sample. Detection of proteins often is based on a limited subset of representative peptides, with three peptides commonly deemed sufficient for reliable identification of the corresponding proteins. In a sense, these limited subsets of peptides serve as "something" that is known about a large number of proteins or "everything" that defines the proteome. Therapeutic applications of peptide mapping for reasons of reproducible manufacturability, safety, and efficacy demand a detailed knowledge on every peptide, or "everything," originating from a relatively pure single protein sample. In order to establish a

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

well-characterized status of a protein drug, a detailed characterization needs to be performed to ensure that the entire sequence is covered and confirmed and that variants are detected and identified. Potential safety, efficacy, and production consistency concerns related to the low-level variants necessitate manufacturers of biotherapeutic proteins to pay careful attention to the reliable detection of these variants.

Peptide mapping is widely used for characterization of therapeutic proteins because of its ability to provide site-specific information at the primary sequence level. Furthermore, peptide mapping with LC-MS is often viewed as the "front line" technology for monitoring and characterizing the quality of biotherapeutics. The ability of the method to monitor multiple attributes in a single assay make it an especially attractive approach. This is increasingly important as the development paradigm shifts from method-centric to attribute-centric strategies, as directed by principals of the quality by design (QbD) and ICH Q8/Q9/Q10 (*112*).

Peptide mapping with LC-MS/MS can provide incredibly detailed information on the protein in question with the ability to fulfill important tasks such as:

- **Sequence confirmation:** primary sequence verification and detection of potential sequence variants due to either mutations or misincorporations.
- **Modification monitoring:** detection and quantitation of product quality attributes (PQAs) such as glycosylation, oxidation, deamidation, and other PTMs.
- **Impurity monitoring:** detection of process-related impurities such as host cell proteins (HCPs) or Protein A.

It is therefore apparent that LC-MS/MS requires rather sophisticated data analysis workflows to process large volumes of information-rich data.

As we discussed earlier in the chapter, the goal of peptide mapping is to analyze a sample in which a single protein (biotherapeutic product) constitutes nearly 100% of that sample's content, whereas MS-based proteomics deals with a mixture of hundreds or thousands of proteins in a same sample at the same time. At a glance, identifying "hundreds of proteins" is a more challenging task than when dealing with a "single protein" biotherapeutic sample. In reality, however, this is not necessarily true considering the differences in the scopes of the two applications. Although MS-based proteomics primarily focuses on the sequence-level information to infer protein identities from a subset of identified peptides, peptide mapping focuses on providing all attainable structural details and a heterogeneity profile for a single protein. A potential presence of *a priori* unknown, low-level modifications can complicate that task. The difference between the two approaches also suggests that the requirements for data analysis can be drastically different, and that most of the well-established proteomics search engines (such as SEQUEST (*80*) and Mascot (*32*)) may not necessarily fit the goals of peptide mapping in the drug development environment. These and other popular proteomics search engines typically are designed for large-scale sequence identification from large databases with limited emphasis on in-depth analysis of multiple modifications. This is in part due to the fact that most

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

proteomic samples are complex mixtures. Modified peptides, which are usually low in abundance, have a rather small chance to provide tandem MS data. In addition, the search space (the number of potential peptide candidates) increases exponentially as more modification types are considered, which significantly hinders the speed and confidence of the search. Although users do have the ability to specify several fixed or variable modifications (such as methionine oxidation and cysteine alkylation), they need to know beforehand which type of modifications may be present in the sample or when modification selection is governed by a specific task, for example, when analyzing for phosphorylation.

Fortunately, there exists a class of programs specifically aimed at modification identifications that use "blind modification search" algorithms (*113–117*). The "blind modification search" algorithm is designed for extensive characterization of modifications (i.e., modification analysis), not requiring users to have *a priori* knowledge about modifications in the sample. To overcome the search space constraint, the searched database must be limited in size, and special algorithms need to be implemented to improve the search efficiency, such as relying on sequence tags to build partial peptide sequences from MS/MS spectra and using modifications to explain unexpected mass shifts, as used in InsPecT (*113*), MODi (*114*), and Byonic (*115*) software. It is claimed that sequence tags can improve search speed by several orders of magnitude. However, other algorithms simply iterate through a larger number of modifications, such as the ones present in the unimod protein modifications database (www.unimod.org), to enable in-depth modification identifications. This strategy is implemented in Mascot Error Tolerance Search (ETS) (*116*) and Peaks PTM (*117*). Clearly, the "blind modification search" algorithms better fit the needs of protein characterization by peptide mapping where a large number of modifications need to be considered.

In the remaining parts of this chapter, a more detailed discussion on the applications of informatics to peptide mapping for characterization of biotherapeutics will ensue. Challenges, data processing workflows, and false discovery rate determination will be discussed with the aim to better understand and verify search results. Peptide mapping data, collected for the NISTmAb, will be used to support the discussion. LC-MS/MS data were collected using Acquity ultrahigh performance liquid chromatography (UPLC) (Waters, Milford, MA) coupled online with an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA) and searched with MassAnalyzer software (*115*).

Challenges in Analysis of Peptide Mapping Data

The main challenge of a peptide mapping experiment becomes apparent when characterizing low- and trace-levels species such as modifications, sequence variants, and impurities. With the exceptional sensitivity offered by modern mass spectrometers, it is not uncommon to detect thousands of species in a single run. In the example of the NISTmAb, 4535 ions ranging several orders of magnitude were detected with a single-to-noise ratio (S/N) of 20. At the same time, a theoretical tryptic digest of a typical IgG1 molecule can generate only 100 to 200 peptides, even accounting for a number of missed cleavages. Understanding the properties and origins of the most commonly detected ions can greatly benefit our

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

interpretation of the data as well as minimize the chances of false identification. In general, the complexity of a peptide mapping data can originate from the following five categories.

Unmodified Peptides

Refers to proteolytic peptides, either fully cleaved, partially cleaved or mis-cleaved. In a non-reduced map there also will be disulfide-linked peptides. These peptides can be predicted easily from the known protein sequence and should account for the majority of the ion signal in peptide maps.

Modified Peptides

- *Post-translational or in-process induced modifications*. Originating from post-translational processes or induced by manufacturing and post-manufacturing processes, these modifications are among the primary targets of peptide mapping. The common ones include oxidation, deamidation, glycation, glycosylation, N-terminal pyroglutamate, C-terminal lysine, hydroxylation, and so forth. Many of these variants can have a direct impact on the potency and quality of the product. In a non-reduced map, disulfide linkages also should be considered.
- *Sample preparation artifacts*. Sample preparation can introduce many unexpected artifacts such as transpeptidation (*118*) or unexpected carboxymethylation on lysine or histidine. Inappropriate sample preparation also can induce artificial levels of deamidation, oxidation and glycation (*119*).
- *In-source induced artifacts*. Electrochemical and chemical reactions occur during the electrospraying process because of the high voltages and heat used in the source. The most common ones are adduction of metal ions (Na^+, K^+, Fe^{2+}), losses of neutrals (NH_3, H_2O, CH_2O), in-source fragmentation (labile amide bonds, glycosidic linkages), and in-source oxidation (tryptophan, tyrosine, and methionine residues) (*120*). Among these, in-source oxidation of tryptophan and tyrosine can be very complex and involves many intermediate species. Inappropriate data analysis can potentially misinterpret these in-source artifacts as being due to PTMs or sequence variants. This issue highlights the importance of using the retention time information when interpreting peptide mapping data because in-source artifacts chromatographically "co-elute" with the precursor peptide. In addition to the in-source induced artifacts, the elevated temperatures and low pH required for chromatographic separation also can contribute to erroneous results; for example, on-column methionine oxidation can significantly overestimate oxidation levels in a biotherapeutic protein.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Sequence Variants

Mutations and mis-incorporations are the prime examples of sequence variants. These usually are observed as a mass shift on a certain amino acid residue and identified as variants using an amino acid substitution table (*121*). Detection of these variants is discussed in the Sequence Variant chapter/Volume 2, Chapter 2).

Product Impurities

Notable examples include HCPs, process-related impurities (e.g., Protein A), and protease autolysis products. Typically, these impurities are not searched along with the comprehensive search for modifications and variants of a target protein, which as was discussed above limits the size of the database. Instead, a separate search (resembling an MS-based proteomics approach) should be performed against a much larger database containing all of the host cell and process-related protein impurities.

Background Ions

Background ions would be present in the environment and chromatographic solvents. These ions usually are observed consistently throughout an LC-MS analysis run and can interfere with the data-dependent selection of low-abundance peptide variant species. Some other impurities, such as surfactants used as excipients in formulations, could carry through all of the sample preparation steps and can contribute to the ion signal during LC-MS. The exclusion of singly charged species is a typical approach to avoid fragmenting background ions during peptide mapping.

The complexity of these ions poses a significant challenge for the software to identify them. For instance, dozens of glycan forms need to be considered for a single glycosylation site. Clearly, this kind of detailed analysis is beyond the scope of traditional proteomic search engines, and even for the specifically designed peptide map software, many ions remain unidentified.

Figure 8 compares the number of identified and unidentified ions from the NISTmAb following data analysis with MassAnalyzer (*121*). It is clear that although the major peaks in the peptide map can be well identified, the majority of lower abundant peaks are not assigned. It should be noted that ions with abundances in the 10^4 to 10^5 range have more than 50% of peaks that remain unidentified. This abundance range is where potential PQAs of interest reside. This example further demonstrates that software and data analysis are the bottleneck in the informative use of peptide mapping. Also, the presence of more contaminating and artifact ions (those discussed above) challenges the software algorithms to reliably extract LC-MS features relevant to the molecule of interest and increases the risk of missing important information. Thus, great care always

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

should be exercised to minimize the presence of contaminants and artifacts in a peptide map.

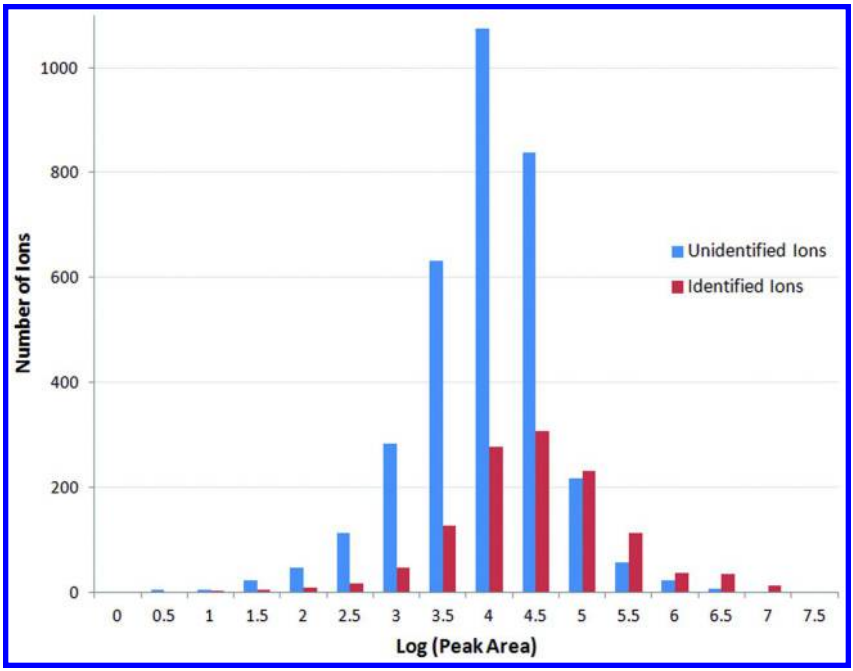

Figure 8. Peak area distribution for identified and unidentified ions in the NIST IgG1 mAb. "Peak" is defined as a mass signal with signal-to-noise ratio (S/N) above 20 and with a well-defined charge state (based on isotopic information). Data searched with MassAnalyzer. In the figure, major peptide peaks are in the 10^6 to 10^7 abundance range, whereas modified peptides are two to three orders of magnitude lower (10^4 to 10^5 abundance range or around 1% in their relative abundance levels).

Software for Processing Peptide Mapping Data

The analysis of peptide mapping data heavily relies on the use of software. For an end user, it is not essential to know every detail about how a certain program works, but a good understanding of some of the components and algorithms involved is beneficial. For instance, knowledge of the identification algorithm help one to decide whether a peak identified without MS/MS should be trusted. The following is a brief discussion for the typical components and algorithms in the software for protein characterization, namely peak detection, identification, quantification, and differential profiling.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Peak Detection

The purpose of peak detection is to generate a list of "peaks" for submission to database searching. In cases of mass spectrometers with low resolution such as ion traps, the charge state often cannot be unambiguously identified and the "peak" merely represents a signal above a certain S/N threshold. The search engine then needs to consider multiple charge states during identification, which typically is error prone. In cases when MS instruments offer high resolution, such as Orbitrap and QTOF, the "peak" represents an isotope cluster with clear charge state information, so that the precursor mass is unequivocally defined. The first peak in the isotope cluster is called the monoisotopic peak, and its *m/z* is used to calculate the monoisotopic precursor mass. This mass is the starting point for the identification: selection of possible peptide candidates within a defined mass tolerance.

Although peak detection is a relatively straightforward process, the selections of the correct monoisotopic peak can be prone to errors. This is especially true for peptides with low abundance or high molecular weight whose monoisotopic peak could be small or overlapping with other clusters. Errors in monoisotopic peak selection can introduce mass shifts of 1.003 Da or multiples of 1.003 Da in assignment of precursor mass, resulting in incorrect identifications or no identifications. One typical consequence of the erroneous identification of the precursor mass would be assignment of "deamidation" status to that peptide (mass shift of 0.985 Da), and in sequence variant analysis, these shifts also can contribute to a number of false positives. Although there is no direct report that compares peak detection algorithms, some software has been acknowledged to be able to pick monoisotopic peaks accurately. Among these, Mascot Distiller (*116*) solves this issue by finding the best fit of the experimental isotope distributions to theoretical distributions.

Identification

Identification is the process of assigning identifications to precursor ions established during the peak detection process. As discussed in Section 2 of this chapter, different algorithms have their unique features to identify a peptide and evaluate the quality of the match. Briefly, the workflow involves *in silico* digestion of a protein sequence (as present in the database), matching experimental precursors with sequence candidates whose masses fall within a specified mass tolerance window, and comparison of the corresponding experimental MS/MS spectra with predicted to determine the best match. In contrast to MS-based proteomics, the number of potential candidates during peptide mapping experiments is very limited, and all of the major peptides in the sample of a recombinant protein are highly predictable based on the enzymatic specificity of the digest. For most of the proteolytic peptides, as well as those with well-characterized modifications, reliable identification can be done solely from the mass of a peptide if a major peak is observed in the chromatogram. But for less-anticipated modifications (e.g., tryptophan oxidation) or sequence variants,

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

precursor mass alone is not sufficient to discriminate between the multiple candidate possibilities. In these cases, MS/MS spectra are necessary to verify the sequence or to locate modification sites. Retention time also can be extremely valuable as an orthogonal identifier or an extra constraint, although its reliability depends highly on the retention time prediction models. Sequence Specific Retention Calculator (*122*) and Normalized Elution Time (NET) predictor, based on using an analytical neural network approach (*123*), are the popular algorithms to predict elution times of peptides taking into account their hydrophobicities. As these methods mature and become more accurate, software packages that employ peptide retention prediction models undoubtedly will simplify analysis of peptide mapping data by contributing to the reduction of false positives, specifically by rejecting identifications whose predicted retention time does not agree with the experimental retention time.

For the purpose of identification of modifications, most peptide mapping software belong to the category of "blind modification search" algorithms, as discussed above. The search is realized by the use of sequence tags or multiple iterations through known modifications. As a general practice, modification identification in a peptide map utilizes precursor mass difference to determine the type of the modification, which is followed by MS/MS verification to locate the modification exact site. Details for several algorithms will be discussed later in this chapter.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Quantitation

Quantitation of modified or variant peptides is an integral part of a typical workflow involving peptide mapping data analysis because the quantitative information often serves as supporting evidence upon which the significance of a given modification can be evaluated. Quantitation from peptide mapping data usually is based on the integration of extracted (or reconstructed) ion chromatograms (XICs) for the native and variant peaks of interest; dividing the peak area of the variant peptide by the total peak area of its native and variant forms gives the relative amount of a certain modification in a protein. It should be noted that the accuracy of such an abundance calculation is significantly affected by the ionization efficiency and width of the dynamic range of an MS instrument. It should not be surprising that there might be a discrepancy between values obtained by orthogonal techniques. For instance, it was reported that when phosphorylation occurs on threonine in a peptide AATAAR, the peptide signal intensity dropped more than 50% due to poorer ionization (*124*). In this case, the quantitative amount of phosphorylation is significantly undervalued. It always should be kept in mind that the quantitative information obtained from peptide mapping data is relative and depends on ion properties. Nevertheless, peptide map-based quantitation remains popular and often is the only available approach to provide site-specific quantitative information on variants in biotherapeutic proteins.

Attention also should be paid to the peak integration approach, especially whether multiple isotopes are combined and whether multiple charge states

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

are considered. In a simple and commonly used practice, XICs for peptide ions are generated using either the monoisotopic or the most abundant isotope peak. However, this process tends to be inaccurate for modifications such as glycosylation inducing significant mass shifts relative to the mass of the peptide or affecting charge (such as amino acid substitution involving charge-bearing residues) because these modifications can change the isotopic or charge state distributions. Thus, a recommended practice consists of summing several isotopic peaks (*121*) and combining multiple charge states of peptides to obtain more accurate peak integration.

Differential Profiling

Comparison of two or more samples often is required during the development of biotherapeutics, targeting the screening of multiple cell line clones for sequence variants, the evaluation of PTM status, the assessment of comparability, and stability monitoring. Of course, performing LC-MS/MS followed by database searching for each sample is a feasible way to detect the differences. However, this approach tends to miss low-abundant species because it relies on data-dependent acquisition during which an MS/MS event is only triggered on ions that pass a certain pre-set criterion (intensity, charge state, m/z range, retention time window, etc.). Precursor ions that cannot trigger an MS/MS event would remain unidentified.

An alternative approach to detect the difference between samples is to overlay their chromatographic profiles (e.g., UV profile) and examine them visually to detect subtle differences that will allow "red flagging" certain samples or triggering further evaluation. The full scan mass spectra at the MS level also can be used for this purpose, offering better sensitivity and additional ability to resolve co-eluting peptides compared to UV detection. Tandem mass spectrometric data, if available, can further reveal the identity for the differentially displayed peak. However, extracting the difference between MS-level data is not a straightforward task and heavily depends on background noise detection and removal, intelligent peak picking and extraction of molecular features, clustering of isotopic and charge-state data and deconvolution of MS spectra, and alignment and normalization of multiple LC-MS profiles.

Alignment of chromatographic peaks is an essential component for successful comparative analysis of peptide maps. By correcting a shift in retention time between multiple runs, peak alignment allows for a direct comparison of different samples and saves resources by searching and sharing the information on common precursors between the runs. Most algorithms adopt non-linear correction by dividing the chromatogram into several slices and aligning each of them. Recently, Zhang reported using a "divide-and-conquer" algorithm to iteratively correct for retention time shifts until the last peak is aligned (*125*). It was implemented in MassAnalyzer software, and it was demonstrated that this algorithm can correct retention time shifts up to five peak widths with a low error rate. Alternately, in the approach described by Zeck et al., the SIEVE program, was employed to align and differentially profile multiple samples, targeting the discovery of low-level

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

sequence variants (*126*). In this latter application, only peaks showing differential abundances between several runs, based on the evaluation of MS-level data alone, were further interrogated for the presence of sequence variants. Although the advantage of the peak alignment is obvious, technically it is a complicated feature that is only available in a limited number of software packages.

Major MS instrument vendors now provide software solutions for comprehensive MS-based differential analysis and statistical evaluation of LC-MS data, including Profile Analysis (Bruker Daltonics), Mass Profiler (Agilent), Protein Pilot (AB Sciex), SIEVE (Thermo Fisher Scientific), and Progenesis QI (Waters Corporation). Because many of these tools are designed and optimized to handle unique features of specific hardware, their use is often data file format-dependent. Several packages, including Expressionist (Genedata) and Progenesis QI (Waters Corporation), recently emerged as stand-alone solutions, allowing data files from multiple vendors to be processed. Many of these packages interface with one or several database searching programs, such as SEQUEST or Mascot, to provide sequence information and characterization details on the detected differences.

The above-mentioned components and algorithms are important but far from enough to build a workable peptide mapping program. An effective scoring scheme, the incorporation of empirical rules, a user-friendly interface, the ease of navigation between MS and MS/MS levels of data, visualization, and annotation all influence the user's experience with a software package. The ideal software not only should allow users to visualize the results, but should provide the opportunity to verify the results easily. The following discussion presents several programs that are frequently used for peptide mapping data analysis based on the literature.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Software for Peptide Mapping

Proteomic Search Engines

Popular MS-based proteomics software packages such as SEQUEST (as stand-alone software or as part of the Proteome Discoverer from Thermo) (*80*) or Mascot (*32*) have been applied successfully to peptide mapping. The full potential of these software packages is realized when screening for HCP process-related impurities in biotherapeutics, when a large database is searched against. However, there are limitations on the use of these software packages because these packages usually limit the number of modification types that can be concurrently searched to around five. This often may be insufficient to cover the scope of a typical peptide mapping analysis, which aims to gain a comprehensive knowledge of the modification profile of a biotherapeutic protein, including numerous glycosylation types and the occurrence of any unexpected modifications or sequence variants. Software packages that use blind modification search algorithms are better fitted for comprehensive modification profiling at the expense of a small database size that can be handled.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Mascot Error Tolerant Search

Mascot Error Tolerant Search (ETS) is a blind modification search program for database matching of uninterpreted tandem MS data can be enabled on a limited subset of proteins following a regular Mascot database search engine (*116*). Utilizing partial sequence tag filtering, it conducts identification of unexpected modifications by iterating through a long list of chemical and post-translational modifications, as well as considering a residue substitution matrix (such as ones contained in the unimod modification database) (*127*). This process fits the purpose of peptide mapping well and in combination with Mascot Distiller for raw data processing enabled researchers from Genentech to conduct in-depth sequence variant analysis of biotherapeutic monoclonal antibodies (mAbs) (*128*). A limitation of the Mascot ETS approach is that it does not provide quantitative data on the relative abundances of variants and its output is overwhelmed by a number of false positives. False positives require great effort for manual data verification because the ETS is unable to conduct a decoy database search to verify the confidence of assignments. In addition, Mascot ETS does not handle the analysis of modifications involving glycosylation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

MassAnalyzer

Developed at Amgen by Z. Zhang, MassAnalyzer is a software package specifically designed for characterization of recombinant proteins and mAbs in particular (*121*). Unlike proteomics software, the search database used is restricted to a few protein sequences. The main feature of the MassAnalyzer is that it predicts ion intensities in a MS/MS scan. Modifications are searched by comparing an experimental spectrum with the predicted spectra of all of the theoretical peptides from that mAb, followed by annotating the precursor mass difference between the experimental and theoretical as a modification. Besides the unique algorithms, multiple empirical rules are incorporated into MassAnalyzer to minimize the chance of false positives. A significant advantage of MassAnalyzer is that it readily provides quantitative information on found modifications, including glycosylation, and the data can be verified easily by navigating between MS and MS/MS spectra. It was demonstrated that with the same number of false positives, MassAnalyzer identified twice as many peptides compared to the regular Mascot search (*121*).

Byonic

This software from Protein Metrics is a novel algorithm utilizing unrestricted blind modification for database searching (*115*, *129*). The Byonic software uses sequence tags extracted from experimental spectra to find gaps in a sequence of a peptide and explain these as potential modifications. The use of sequence tags for the unrestricted blind modification searching in a sense is a hybrid approach between *de novo* sequencing and database searching. Whereas *de novo*

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

sequencing makes it possible to discover unexpected modifications, database searching narrows down the search space and reduces the rate of false positives. It is claimed that Byonic identifies 20% to 300% more peptides than most search engines at the same false discovery rate (FDR).

BiopharmaLynx

Developed by Waters, BiopharmaLynx is an automated software to process LC-MS data for intact proteins and peptide maps (*130*). Unlike other algorithms, it involves no data-dependent MS/MS but primarily depends on precursor masses and retention time, as well as MS^E (a way to fragment all the ions without relying on data-dependent selection), for peak identification. It also incorporates extensive visualization and annotation tools that can demonstrate the difference between samples.

MassHunter BioConfirm

Like BiopharmaLynx, MassHunter bioConfirm (Agilent) software is designed for biopharmaceutical applications to confirm the identity of a protein and to identify variants (*131*). It typically is used with Agilent TOF and QTOF instruments, enabling the analysis of intact mass and LC-MS/MS data using the same software package. With the information-rich user interface, analysts can automatically compare mirror plots of multiple samples, generate sequence coverage maps, and highlight modified locations. The qualitative comparison feature is a great tool to view XIC, full MS, and MS/MS in the same window for data verification.

Understanding the Confidence of the Search Results

Although modern software packages greatly facilitate the analysis of peptide mapping data, a certain rate of false positives is still inevitable in the reported results. In fact, one of the most time-consuming steps during data interpretation following the Mascot ETS search, for example, is to filter out an overwhelming number of false positive assignments. For analysts, a good understanding of the confidence of search results is essential to ensure the authenticity of interpreted data. For database searching algorithms, a common approach to assess the confidence of results is by determining the FDR, as was discussed in Section 2 of this chapter. The following discussion will emphasize the application of FDR methodology to peptide mapping. To remind ourselves, 1% FDR means that for each 100 identified peptides, 1 is likely to be wrong. An identification without FDR is usually meaningless because there is no way to know whether it can be trusted. In practice, it is often more challenging to provide a convincing argument that the software-generated identification is correct than assigning that identification in the first place. Although the target-decoy database search strategy (*132*) is simple to implement and works quite well in MS-based proteomics

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

studies, a researcher must proceed with caution when applying it to peptide mapping.

A target-decoy search is a statistical approach and therefore requires large "search space," defined by using a large database and a large data set, to provide accurate FDR estimation. It was reported that when the number of proteins in the database is less than 1000, the FDR estimation can be inaccurate (*133*). Zhang also mentioned that in MassAnalyzer results, the confidence level determined by using score distribution is 80% (similar to a 20% FDR), whereas the FDR determined using a decoy search is only 0.4%. This difference in confidence implies that a small search space may underestimate FDR (*121*). This emphasizes the need for extensive manual verification of the search results to remove false positives, especially for critical data and PQAs.

Another challenge for peptide mapping is that the FDR will not be consistent for all types of peptides. In an insightful discussion on FDR, Chalkley et al. emphasized that in a target-decoy database search, random matches always reflect the composition of the search space (*134*). For instance, when searching for phosphorylation in proteomics, the authors noticed that the FDR for phosphopeptides is actually 7.5%, whereas the FDR for the total identifications is set at only 2%. This difference is due to the fact that when specifying phosphorylation as variable modifications, theoretically the majority of the peptide candidates are phosphorylated. When a random match happens, it has a higher chance to match to a phosphopeptide. The issue becomes more profound in peptide mapping where a much larger number of modification types are often considered. If a rough estimate is that each variable modification doubles the search space, then by allowing 10 modification types, the search space becomes as large as 2^{10} (1024 combinations), whereas only 1 out of the 1024 is an unmodified peptide. This means that when a random match happens, the resulting identification has a 99.9% chance of being a modified peptide. In the MassAnalyzer search results of the NISTmAb, 1142 peptides were identified at an 80% confidence level, and 741 of them were with modification. By searching the decoy database, 131 peptides were identified, and 123 peptides were classified as modified. Whereas the total FDR is only 11% (131/1142), the FDR for an unmodified peptide is only 2% (8/401), compared with 17% (123/741) for modified peptides. Clearly, greater attention should be paid to the modifications during data verification. Orthogonal criteria, such as mass error, retention time, modification type, and peak size, should be used in addition to MS/MS to verify whether the identification is true or false.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

FDR in Sequence Variant Analysis

One extreme of the above discussion is the analysis for sequence variants, which is equivalent to searching more than 100 different kinds of variable modifications. When a random match is encountered, it has almost a 100% chance of matching to a peptide with a sequence variant. Even using various kinds of filters to remove false positives, it is not uncommon to see more than an 80% false positive rate for peptides with potential sequence variants. This number of

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

hits would require a significant amount of time on the part of the researcher to manually verify the MS/MS spectrum and orthogonal criteria, making sequence variant analysis among the most challenging tasks in protein characterization (*123*).

FDR in HCP Analysis

Detecting HCP in mAb samples is essentially a proteomics experiment, but requiring a very wide dynamic range. Because abundances of mAb peptides are dominating, it is common to observe that nearly all of the identified peptides are from that mAb and only a few belong to HCPs, despite the database searched actually consisting of a single mAb sequence along with thousands of HCP sequences (e.g., all of the known sequences from the Chinese hamster ovary cells). The question is if a software identified 1000 peptides at 1% FDR and 10 of them are HCP peptides, can these HCP identifications be trusted? Clearly not. The database composition determines that a random match has an almost 100% chance of falling onto an HCP sequence. These 10 HCP identifications are not statistically significant because at 1% FDR, the expectation is about 10 false positives or random matches. It is important to use orthogonal criteria (like number of unique peptides per protein, mass accuracy, retention time, etc.) to further verify any HCP identifications.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Concluding Remarks

The term "bioinformatics" originally was introduced by P. Hogeweg and B. Hesper in 1970 to refer to the "study of informatic processes in biological systems" (*135*). In its original concept, the focus of bioinformatics was on information processing as a useful metaphor for understanding living systems (e.g., information accumulation during evolution, information transmission from DNA to intra- and intercellular processes, the interpretation of such information at multiple levels). Later, in the 1980s, the term "bioinformatics" transformed to refer to a dynamic interdisciplinary scientific field focusing on developing computational methods for storing, retrieving, organizing, analyzing, and comparing biological data.

While recognizing the diverse range of tasks of modern bioinformatics, the focus of this chapter was on its use for MS-based analysis of proteins, which is one of the fastest growing and arguably the most popular technology in modern protein research. Based on a literature analysis, starting from the 1980s, mass spectrometry has experienced a nearly exponential growth in a number of its applications to protein analysis, as shown in Figure 9. Primarily coinciding with the availability of mass sequence data and whole genomes for a number of species, the influence of bioinformatics solutions to analysis by mass spectrometry has been growing steadily. Nowadays, informatics plays a center role in data collection, management, and analysis. The development of bioinformatics tools has influenced and became an integral part of instrument development.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Here we reviewed the basic concepts of searching and matching mass spectrometry data to known sequences in a database, as well as popular bioinformatics algorithms, as applied to both MS-based proteomics and peptide mapping of recombinant proteins. MS-based proteomics applications have been the major driver for the development of sophisticated computer-aided algorithms for the analysis of mass spectrometry data. Our discussion started with an introduction to the popular software solutions available to assist researchers in the field of proteomics.

The second part of the chapter focused on the state of the informatics for peptide mapping applications as one of the most insightful tools for the characterization of recombinant proteins. While recognizing methodological similarities between the MS-based proteomics and peptide mapping of recombinant proteins applications and, thus, allowing the use of bioinformatics tools interchangeably, we pointed out the principal differences in the tasks and deliverables of the two applications. The current utility of the informatics tools for peptide mapping is far from ideal and requires a significant amount of manual involvement and interpretation. The manual process of data analysis is time- and resource-consuming. For example, verification of the Mascot ETS results for the purpose of sequence variant analysis in a recombinant mAb can take multiple days or weeks to complete, while sample preparation and data acquisition takes a couple of days at the most. This limitation in data analysis requires extensive experience on the part of the analyst, often prohibiting high-throughput application of peptide mapping. As more orthogonal information, chemical knowledge, and empirical rules are incorporated into modern software, we anticipate the efficiency of this technique will be significantly improved. After all, it is the only technique that can provide site-specific detail on a protein.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

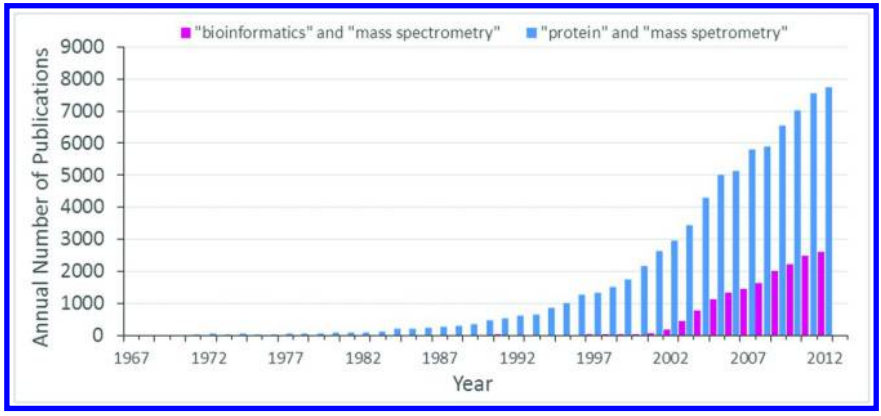

Figure 9. Annual number of publications with keywords "bioinformatics" and "mass spectrometry" and "protein" and "mass spectrometry" according to the publication database Web query with the National Center for Biotechnology Information (NCBI) (www.pubmed.gov).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Disclaimer

Hua Xu is an employee of PepsiCo, Inc. The views expressed in this chapter are those of the author and do not necessarily reflect the position or policy of PepsiCo, Inc.

Acknowledgments

W.L. and O.B. acknowledge Dr. Robert J. Duff and Dr. Izydor Apostol from Amgen Inc. for critical reading of the manuscript and providing valuable comments and revisions. H.X. acknowledges Dr. Eric Milgram and Dr. John St. Peter from PepsiCo, Inc. for helpful discussion and critical reading of the manuscript.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

References

1. Kanehisa, M.; Bork, P. *Nat. Genet.* **2003**, *33*, 305–310.
2. Van Riper, S. K.; de Jong, E. P.; Carlis, J. V.; Griffin, T. J. *Adv. Exp. Med. Biol.* **2013**, *990*, 1–35.
3. Cappadona, S.; Baker, P. R.; Cutillas, P. R.; Heck, A. J.; van Breukelen, B. *Amino Acids* **2012**, *43*, 1087–1108.
4. Wilkins, M.; Gooley, A.; Williams, K.; Humpherey-Smith, I.; Cordwell, S.; Appel, R. D.; Sanchez, J. C.; Golaz, O.; Pasquali, C.; Hochstrasser, D. F. In *Towards the Protein Genome: Rapid Identification of 2D Spots by Amino Acid Analysis*, Siena Meeting on 2D Electrophoresis, Siena, Italy, 1994; pp 35−36.
5. Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. *Genet. Eng. Rev.* **1995**, *13*, 19–50.
6. Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. *Electrophoresis* **1995**, *16* (7), 1090–1094.
7. Kellner, R. *Fresenius J. Anal. Chem.* **2000**, *366* (6-7), 517–524.
8. Issaq, H. J. *Electrophoresis* **2001**, *22*, 3629–3638.
9. Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269–295.
10. Alaiya, A. A.; Franzen, B.; Auer, G.; Linder, S. *Electrophoresis* **2000**, *21* (6), 1210–1217.
11. Stulik, J.; Hernychova, L.; Porkertova, S.; Knizek, J.; Macela, A.; Bures, J.; Jandik, P.; Langridge, J. I.; Jungblut, P. R. *Electrophoresis* **2001**, *22* (14), 3019–3025.
12. Masumori, N.; Thomas, T. Z.; Chaurand, P.; Case, T.; Paul, M.; Kasper, S.; Caprioli, R. M.; Tsukamoto, T.; Shappell, S. B.; Matusik, R. J. *Cancer Res.* **2001**, *61* (5), 2239–2249.
13. Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M. *Nature Med.* **2001**, *7* (4), 493–496.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

14. Hondermarck, H.; Vercoutter-Edouart, A. S.; Revillion, F.; Lemoine, J.; El-Yazidi-Belkoura, I.; Nurcombe, V.; Peyrat, J. P. *Proteomics* **2001**, *1* (10), 1216–1232.
15. Lubman, D. M.; Kachman, M. T.; Wang, H.; Gong, S.; Yan, F.; Hamler, R. L.; O'Neil, K. A.; Zhu, K.; Buchanan, N. S.; Barder, T. J. *J. Chromatogr. B* **2002**, *782*, 183–196.
16. Giometti, C. S.; Williams, K.; Tollaksen, S. L. *Electrophoresis* **2000**, *18* (3-4), 573–581.
17. Bergman, A. C.; Benjamin, T.; Alaiya, A.; Waltham, M.; Sakaguchi, K.; Franzen, B.; Linder, S.; Bergman, T.; Auer, G.; Appella, E.; Wirth, P. J.; Jornvall, H. *Electrophoresis* **2000**, *21* (3), 679–686.
18. Seow, T. K.; Ong, S. E.; Liang, R. C. M. Y.; Ren, E. C.; Chan, L.; Ou, K.; Chung, M. C. M. *Electrophoresis* **2000**, *21* (9), 1787–1813.
19. Chong, B. E.; Hamler, R. L.; Lubman, D. M.; Ethier, S. P.; Rosenspire, A. J.; Miller, F. R. *Anal. Chem.* **2001**, *73* (6), 1219–12227.
20. Wulfkuhle, J. D.; McLean, K. C.; Paweletz, C. P.; Sgroi, D. C.; Trock, B. J.; Steeg, P. S.; Petricoin, E. F. *Proteomics* **2001**, *1* (10), 1205–1215.
21. Minamoto, T.; Buschmann, T.; Habelhah, H.; Matusevich, E.; Tahara, H.; Boerresen-Dale, A. L.; Harris, C.; Sidransky, D.; Ronai, Z. *Oncogene* **2001**, *20* (26), 3341–3347.
22. Unger, H. A.; Kane, R. D.; Fox, K. M.; Gandhi, S.; Alzola, C.; Lamerato, L.; Newling, D.; Kumar, S. *Urol. Oncol.: Semin. Orig. Invest.* **2005**, *23*, 238–245.
23. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207.
24. Griffin, T. J.; Goodlett, D. R.; Aebersold, R. *Curr. Opin. Biotechnol.* **2001**, *12*, 607–612.
25. Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
26. Yates, J. R. *Trends Genet.* **2000**, *16* (1), 5–8.
27. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3* (6), 327–332.
28. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Protein Sci.* **1994**, *3*, 1347–1350.
29. Henzel, W. J.; Watanabe, C.; Stults, J. T. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 931–942.
30. Akrigg, D.; Bleasby, A. J.; Dix, N. I. M.; Findlay, J. B. C.; North, A. C. T.; Parry-Smith, D.; Wootton, J. C.; Blundell, T. L.; Gardner, S. P.; Hayes, F.; Islam, S.; Sternberg, M. J. E.; Thornton, J. M.; Tickle, I. J. *Nature* **1988**, *335*, 745–746.
31. Bleasby, A. J.; Wootton, J. C. *Protein Eng.* **1990**, *3* (3), 153–159.
32. Fenyo, D.; Beavis, R. C. *Anal. Chem.* **2003**, *75*, 768–774.
33. Gras, R.; Muller, M.; Gasteiger, E.; Gay, S.; Binz, P. A.; Bienvenut, W.; Hoogland, C.; Sanchez, J. C.; Bairoch, A.; Hochstrasser, D. F.; Appel, R. D. *Electrophoresis* **1999**, *20* (18), 3535–3550.
34. Fenyo, D. *Curr. Opin. Biotechnol.* **2000**, *11*, 391–395.
35. Zhang, K.; Tang, H.; Huang, L.; Blankenship, J. W.; Jones, P. R.; Xiang, F.; Yau, P. M.; Burlingame, A. L. *Anal. Biochem.* **2002**, *306* (2), 259–269.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

36. Feng, Q.; Wang, H.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y. *Curr. Biol.* **2002**, *12* (12), 1052–1058.
37. Ng, H. H.; Feng, Q.; Wang, H.; Erdjument-Bromage, H.; Tempst, P.; Zhang, Y.; Struhl, K. *Genes Dev.* **2002**, *16* (12), 1518–1527.
38. van Leeuwen, F.; Gafken, P. R.; Gottschling, D. E. *Cell* **2002**, *109* (6), 745–756.
39. Link, A. J. Multidimensional peptide separations in proteomics. *Trends Biotechnol.* **2002**, *20* (12), S8–S13.
40. Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871–2882.
41. Zhang, L.; Eugeni, E. E.; Parthun, M. R.; Freitas, M. A. *Chromosoma* **2003**, *112*, 77–86.
42. Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513–523.
43. Ferguson, P. L.; Smith, R. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 399–424.
44. Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. *Chem. Soc. Rev.* **2012**, *41*, 3912–3928.
45. Hunt, D. F.; Bone, W. M.; Shabanowitz, J.; Rhodes, J.; Ballard, J. M. *Anal. Chem.* **1981**, *53*, 1704–1706.
46. Hunt, D. F.; Buko, A. M.; Ballard, J. M.; Shabanowitz, J.; Giordani, A. B. *Biomed. Mass Spectrom.* **1981**, *53*, 397–408.
47. Hunt, D. F.; Shabanowitz, J.; Yates, J. R.; Meiver, R. T.; Hunter, R. L.; Syka, J. E. P.; Amy, J. *Anal. Chem.* **1985**, *57*, 2728–2733.
48. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *57*, 6233–6237.
49. Biemann, K. *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 99–111.
50. Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601–601.
51. Jankowski, K.; Pare, J. R. J.; Belanger, J. *Biomed. Mass Spectrom.* **1985**, *12* (10), 631–631.
52. Roepstorff, P. *Biomed. Mass Spectrom.* **1985**, *12* (10), 631–631.
53. Johnson, R. S.; Matrin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal. Chem.* **1987**, *59* (21), 2621–2625.
54. Biemann, K. *Annu. Rev. Biochem.* **1992**, *61*, 977–1010.
55. Purvine, S.; T., E. J.; Yi, E. C.; Goodlett, D. R. *Proteomics* **2003**, *3* (6), 847–850.
56. Bakhtiar, R.; Guan, Z. Q. *Biotechnol. Lett.* **2006**, *28* (14), 1047–1059.
57. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Natl. Acad. Sci. USA* **2004**, *101*, 9528–9533.
58. McAlister, G. C.; Phanstiel, D.; GOOd, D. M.; Berggren, W. T.; Coon, J. J. *Anal. Chem.* **2007**, *79* (10), 3525–3534.
59. Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 477–485.
60. Nikolaev, E. N.; Somogyi, A.; Smith, D. L.; Gu, C. G.; Wysocki, V. H.; Martin, C. D.; Samuelson, G. L. *Int. J. Mass Spectrom.* **2001**, *212*, 535–551.
61. Laskin, J.; Beck, K. M.; Hache, J. J.; Futrell, J. H. *Anal. Chem.* **2004**, *76*, 351–356.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

62. Dongre, A. R.; Somogyi, A.; Wyscoki, V. H. *J. Mass Spectrom.* **1996**, *31*, 339–350.
63. Sadygov, R. G.; Cociorva, D. C.; Yates, J. R. *Nat. Methods* **2004**, *1* (3), 195–202.
64. Sleno, L.; Volmer, D. A. J. *Mass Spectrom.* **2004**, *39*, 1091–1112.
65. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.
66. Paizs, B.; Suhai, S. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.
67. Schütz, F.; Kapp, E. A.; Simpson, R. J.; Speed, T. P. *Biochem. Soc.Trans.* **2003**, *31*, 1479–1483.
68. Zhang, Z. *Anal. Chem.* **2004**, *76*, 3908–3922.
69. Steen, H.; Mann, M. *Mol. Cell. Proteomics* **2004**, *5*, 699–711.
70. Nesvizhskii, A. I.; Aebersold, R. *Drug Discovery Today* **2004**, *9* (4), 173–181.
71. Taylor, J. A.; Johnson, R. S. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1067–1075.
72. Dančík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J. Comput. Biol.* **1999**, *6*, 327–342.
73. Fernandez-de-Cossio, J.; Gonzalez, J.; Satomi, Y.; Shima, T.; Okumura, N.; Besada, V.; Betancourt, L.; Padron, G.; Shimonishi, Y.; Takao, T. *Electrophoresis* **2000**, *21*, 1694–1699.
74. Münchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047–4057.
75. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (19), 10313–10317.
76. Standing, K. G. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 595–601.
77. Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R. *Anal. Chem.* **2004**, *76*, 2220–2230.
78. Searle, B. C.; Dasari, S.; Wilmarth, P. A.; Turner, M.; Reddy, A. P.; David, L. L.; Nagalla, S. R. *J. Proteome Res.* **2005**, *4*, 546–554.
79. Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. *Anal. Chem.* **2005**, *77*, 7265–7273.
80. Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
81. Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390–4399.
82. Bafna, V.; Edwards, N. *Bioinformatics* **2001**, *17*, S13–S21.
83. Hansen, B. T.; Jones, J. A.; Mason, D. E.; Liebler, D. C. *Anal. Chem.* **2001**, *73*, 1676–1683.
84. Zhang, N.; Aebersold, R.; Schwikowshi, B. *Proteomics* **2002**, *2*, 1406–1412.
85. Field, H. I.; Fenyö, D.; Beavis, R. C. *Proteomics* **2002**, *2*, 36–47.
86. Havilio, M.; Haddad, Y.; Smilansky, Z. *Anal. Chem.* **2003**, *75*, 435–444.
87. Sadygov, R. G.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 3792–3798.
88. Tabb, D. L.; Saraf, A.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 6415–6421.
89. Hernandez, P.; Gras, R.; Frey, J.; Appel, R. D. *Proteomics* **2003**, *3*, 870–878.
90. Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T.; Magnin, J. *Proteomics* **2003**, *3*, 1454–1463.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

91. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J. Proteome Res.* **2004**, *3*, 958–964.
92. Sadygov, R.; Wohlschlegel, J.; Park, S. K.; Xu, T.; Yates, J. R. *Anal. Chem.* **2006**, *78*, 89–95.
93. MacCoss, M. J.; Wu, C. C.; Yates, J. R. *Anal. Chem.* **2002**, *74*, 5593–5599.
94. Fenyo, D.; Qin, J.; Chait, B. T. *Electrophoresis* **1998**, *19*, 998–1005.
95. Craig, R.; Cortens, J. P.; Beavis, R. C. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.
96. Yates, J. R. *Electrophoresis* **1998**, *19* (6), 893–900.
97. Kapp, E. A.; Schütz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J. *Proteomics* **2005**, *5* (13), 3475–3490.
98. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 4646–4658.
99. States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. *Nat. Biotechnol.* **2006**, *24* (3), 333–338.
100. Mann, M.; Wilm, M. *Trends Biochem. Sci.* **1995**, *20*, 219–22.
101. Keller, A.; Nesvizhskii, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383–5392.
102. Cox, J; Neuhauser, N; Michalski, A; Scheltema, R. A; Olsen, J. V; Mann, M *J. Proteome Res* **2011**, *10* (4), 1794–1805.
103. Zhang, J; Xin, L; Shan, B; Chen, W; Xie, M; Yuen, D; Zhang, W; Zhang, Z; Lajoie, G. A; Ma, B. *Mol. Cell. Proteomics* **2012**, *11* (4), M111-010587.
104. Tabb, D. L; Fernando, C. G; Chambers, M. C. *J. Proteome Res* **2007**, *6* (2), 654–661.
105. Sadygov, R. G.; Liu, H.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 1664–1671.
106. Xu, H.; Freitas, M. A. *BMC Bioinf.* **2007**, *8*, 133.
107. Xu, H.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (7), 2605–2615.
108. Xu, H.; Freitas, M. A. *Proteomics* **2009**, *9* (6), 1548–1555.
109. Xu, H.; Hsu, P. H.; Zhang, L.; Tsai, M. D.; Freitas, M. A. *J. Proteome Res.* **2010**, *9* (7), 3384–3393.
110. Xu, H.; Zhang, L.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (1), 138–144.
111. Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4* (3), 207–214.
112. *Guidance for Industry: Q8, Q9, and Q10 Questions and Answers (R4)*; U.S. Food and Drug Adminstration: Washington, DC, 2011, Revision 1.
113. Tanner, S.; Shu, H; Frank, A.; Wang, L. -C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. *Anal. Chem.* **2005**, *77*, 4626–4639.
114. Kim, S.; Na, S.; Sim, J. W.; Park, H.; Jeong, J.; Kim, H.; Seo, Y.; Seo, J; Lee, K.-J.; Paek, E. *Nucleic Acids Res.* **2006**, *34*, 258–263.
115. Bern, M.; Cai, Y.; Goldberg, D. *Anal. Chem.* **2007**, *79*, 1393–1400.
116. *Matrix Science*. http://www.matrixscience.com (accessed June 2015).
117. Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B. *J. Proteome Res.* **2011**, *10* (7), 2930–2936.
118. Fodor, S.; Zhang, Z. *Anal. Biochem.* **2006**, *356* (2), 282–290.
119. Dick, L. W., Jr; Mahon, D.; Qiu, D.; Cheng, K. C. *J. Chromatogr. B* **2009**, *877* (3), 230–236.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

120. Morand, K.; Talbo, G.; Mann, M. *Rapid Commun. Mass Spectrom.* **1993**, *7* (8), 738–743.
121. Zhang, Z. *Anal. Chem.* **2009**, *81* (20), 8354–8364.
122. Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. *Mol. Cell Proteomics* **2004**, *3*, 908–919.
123. Petritis, K.; Kangas, L. J.; Yan, B.; Monroe, M. E.; Strittmatter, E. F.; Qian, W. J.; Adkins, J. N.; Moore, R. J.; Xu, Y.; Lipton, M. S.; Camp, D. G., 2nd; Smith, R. D. *Anal. Chem.* **2006**, *78*, 5026–5039.
124. Gao, Y.; Wang, Y. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (11), 1973–1976.
125. Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (4), 764–772.
126. Zeck, A.; Regula, J. T.; Larraillet, V.; Mautz, B.; Popp, O.; Göpfert, U.; Papadimitriou, A. *PLoS one* **2012**, *7* (7), e40328.
127. Creasy, D. M.; Cottrell, J. S. *Proteomics* **2002**, *2*, 1426–1434.
128. Yu, X. C.; Borisov, O. V.; Alvarez, M.; Michels, D. A.; Wang, Y. J.; Ling, V. *Anal. Chem.* **2009**, *81*, 9282–9290.
129. Kil, Y. J.; Becker, C.; Sandoval, W.; Goldberg, D.; Bern, M. *Anal. Chem.* **2011**, *83* (13), 5259–5267.
130. Ahn, J.; Gillece-Castro, B. L.; Berger, S. *BiopharmaLynx: A New Bioinformatics Tool for Automated LC/MS Peptide Mapping Assignment*; Application Note;Waters Corp.: Milford, MA, 2008.
131. Hariharan, R.; Babu, S.; Gudihal, R. Identification of Oxidation Sites on a Monoclonal Antibody Using an Agilent 1260 Infinity HPLC-Chip/MS System Coupled to an Accurate-Mass 6520 Q-TOF LC/MS; Application Note; Agilent Technologies, Inc: Santa Clara, CA, 2011.
132. Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4* (3), 207–214.
133. Revised publication guidelines for documenting the identification and quantification of peptides, proteins, and post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics*. http://mcponline.org/site/misc/PhiladelphiaGuidelinesFINALDRAFT.pdf (accessed June 2015).
134. Chalkley, R. J. *J. Proteome Res.* **2013**, *12* (2), 1062–1064.
135. Hogeweg, P. *PLoS Comput. Biol.* **2011**, *7*, e1002021.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch007

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Chapter 8

Adventitious Agent Testing of Biologicals: Changing to a New Frontier of Technology, Cell-Based to Nucleic Acid-Based Detection

Ivar Kljavin, Kevin McCarthy, and Dieter Schmalzing*

Global Biologics QC, Genentech, Member of the Roche Group, One DNA Way, South San Francisco, California 94080, United States
***E-mail: schmalzing.dieter@gene.com**

The production of biotechnology products using mammalian cell lines offers an inherent risk of viral contamination due to the scale of the process and the complexity of the materials employed. Thus, detection of adventitious agents in biopharmaceutical manufacturing is both a regulatory requirement and essential to protect both facility and product supply chain to the patient. Within the repertoire of detection methods for adventitious agents are those methods that are based on cellular technology, where the test article, which is derived from appropriate stages of the manufacturing process, is incubated on indicator cells. These *in vitro* tests provide an output where an operator examines the cells for microscopic evidence of an infectious agent. The assays, which require weeks to execute, are indeed effective but do have limitations, including the potential for providing variable final reportable results due to sample matrix factors. Newer, nucleic acid-based methods such as polymerase chain reaction (PCR) can be significantly more rapid, specific to particular infectious agent targets, and more sensitive. However, such method technologies demonstrate the presence of nucleic acid and not direct evidence of an active infectious agent. In this chapter, we discuss the uses and limitations of the *in vitro* virus methods, as well as provide an example of the value of PCR and considerations that guide the design of such methods for an appropriate intended use.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Introduction

Biopharmaceutical production is typically a large-scale process that is considered very complex and prone to contamination. Many different culture media components are used in large amounts to support the growth of host cells that have been engineered to produce a desired biological product. Because of the use of such a complex cellular substrate, infection of the cells and or cell culture media from a number of adventitious agents should be considered a threat to the biopharmaceutical production process. An adventitious agent may be considered any microorganism—including virus, mycoplasma, spiroplasma, mycobacteria, rickettsia, bacteria, fungi, protozoa, parasites, or transmissible spongiform encephalopathy (TSE)—that may have inadvertently been introduced into the production process. As an example, viral contamination can theoretically be spread from a single infected cell or by a single infectious entity through a process that involves expansion of the engineered cell lines into master and working cell banks. These contaminated cell banks can then be expanded into a cellular population for large-scale bioreactor production at the end of which the cell culture fluid is harvested for further processing and purification activities that lead to the final drug product.

Because of the basic design of the manufacturing process, which may be several weeks in duration, one must consider that there is an inherent risk for realizing an infectious agent throughout any step along the manufacturing process, from the raw materials and cells used in production to the purification activities for the final drug product. Thus, a biologics manufacturing viral contamination may have serious impact to an entire facility, including the bioreactor and downstream purification processes, which all require decontamination and verification of effective eradication of the agent.

The biopharmaceutical industry follows regulatory documents that in principle follow a three-strategy approach involving (i) thorough testing of the manufacturing cell substrate and of the raw materials for viral and other adventitious agents; (ii) assessment of the capacity of downstream processing to clear infectious viruses; and (iii) testing the product at appropriate steps for contaminating adventitious viruses or other agents (*1*–*7*). With regard to the use of adventitious agent testing methods that are employed for either raw materials, the cell substrate, or samples derived from the cell culture process itself, regulations that govern the testing may be derived from a broad range of worldwide sources. These include regulations from the U.S. Food and Drug Administration (FDA), the U.S. Pharmacopeial Convention (USP), the European Pharmacopoeia (EP), and the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guidelines. These regulations often do not specify the exact method description such as what would be found in an operational procedure, but instead may provide higher level parameters that are based on principles of virology. For example, in the FDA and ICH guidance for viral safety, testing execution requires that the test article be incubated for 14 to 28 days on a human and a primate cell line, as well as being tested on a cell line of the same species and tissue type that is used for production of the biological product (*1*, *5*). Such parameters provide a broad range of virus

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

detection potential with such a combination of indicator cells that themselves each show a different infectivity potential. On the other hand, for mycoplasma testing, more detail as to sample inoculation volume, incubation conditions, and duration, along with detection medium formulations, are described (*8–13*). However, not all regulatory documents are completely aligned with regard to method execution differences or level of detail, as is the case between the EP, the USP, the FDA Points to Consider document, and the Japanese Pharmacopeia (*8*, *11–13*).

Nevertheless, application of a three-tiered strategy described above has indeed successfully mitigated a negative clinical consequence directly to patients. However, it should be noted that even with extensive testing in place at the appropriate steps in the process most likely to detect virus or other adventitious agents, contaminations of biopharmaceutical manufacturing facilities do occur. This is primarily due to the fact that test results are too late or have been relied on too heavily without other more holistic approaches of preventing introduction of adventitious agents (e.g., employing barriers to prevent entry in the manufacturing process like removal of high-risk raw materials from the process, heat inactivation or gamma irradiation of cell culture media components, proper facility design and controls, and personnel flows that mitigate the risk of adventitious agent introduction). Moreover, we must keep in mind that using any of the adventitious agent detection methods available should not be considered a means to demonstrate absence of the target infectious agent. Absolute freedom of virus, for example, is not possible to confirm based on a final reportable result of "non-detected" because the sensitivity and specificity of the method cannot assure that all possible viral agents will be found.

Although mycoplasma and other organisms such as *Leptospira* have been detected in a biopharmaceutical processes and are considered threats to the manufacture of biopharmaceutical products, this chapter will focus on viral agents and their detection for purposes of illustrating the evolution of a testing technology based on microscopic observation of cells for evidence of infection to more advanced, nucleic acid-based detection platforms. *In vitro* adventitious virus detection, which utilizes indicator cells and is a good example of a currently used effective testing strategy for screening samples from the bioreactor, is evolving to nucleic acid-based detection platforms. Although this evolution is in part due to limitations in cell-based viral detection itself, which will be described below, there also are practical advantages to moving away from cell-based detection. These include increased sensitivity, shorter time to derive a final reportable result, and overcoming potential interference of indicator cell infection from the sample matrix itself.

One note worth mentioning here is that the *in vitro* adventitious virus detection methods will and must be used even after the new technologies have been implemented. Nucleic acid-based test results do not represent direct evidence of an infectious agent, but rather such methods provide evidence of nuclear material of those agents. The nucleic acid-based methods would then indeed be augmented by the *in vitro*-based methods to derive evidence of an infectious agent, and thus together be a part of the overall testing strategy.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Available Testing Strategies

Figure 1 summarizes examples of different types of viral detection assays that are generally used in the control system for the manufacturing of biologics. These include assays that are cell-based *in vitro* and *in vivo* methods, electron microscopy, and polymerase chain reaction (PCR). Such methods serve the purpose of demonstrating the presence or absence of virus in raw materials, manufacturing cell banks, and unprocessed bulk for release (*5*).

Figure 2 illustrates the various points at which these testing activities take place within appropriate stages of a typical manufacturing process. For example, the testing of the large-scale bioreactor material for infectious agents at harvest is not only a regulatory requirement, but it also is considered the best opportunity for detection because an infectious particle would have propagated to its maximum titer. Additionally, employing viral detection methods to "high-risk" raw materials affords the opportunity to assess the risk of those raw materials prior to being used in the manufacturing process. For production of biologicals, ICH Topic Q5A (1999) suggests extensive screening of a master cell bank for both endogenous and non-endogenous viral contaminants (*5*). The appropriate control strategy involves implementing the correct testing strategy at the appropriate time points in a production process. Cell-based *in vitro* and PCR-based methods will be discussed below to highlight their utility in a representative process.

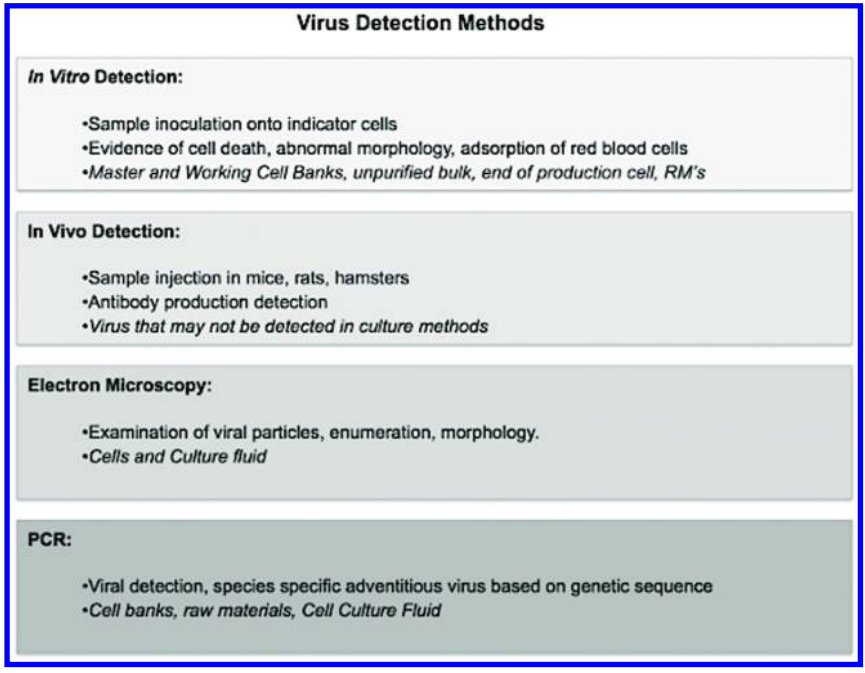

Figure 1. Viral detection methods that are employed within the biopharmaceutical manufacturing control system. RM: raw materials.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

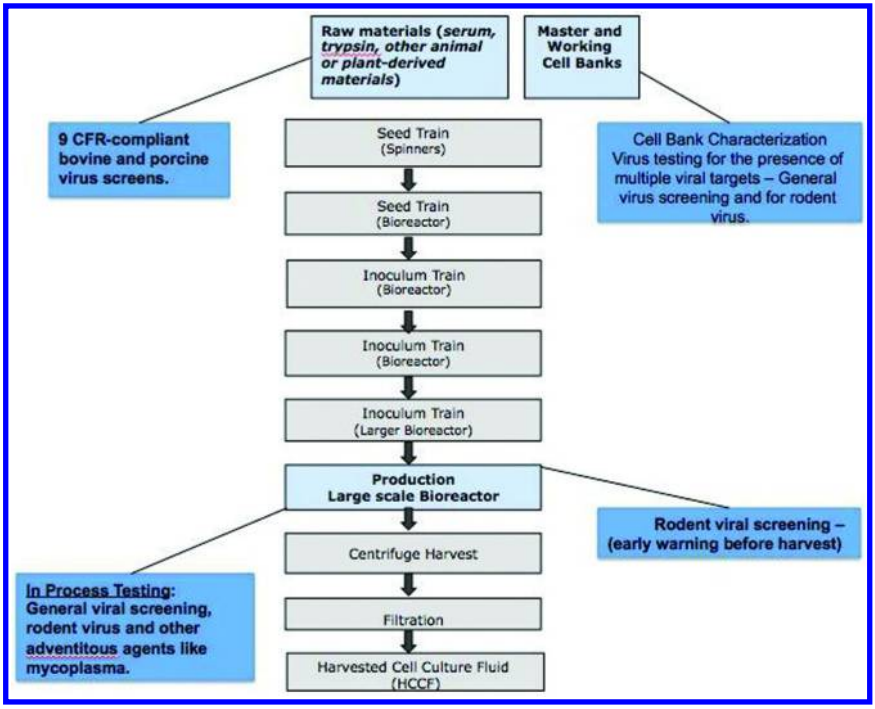

Figure 2. Illustration of the various points at which virus detection methods take place within appropriate stages of the biopharmaceutical manufacturing process.

In Vitro Virus Detection

The *in vitro* viral screening method is based on the principle that a virus will replicate in a suitable host cell, demonstrating its presence through morphological changes observed in the infected cells and/or by hemadsorption of erythrocytes onto the indicator cell surfaces or hemagglutination. Suitable indicator cells are used based on their ability to become infected by target viruses. For the general viral screening assay, three different indicator cells are generally employed: a human diploid cell line (MRC-5), a primate cell (Vero), and a cell line that represents the same species as the cell substrate used in the manufacturing process (*1*). This design affords the opportunity to detect a wide spectrum of viruses infectious to humans and to the manufacturing cell substrate. For other specific viruses such as minute virus of mice (MMV), a cell line demonstrated to be sensitive to this virus (324K cells) is employed. For the 9 Code of Federal Regulations testing of raw materials, bovine or porcine viruses are screened using species-appropriate indicator cells (*10*).

The actual execution of the *in vitro* screening method involves the inoculation of the test article (i.e., unprocessed bulk fluid from the bioreactor) onto the indicator cells for a defined period of time. The cells are subsequently incubated for 2 to 4 weeks or more and observed under the microscope for evidence of

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

infection throughout the duration of the assay as described above. Evidence of infection may be total lysis of the cells, syncytia formation, or other morphological changes not seen within the non-sample inoculated controls. Such morphological signs are termed cytopathic effect (CPE).

Detection of a virus is typically based on the utilization of *in vitro*-based methods that require 2 or more weeks to derive a final reportable result. Because the test article would be derived from the bioreactor when the cell culture fluid is harvested, a positive result would be reportable after much of the purification process has been completed. Furthermore, although considered the "gold-standard," *in vitro* viral detection methods have innate limitations to their design, which is based on basic virology and cell culture principles. Because such assays are widely in use and are accepted by health authorities worldwide, their limitations will be described in more detail below.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Limitations of in Vitro Viral Detection

Although testing for virus is a requirement and the basic design of the methods are outlined in several regulatory documents, the limitations of those methods should be understood. Because there have been reported false positives, false negatives, and conflicting final reportable results demonstrated between different testing laboratories on the same sample aliquot (*14*), *in vitro* viral screening assays have, over the past several years, been the focus of attention for several manufacturing firms. The detection of some viruses may not be realized or may be inconsistent because of two contributing limiting factors: (1) the nature of viral infection itself; and (2) the nature of the methods' dependency on the cells to become infected. Therefore, one should consider the following when employing cell-based viral detection methods:

- Not all viruses induce microscopically noticeable morphological changes in the indicator cells (CPE) and/or hemadsorption. This is because not all virus replicate in test method systems, or the virus may replicate without any visible effect on the indicator cells.
- The panel of indicator cell lines used in the methods will show different susceptibilities to infection for a particular virus due to the mechanism of infection.
- Infectivity potential of the indicator cells may be affected by how the indicator cells are handled and cultured and/or impacted by the test article itself.

All three of the above considerations may contribute to either a false positive or false negative that, as described above, have been reported to have significant impact to a pharmaceutical manufacturer and ultimately the patents. The bullet points below indicate those detailed factors that can induce such false responses of the indicator cells.

False negatives:

- Inactivation of the virus itself by the sample matrix or the product itself.
- Decrease infectivity potential of the indicator cells:
    - Effects from the sample matrix changing the ability of the cells to become infected.
    - Specific impact from the biological activity of the product changing the ability of the cells to become infected.
    - Variations in cell culture maintenance. How the indicator cells are handled or cultured may impact the infectivity potential of the indicator cells.

False positives:

- Sample matrix may induce morphological changes of the indicator cells that may be interpreted as caused by a virus.
- Product biological activity may induce observations of cell death or other morphological changes that mimic a virus-induced morphological change.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Contamination Experience That Led to the Realization of the Value of PCR-Based Viral Detection

Genentech's experience with two large-scale manufacturing contaminations in 1993 and 1994 by a rodent virus, MMV, provided an opportunity to demonstrate the value of PCR-based detection because the two events had significantly different outcomes (*15*). Comparing the two events that occurred between many successful bioreactor runs led to the development and implementation of a viral barrier approach involving the supplementation of cell-based MMV detection methods with PCR-based methods.

The initial 1993 event was considered to be a "late" detection, or rather too late for preventing the spread of contamination to downstream processes. Virus was first indicated in an *in vitro* viral screening assay several days post-sample inoculation and followed by further testing to verify MMV as the contaminant. The time required for completing these testing activities allowed the spread of the contamination to multiple bioreactors and downstream production activities. The overall impact of this event was a requirement for extensive decontamination of the facility along with significant loss of product. The widespread contamination due to delayed assay readout time clearly demonstrates the need for an orthogonal assay for improved real-time process control for which PCR-based assays are well suited.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Considerations for Design of PCR Methods for Adventitious Agent Detection

PCR methods are useful in overcoming the limitations associated with media-based methods. The first PCR is described in a paper by Kleppe et al. in 1971 (*16*) in which they describe replicating short sequences of DNA using a DNA polymerase; however, the DNA polymerase they used was heat-labile and had to be re-introduced with each new amplification cycle, making the process essentially impractical for routine use. Modern, efficient PCR was developed as a result of the discovery of a thermally stable DNA polymerase from the bacterium *Thermus aquaticus* found in the sulfur-rich hydrothermal vents of Yellowstone National Park in 1983 (*17*). This wild-type polymerase is commonly referred to as *Taq* polymerase, and it is the progenitor of many commercially engineered *Taq* polymerases available today. The EP chapter 2.6.21 (*18*), Nucleic Acid Amplification Techniques (NAT), contains guidance and recommendations for PCR assay development and considerations for validation of nucleic acid amplification techniques.

PCR is an *in vitro* assay that amplifies a limited quantity of DNA template that it not readily detectable into several billion "copies" of DNA template that can be easily visualized on an agarose or polyacrylamide gel electrophoresis (PAGE) and/or used for subsequent molecular manipulations such as cloning. With today's technology, detection of the amplified DNA is most often accomplished using instrumentation such as Real-Time PCR instruments. A typical PCR cycle involves denaturing the template DNA into single strands at ~94 °C. The reaction is then cooled to allow for a synthesized forward and reverse primer set (of known sequence) to anneal to complementary sequences on the template; this temperature is sequence-specific and often requires optimization. A lower annealing temperature yields less fidelity of the primer-template annealing and therefore less specificity of the reaction. The final step is the polymerization step, which entails increasing the reaction temperature to the optimal temperature for *Taq* polymerase processivity; for wild-type *Taq* polymerase it is ~72 °C. At this temperature, the *Taq* polymerase polymerizes a complementary sequence using the primer-template hybrid as a substrate. These 3 steps are typically repeated for 25 to 45 cycles to achieve the level of sensitivity desired based on logarithmic amplification of the template DNA. The basic PCR components include: template DNA, sequence-specific primers, $MgCl_2$, deoxynucleotide triphosphates (dNTPs), water, and *Taq* polymerase. PCR assays usually are developed to be highly specific to a target adventitious agent. This specificity can be optimized to allow for a wider detection range of similar species by using degenerate primers (different primer sequences in the same reaction that take into account sequence variability between species), adjusting the concentration of PCR components such as $MgCl_2$, or a combination of these.

There have been many iterations of PCR since its inception in 1983, some of these include (but are not limited to):

- Touch-Down PCR is used to increase PCR specificity by starting with an initial annealing temperature above the projected melting temperature (T_m) of the primers being used and then reducing the annealing

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

temperature by one or two degrees with each cycle. Any difference in T_m between correct and incorrect annealing will produce an exponential advantage of twofold per cycle. This allows the template-primer hybridizations with the highest fidelity to outcompete those that have less fidelity. This assay usually relies on gel electrophoresis for detection of the product.

- Real-Time PCR using a probe in which the PCR reaction includes a sequence-specific probe labeled with a reporter dye at the 5′ end and a quencher molecule at the 3′ end. Real-Time PCR uses an engineered *Taq* polymerase such as TaqMan®; these polymerases possess a 5′ to 3′ exonuclease activity that cleaves (one base at a time) any DNA (in this case the probe) it encounters during polymerization. When the reporter molecule is cleaved off it becomes free of the steric quenching effect of the quencher molecule and can now fluoresce and be detected when excited at the appropriate wavelength. The sequence-specific probe also adds another level of specify to the assay. This detection involves instruments that excite (laser) and detect (charge-coupled device [CCD] camera) available fluorescent signal during each amplification cycle.
- Multi-Plex Real-Time PCR uses two different primer probe sets to detect different target sequences. The reporter molecules on the 5′ end of both probes will be excited and detected at different wavelengths to avoid interference with each other. This method usually is used to detect internal control (IC) spikes.
- Real-Time PCR using a Hoechst dye (such as SyberGreen®). The Hoechst dye is a double-stranded, DNA-binding molecule that fluoresces when attached to double-stranded DNA (e.g., at the end of an elongation cycle) and is nonfluorescent when it dissociates when the DNA becomes single stranded again (e.g., at the denaturation step). This detection involves instruments that excite (laser) and detect (CCD camera) available fluorescent signal after each amplification cycle.
- Reverse-Transcriptase PCR uses RNA as a starting template. The RNA is first transcribed into DNA using reverse transcriptase and then subjected to other PCR techniques as a complementary DNA (cDNA) template.
- Quantitative PCR (q-PCR) is a type of Real-Time PCR that includes a DNA standard curve that is a dilution series of known quantities of template DNA. The amount of fluorescence from an unknown sample can be compared to the standard curve fluorescence to derive a starting quantity of template DNA. Bustin et al. (*19*) provides a thorough background and an in-depth discussion about specific considerations when designing a quantitative PCR assay.
- Uni-Directional PCR uses one primer to amplify one strand of template DNA. This provides a linear amplification of the template and is used in DNA sequencing applications.

Further discussion here will focus on using Real-Time PCR or reverse-transcriptase PCR methods for detecting and identifying adventitious agents such as mycoplasma, *Leptospira*, and viruses. In the biopharmaceutical industry (e.g.,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

processes that use Chinese hamster ovary [CHO] cell lines to synthesize drug products), the host cells (or properties of the host cells) can provide nutritional co-factors that are required for certain adventitious agents to propagate. As such, the adventitious agents tend to associate (attached to or in close proximity to) with the host cells in highly variable numbers. If the adventitious agents are cultivable on solid media, standard plate count methods would not necessarily provide a reliable estimation of the number of adventitious agents present in the media because, for example, if 10 mycoplasma were attached to one CHO cell, that would "plate" out as one colony-forming unit (CFU). Conversely, if 1,000 mycoplasma were attached to the CHO cell, they would also "plate" out as one CFU. Both scenarios yield one CFU but have severely different mycoplasma titers. Both *Leptospira* and mycoplasma are adventitious agents that associate with their host cell in this way so that they may glean required nutrients for growth. Additionally, as mentioned above, the culture method would not detect the significant presence of non-viable adventitious agents.

For quantitation and or limit-of-detection determination for adventitious agents that display varying quantities of association with the host cell, a method should be developed and validated to establish a qualified control standard. This standard would typically be generated by extracting large quantities of DNA (e.g., microgram and nanogram quantities) from a healthy culture of the target organism. The DNA would then be thoroughly quantified using a method such as PicoGreen® or another reliable and valid method (*20*). The qualified and quantified DNA could then be diluted to the picogram and femtogram levels and used as a standard curve to quantitate the sample DNA results or determine the limit of detection during validation. The quantitated DNA of the adventitious agents in the sample then can be used to calculate the approximate number of organisms present in the sample using the published copy number of the target gene, the efficiency of the amplification, the molecular weight of each base, the published genome sequence, and the number of each base in the sequence.

PCR can be leveraged to overcome many issues associated with adventitious agent culturing. The most common PCR methods employed for detection of adventitious agents are Real-Time PCR, Multi-Plex Real-Time PCR, and reverse transcriptase-PCR. Additionally, DNA sequencing may be modified to detect and identify adventitious agents.

The following is an example of a Multi-Plex Real-Time PCR sample detection method outline for DNA-based adventitious agents (Note: As with all PCR methods, this procedure will need to be customized and optimized for specific user requirements.):

- Pre-treat samples if needed (i.e., formulate with EDTA to overcome nuclease activity or dilute to overcome matrix effects).
- Include an IC of known sequence and quantity that is spiked into the sample matrix. Ideally, an IC should emulate the pretreatment, extraction, and PCR profile of the sample but have a different PCR target sequence (e.g.. an engineered plasmid or innocuous non-interfering organism from a separate genus).
- Include a negative extraction control (NEC) spiked with the IC.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

- Extract the DNA or RNA using industry standards. Today's commercial liquid handling technologies (e.g., the MagMAX™ Express from Life Technologies, the MagNA Pure 96 from Roche Applied Science, Qiagens QIAcube technology), afford many automated options for DNA or RNA extraction (following manufactures recommendation). Many of the current extraction methods are guanidine HCl-based cell lysis followed by DNA binding magnetic bead purification.
- For RNA, this is the step when the reverse transcriptase reaction is conducted according to the instructions provided with the enzyme.
- Obtain two different commercially available primer-probe sets or design primers and probe to individual needs using commercial primer synthesizing suppliers. One primer-probe set should be designed to amplify the sample target sequence, and the other set should be designed to amplify the IC target sequence.
- In a class II biosafety cabinet (BSC) of a dedicated clean room, prepare a Multi-Plex Real-Time PCR master mix and aliquot into the appropriate (pre-determined) wells of a 96-well Real-Time PCR plate. Add PCR-grade water to at least 2 wells to act as a PCR-negative control (NC). Cover the plate with optical adhesive cover (sealing the NC wells). The master mix should be prepared to account for all reaction aliquots plus 10 percent additional volume.
- Move the covered plate to the BSC of a dedicated DNA extraction room and add the DNA from the extracted samples and NEC. Seal those wells.
- Move the covered plate to the BSC of a dedicated positive control room and add DNA from a known positive sample target (PCR positive control [PC]), contained in at least 2 wells. Seal the plate and centrifuge to dislodge bubbles in the bottom of the wells.
- Program the appropriate cycle parameters into the Real-Time PCR instrument, place the PCR plate in the instrument, and start the run.
- The data are usually reported as threshold cycle (Ct), crossing point (Cp), or take-off point (TOP) values. These values refer to the same value from the real-time instruments; the nomenclature difference is due to real-time instrument manufacturers assigning these names so there is no conflict with manufacturing competitors. They are defined as the number of cycles required for the fluorescent signal to cross the threshold (i.e., exceed background level) and are inversely proportional to the amount of target nucleic acid in the sample (i.e., the lower the Ct level, the greater the amount of target nucleic acid in the sample).
- System suitability: The PCR PC and all IC values should be positive with known acceptance criteria assigned (based on validation); the NEC should be negative for sample signal and the IC signal positive with pre-determined acceptance criteria; the NC should be negative for both fluorescent wavelengths.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Contamination Following PCR Implementation Reduced Impact on Production

Following the 1993 contamination event discussed above, PCR-based testing was implemented as an orthogonal assay for MMV. The outcome of the 1994 contamination, however, was different and had less of an impact on the facility and product supply because of the use of a MMV PCR test as an early-warning, in-process hold step. With this strategy, the PCR test was executed 48 hours prior to harvest of the cell culture fluid. Thus, because the PCR test usually yields results on the same day of the test execution, the bioreactor activities were stopped to prevent further processing in the event of a positive MMV indication. The 1994 MMV contamination, therefore, demonstrated a significantly different outcome in that the contamination was isolated to a single bioreactor with minimal loss of product. It is important to note that after the 1994 contamination, other risk mitigation activities were developed that are currently used today, including heat inactivation of select high-risk raw materials, virus filtration, and personnel/equipment segregation practices. Therefore, the PCR-based test for MMV, together with the above risk mitigation strategies, now provide a holistic means to mitigate the risk of adventitious agent contaminations.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Considerations for Moving From Cell-Based Detection to the Future State of PCR-Based Detection of Adventitious Agents

As illustrated in the discussion above, the cell-based methodology has limitations. However, one must consider that PCR methods also have elements of limitations, and these limitations require understanding to ensure a good fit for the intended testing purpose.

Below is a summary of the most prevalent considerations for both cell-based and PCR-based methods used for virus screening.

a. In vitro viral detection relies on infectivity of susceptible cell lines.

Advantages: Direct assessment of infectivity, industry and health authority "gold-standard," direct assessment of infectious nature and of viable organisms.

Disadvantages: Lengthy time requirement, agent may not be seen microscopically, sample matrix impact, complex, can display low sensitivity, prone to operator ergonomic issues.

b. PCR detection based on specific amplification of target within infectious agent genome.

Advantages: Enhances ability to detect infectious agents from a variety of sample matrix types without interference, rapid detection, sensitive and more automated.

Disadvantages: No assessment to determine infectivity or viability uses lower sample volume than cell-based methods, requiring assessment of impact to target detection sensitivity; multiple viral detection may be difficult for

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

general screening; only recently becoming more accepted by health authorities worldwide.

An adventitious agent detection method's "fit-for-purpose" is an important consideration. Although PCR is rapid and sensitive, the information first available for a positive result would, in many cases, require further verification and study for viability of the detected organism. This is particularly a problem for biologics, where a positive PCR result only indicates the presence of DNA of the adventitious agent target in the sample, leaving the question of whether the result was due to a "real" contamination of the process or the detection of DNA that was carried into the production process from a raw material. This aspect of PCR is important for deciding the intended purpose of the testing activity (e.g., Is evidence of an adventitious agent nucleic acid in a test article, without evidence of a viable agent, acceptable for release of a biologics?).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Considerations for Application of PCR for Adventitious Agent Detection as a Primary Tool

The detection and subsequent identification of bacteria, molds, and viruses is critical in the biopharmaceutical industry to insure patient safety and product efficacy. To that end, there is a continual endeavor to improve the available detection methods; evolution of new technologies affords these opportunities for method improvements and new method developments. The following section identifies possible considerations when deciding on the replacement of a culture-based method with a PCR-based method for detection of adventitious agents:

- *In vitro* detection methods for bacteria, molds, and viruses that involve culturing the organisms in broth and/or on solid media can in some cases enhance growth of adventitious agents, and, in turn, increase the sensitivity of detection. These media are often supplemented for optimal growth of the adventitious agent (e.g. viruses when cultured in the presence of living host cells). The selection and/or enrichment opportunity available in formulation of the *in vitro* assay culture media is broad and often allows for selective growth of target organisms, which may be advantageous in some cases.
- In some cases, as is seen with *Leptospira licerasiae*, some species do not cultivate well on solid media (Roche Genentech Inc., unpublished data), and although detection using increased optical density of the media over time is useful, one is not able to reliably generate CFU/mL results from that data. Additionally, as seen with *Leptospira licerasiae*, an inoculated media may require 7 to 10 days to optically detect growth; this would not suffice to insure sterility and purity of downstream processes. In such a case, PCR-based methods would supplement *in vitro* techniques by providing more quantitative results and more rapid detection of potential contamination.
- Broth- and/or agar-based culture methods detect viable organisms and do not usually detect non-viable organisms, PCR does not typically discern

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the two. The ability of *in vitro* assays to detect viability is important for understanding whether or not the infection is active and is a threat to the facility and downstream processing that must be mitigated.

- Detection of non-viable contaminations (e.g., by PCR-based methods) can be critical in biopharmaceuticals as well, in that:
    - A previously viable adventitious agent may have, as a by-product of its growth, altered the biopharmaceutical product that has already been produced in the bioreactor (e.g., for protein- and/or antibody-based products, the quaternary structures or other post-translational modifications such as glycosylation or phosphorylation events might be changed).
    - Parts of these no longer viable adventitious agents that are not easily detected with other methods (e.g., exotoxins, endotoxins) may carry over into downstream purification and negatively affect a patient health. For example, some *Leptospira licerasiae* species cell walls do not reliably react when used as a spike in a Limulus amoebocyte lysate (LAL) assay, the historically accepted assay used to detect endotoxin (Roche Genentech Inc., unpublished data); this is undoubtedly due to the variable amount of lipopolysaccharide found in *Leptospira licerasiae* cell walls between species.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Future Perspectives

The ability to detect and identify viable and nonviable adventitious agents in a biopharmaceutical bioreactor is critical for patient safety and product efficacy. Biopharmaceutical companies should make every effort to ensure that they are leveraging the most current and reliable technologies available by continually reviewing the latest methods and information available to the industry, developing and validating new (improved) methods for detection, auditing and trending currently employed methods, and optimizing current methods as appropriate. PCR affords a wide array of opportunities for the design of efficient, effective, and sensitive detection assays.

Suggestions To Improve Sensitivity

A lower starting template DNA concentration (i.e., 1 copy per 20 µL) makes it less likely that the target molecule will be successfully pipetted into the PCR reaction tube. To alleviate this effect and improve sensitivity, when larger volumes of sample (e.g., 10 mL) are available for testing, it should be determined if centrifuging the larger volume and re-suspending the resulting pellet(s) into a more manageable volume (e.g., 1 mL total volume) is feasible. Also, treatment of the sample with EDTA to reduce metal co-factor mediated nuclease activity can be considered to help ensure that the target DNA remains intact throughout the DNA extraction.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Whether developing a new PCR assay, modifying an existing PCR assay due to vendor reagent changes, or incorporating new technologies into an existing PCR assay, one of the early steps in these processes should be assay optimization. Optimization of a PCR assay is important for improving sensitivity and specificity of the assay. Some of the PCR parameters that should be considered for optimization include $MgCl_2$ concentration, primer probe concentrations and sequences, dNTP concentrations, thermal cycling temperatures, and cycle ramp rates. Additional sensitivity improvements include basic procedures such as increasing the master mix component stock concentrations such that less volume is used per reaction to allow for larger volumes of template DNA to be used, as well as using more than one PCR reaction per sample if the volume of extracted DNA is sufficient so that most of the extracted DNA is used.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

One Assay, Many Uses

Although specificity of a detection assay is critical, designing an assay that can detect and identify many different types of contaminants while maintaining sensitivity would be a significant benefit to patients, product and facility safety, and enhancement of business processes. For example, Microbial ID Sequencing using 16s rDNA sequencing is an assay that is typically used for identification of bacteria based on their 16s rDNA gene sequences. The 16s ribosomal RNAs are the components of prokaryotic ribosomes. These RNAs contain very conserved primer sequence binding sites as well as hypervariable regions that provide species-specific differentiation. This makes the 16s rDNA an optimal target for microbial identification for many species simultaneously.

The following procedure can be used to modify the basic Microbial ID assay using 16s rDNA sequencing so that it can function as a detection assay as well.

Improved sample preparation can be achieved by concentrating large volume samples (e.g., 10 mL), using centrifugation prior to DNA extraction. After ensuring a homogeneous suspension, half of the sample can be tested in a standard bioburden assay format, and DNA extracted from the second half of the sample can be analyzed by the Microbial ID sequencing assay.

The Microbial ID sequencing assay includes an initial PCR amplification of a sample using degenerate and/or multiple primer sets that increases the availability of the target DNA for the sequencing reaction. Increasing the number of PCR cycles from the typical 30 cycles to 45 cycles offers an opportunity to amplify very low levels of contaminants, thereby improving sensitivity. A portion of the PCR reaction product can be analyzed on a precast agarose gel with a quantitative DNA ladder, and based on the comparison to the DNA quantitation ladder, the remaining sample can be diluted to the appropriate concentration needed for the DNA sequencing reaction. At this point, a positive result (as indicated by an appropriate size of the PCR product band on the agarose gel), although not yet identified, can be used as a cue to stop or pause the manufacturing process that has been determined to be contaminated. The sequence of the PCR product can be determined using a version of the Sanger's dideoxy sequencing method and compared to validated sequences, contained in the well-characterized database to identify contaminant species.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Undetermined sequences and/or nested sequencing chromatogram peaks typically indicate a "mixed" starting culture. In this case, identification of all the contaminating species is still required despite the sample contamination having been confirmed with the PCR part of the assay. If this occurs, the Bioburden culture inoculated from half of the original sample can be used to obtain isolates of each of the different contaminants for use for confirmation with Microbial ID using 16s rDNA sequencing.

PCR-based detection and identification methods are very dynamic tools available to the biopharmaceutical industry. They offer many potential avenues for future assay development, optimization, and implementation that will help ensure product and patient safety. In this section, we highlighted ways to optimize current assays, as well as provide a basic example of an assay redesign such that an ID method also could be used as a detection method. New technologies, discoveries, and reagent improvements continuously offer the opportunities to develop and/or optimize quality control assays. If recognized and capitalized upon, those opportunities can help insure patient safety, shorter lot release times, and a more streamlined business process.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Conclusion

This chapter has provided a brief background into adventitious agents in the biopharmaceutical industry, outlined an example of a type of older adventitious agent detection method that still is considered an industry standard today, and given a brief description of a possible replacement method. Although well-established methods are validated and relied upon to help ensure the safety of a product, quality control organizations should continually assess the applicability of new technologies that may improve the sensitivity, specificity, and/or robustness of their current methods. We have introduced one possible replacement and/or supplemental technology for an industry standard viral detection method based on PCR. Because PCR also may have possible difficulties that must be overcome, as with any new method, it will require optimization and validation prior to implementation. Keep in mind that a new technology assessment may determine that the legacy method is indeed the better of the two. All biopharmaceutical entities should continually strive to improve on all of their current quality control methods by proactively assessing new technology and emerging scientific knowledge with the end goal of improving the sensitivity, specificity, and or robustness of quality control test methods. This helps ensure that the safety of our patients is held in the highest regard.

References

1. *Points to Consider in Characterization of Cell Lines to Produce Biologicals*; Center for Biologics Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, 1993.
2. *Note for Guidance on Virus Validation Studies: The Design, Contribution and Interpretation of Studies Validating the Inactivation and Removal of*

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Viruses; CPMP/BWP/268/95l; European Agency for the Evaluation of Medicinal Products: London, 1996.
3. *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*; Center for Biologics Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, 1997.
4. *Requirements for the Use of Animal Cells as in Vitro Substrates for the Production of Biologicals*; World Health Organization:Geneva, 1998.
5. *Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin*; ICH Q5A; ICH: Geneva, 1999.
6. *Guideline on Virus Safety Evaluation of Biotechnological Investigational Medicinal Products*; EMEA/CHMP/BWP/398498/2005; European Medicines Agency: London, 2008.
7. *Guidance for Industry: Characterization and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications*; Center for Biologics Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, February 2010.
8. *European Pharmacopoeia*, 7th ed.; Section 2.6.7, Mycoplasmas, January 2008, 20607 corrected 6.1.
9. *Japanese Pharmacopoeia XV*; Section 14. Mycoplasma Testing for Cell Substrates used for the Production of Biotechnological/Biological Products, 2006.
10. Food and Drugs: General Biological Product Standards; Test for Mycoplasma. *Code of Federal Regulations*, Part 610, Title 21, Section 610.30, 2014.
11. *Mycoplasma Tests*; General Chapter 63, USP 33-NF 28 Reissue; U.S. Pharmacopoeial Convention, Rockville, MD, 2010; pp 88−91.
12. *Points to Consider in Characterization of Cell Lines to Produce Biologicals*; Center for Biologics Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, 1993; Attachment #2, pp 38–40.
13. *Japanese Pharmacopoeia XV*; Section 14. Mycoplasma Testing for Cell Substrates used for the Production of Biotechnological/Biological Products, Supplement II, September 30, 2009.
14. Hendricks, L. C.; Jordan, J.; Yang, T.-Y.; Driesprong, P.; Haan, G. J.; Viebahn, M.; Mikosch, T.; Van Drunen, H.; Lubiniecki, A. S. Apparent virus contamination in biopharmaceutical product at Centocor. *PDA J. Pharm. Sci. Technol.* **2010**, *64* (5), 471–480.
15. Garnick, R. L. Viral safety and evaluation of viral clearance from biopharmaceutical products. *Dev. Biol. Stand. (Basel)* **1996**, *88*, 49–56.
16. Kleppe, K; Ohtsuka, E; Kleppe, R; Molineux, I; Khorana, HG Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* **1971**, *56*, 341–361.
17. Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Springs Harbor Symp. Quant. Biol.* **1986**, *51*, 263–273.
18. Nucleic Acid Amplification Techniques. In *European Pharmacopoeia*; Chapter 2.6.21, 2014; 07/2014:20621.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

19. Bustin, S.; Benes, V.; Garson, J.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.; Shipley, G.; Vandersompele, J.; Wittwer, C. The MIQE Guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55* (4), 611–622.
20. Ahn, S; Costa, J; Emanuel, J PicoGreen quantitation of DNA: Effective evaluation of samples pre or post-PCR. *Nucleic Acids Res.* **1996**, *24*, 2623–2625.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch008

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Chapter 9

Monoclonal Antibody Analysis Using Microfluidic Technologies

G. O. Staples,* O. G. Potter, and H. Yin

Agilent Technologies, 5301 Stevens Creek Blvd., Santa Clara, California 95051
***E-mail: gregory_staples@agilent.com**

Microfluidic devices are enabling technologies that have delivered tangible benefits to the bioanalytical community in the form of reduced sample consumption, massively parallel assays, cost-savings, portability, and automation and integration of multi-step protocols. Those researchers charged with the characterization of therapeutic proteins, namely monoclonal antibodies, have also benefited from microfluidics. In these experiments, where sample quantity is rarely a limiting factor, microfluidics offers considerable reduction of sample manipulation and thus time to measurement. This chapter highlights some of the contributions of microfluidic analytical tools to monoclonal antibody analysis. In addition, data generated from characterization of the NISTmAb reference material using microfluidic technologies such as lab-on-a-chip and HPLC-chip MS are presented.

Introduction

Recombinant monoclonal antibodies (mAbs) are the fastest growing group of therapeutic drugs in the pharmaceutical market (*1*). The power of mAbs in the treatment of disease stems from their exquisite specificity and high affinity for their targets. Such characteristics can be exploited by generating mAbs against features expressed by cancerous cells but not by normal cells (*2*). A great deal of progress in the production of mAbs has been made since the first attempt at doing so in the early 1980s (*3*). Indeed, a major victory was achieved when the first mAb for cancer treatment was approved by the U.S. FDA in 1997 (*4*). Continual

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

improvements in the development of mAb therapeutics have been made since the appearance of that first product. For example, we now understand and have overcome the limitations of using murine-derived mAbs, which can exhibit toxicity (*5*) and reduced half-life, and have developed methods for producing humanized and totally human versions of molecules with desired specificities.

As of 2012, twelve mAbs have been approved by the U.S. FDA for treatment of solid tumors and blood malignancies (*6*). There are also mAbs approved for the treatment of other diseases, such as multiple sclerosis, macular degeneration, and Crohn's (*7*). There are hundreds of mAbs currently in development around the world, with a trend toward customization for increased functionality. Deeper understanding of mAb biochemistry has led to the use of mAbs to deliver cytotoxic drugs (in the form of antibody drug conjugates, ADCs) to tumor cells via endocytosis (*8*). The developments continue towards greater efficacy, specificity, and lower toxicity in the form of bispecific mAbs (*9*, *10*), antibody fragments (*11*, *12*), and mAbs with engineered glycosylation (*13*). These efforts to create custom molecules, among others, may hold the key to a therapeutic that can act as the sole treatment for cancer therapy without the use of potentially harmful co-treatments such as radiation or chemotherapy.

A significant portion of the cost of developing and producing a mAb therapeutic is the analytical and physicochemical characterization of the molecule. As the market grows, increasing demands continue to be placed on the analytical technology used for assessing mAb structure. It is in these situations, where speed and ease of use are of increasing importance, that the use of microfluidic devices is greatly beneficial. Microfluidic approaches have shown particular utility in the integration of multiple steps of complex analytical protocols. This chapter will highlight some of the key implementations of microfluidics in the area of mAb characterization. As examples, we present an analysis of the NISTmAb reference material (RM) using three such analytical platforms: the Agilent HPLC-chip liquid chromatography/mass spectrometry (LC/MS) system, the Agilent 2100 Bioanalyzer system, and the Agilent 7100 Capillary Electrophoresis system. The use of these platforms enabled rapid acquisition of structural data on the RM with short turnaround time and minimal manual sample manipulation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Structural Feature of mAbs

mAbs are large (~150 kDa) molecules with complex three-dimensional structure. They are composed of four polypeptide chains, including two heavy chains (~50 kDa each) and two light chains (~25 kDa each). Both the heavy chains and the light chains are comprised of variable and constant domains. mAbs can be described by their component functional units, which include the Fab (fragment antigen-binding) region and the Fc (fragment crystallizable) region. mAb subclasses (IgA, IgD, IgE, IgG, IgM) are based on the amino acid sequences of the heavy chain (*14*), and most therapeutic mAbs belong to the IgG subclass. A detailed review of mAb structure can be found in (*15*) and (*16*).

One of the most compelling features of mAbs as therapeutic molecules is their ability to target specific antigens, which is a function of the enormous diversity

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

in the complementarity determining regions (CDRs). Other features, such as effector functions important in antibody-dependent cell-mediated cytotoxicity (ADCC) (*17*) and complement-dependent cytotoxicity (CDC), two of the principal mechanisms by which mAbs function as therapeutics (*18*), are dependent on the Fc region. Indeed, the unique function of mAbs and in fact their success as drugs can be attributed to their complexity. While this feature is certainly appealing from a functional standpoint, it has posed significant challenges in the characterization of these drugs, especially when compared to their small-molecule counterparts.

Ideally, a mAb population would be completely homogeneous. That is, every molecule would have the exact same amino acid sequence, post-translational modifications, 3D structure, etc. Owing to the fact that mAbs are biosynthesized from cell culture, though, they are structurally heterogeneous. For example, sequence variations can be present in low levels, resultant from DNA mutations or errors in translation (*19*). Furthermore, chemical modifications can occur during any of the many steps of bioprocessing or formulation/storage. Disulfide bond shuffling, N-terminal cyclization, variations in C-terminal lysine processing, and oxidation are all examples of myriad modifications that can be found on mAbs (*20*).

A particularly important modification that contributes to mAb heterogeneity is glycosylation. Human IgG molecules contain a conserved *N*-linked glycosylation site, located in constant domain 2 of the Fc region. Typical of protein glycosylation in general, mAb glycosylation sites are occupied by a mixture of glycans, a feature known as glycosylation site microheterogeneity. The majority of therapeutic mAbs are modified by relatively simple glycan distributions that are typically complex-type, biantennary, core fucosylated (α1,6-linked), and which terminate with 0-2 galactose units. Structures containing one terminal sialic acid are sometimes found in low abundance, while structures containing two terminal sialic acids are rarely detected (*21*). For a detailed review of mAb glycosylation the reader is directed to a review by Jefferis (*22*).

mAb glycosylation is critical for the interaction of the Fc region with Fc receptor γ, and deglycosylated antibodies completely lose their effector function (*23*). mAbs containing high mannose *N*-glycans are cleared more quickly from the bloodstream than those with complex-type structures (*24*). Glycans containing bisecting GlcNAc (*25*) or those that lack core fucose have been shown to induce strong ADCC (*26*). Importantly, non-human glycans or glycan epitopes can confer immunogenicity to mAbs. Such modifications result from mAb production in non-human cells, but even when human cells are used, glycosylation can be drastically influenced by the particular cell culture conditions (*27*). Highly immunogenic Galα1,3Gal epitopes were found on a mAb produced in murine hybridoma cells (*28*). Chinese hamster ovary cells are capable of producing *N*-glycolylneuraminic (NeuGc) acid containing glycoproteins, and cell culture conditions have been shown to alter the proportion of this monosaccharide on a recombinant protein produced in this expression system (*29*). Special attention must be paid to NeuGc levels in mAbs, since humans contain antibodies to this non-human monosaccharide (*30*). Prolonged treatment with mAbs, particularly those with high concentration formulations, could thus be harmful to patients who

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

happen to have high levels of anti-NeuGc antibodies. An additional possibility exists that initially low levels of anti-NeuGc antibodies can be raised by repeated exposure to NeuGc containing proteins. Even in the absence of an immune response, NeuGc incorporation into mAbs could be undesirable, as it has been shown to increase the clearance rate of Cetuximab (a mAb which contains NeuGc) in mice which are deficient in NeuGc production but which were previously immunized with a NeuGc containing epitope (*31*).

While many of the modifications observed on mAbs are natural, a subset can be introduced in a variable fashion as a function of cell culture conditions, formulation, or storage. It is important that all modifications, however introduced, are measured as they can ultimately affect the immunogenicity and efficacy of the molecule. The need for structural characterization of mAbs is intensified by the fact that the effects of many modifications on the safety of an individual drug cannot be predicted. mAb analysis is a challenging task from an analytical standpoint, especially when compared to the analysis of small molecule drugs. The large size of mAbs certainly increases the difficulty of analysis, but moreover, mAbs can be considered as a mixture of molecules having various modifications. It is this feature in particular that places the greatest demand on the analytical tools used to characterize mAb therapeutics.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Microfluidics and Analysis of Monoclonal Antibodies

Microfluidic devices offer many advantages in a variety of analytical contexts. In some cases, microfluidic technologies provide a cost effective means of performing analyses, especially on disposable devices with low cost to manufacture. In others, the use of microfluidics confers portability, which is useful for applications such as point-of-care diagnostics. In the bioanalytical fields, a major driving force for the adoption of microfluidics-based approaches is the minimization of sample consumption. An additional driving force is the need for higher sample throughput, and the miniaturization possible in microfluidic devices yields platforms capable of processing many samples in a highly parallel fashion.

In the context of recombinant protein analysis, sample limitation is an exception rather than a rule. Nonetheless, microfluidics-based research equipment has made considerable contributions to the field. The main benefit in this context is the integration of multiple steps of sample preparation or analysis. Such "plug and play" devices are generally simple to operate, and relieve the researcher of laborious bench-based protocols. It should be noted that antibodies in general have been integrated into a variety of microfluidic devices for research and measurement purposes, for example in the field of biomarker discovery (*32*). The focus in this chapter, however, is on microfluidic devices used to characterize biotherapeutic mAbs. In the following section, we highlight as examples the GE Biacore, the Agilent 2100 Bioanalyzer, and the Agilent HPLC-chip cube.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

mAb Analysis Using Microfluidics-Based Surface Plasmon Resonance

A critical piece of information which is necessary early in the development of a particular mAb therapeutic is the nature of its binding to the intended target. In particular, the measure of dissociation constant (K_D) is vital, as this is a key parameter that affects the pharmacokinetics of the drug. Surface plasmon resonance (SPR) has been particularly useful for antibody interaction analysis (*33–35*), and the first commercial instrument making use of this approach, the GE Biacore, was introduced in 1990. In a Biacore experiment, a molecule of interest (for example an antigen or a mAb) is covalently bound to the surface of a sensor chip, which is then pressed against an integrated microfluidic cartridge (IFC) to form a flow cell. In this arrangement, the sensor chip surface forms one wall of the flow cell while the IFC provides the remaining three sides. The design of the IFC allows delivery of continuous, smooth fluid flow of constant analyte concentration across the sensor chip. An optical detector measures changes in the refractive index of the solution in contact with the sensor chip that occurs upon binding of a molecule in solution to the immobilized molecule. The system thus permits real-time study of molecular interactions. Association and dissociation events can be measured and used to assess specificity, concentration, and binding kinetics.

A variety of mAb analysis assays have been developed on the Biacore platform. For example, Biacore has been used to analyze how succinimide formation in the CDR affects mAb target binding (*36*). Assays to investigate the immunogenicity of therapeutic mAbs have also been developed and have been shown to be useful for the sensitive detection of anti-drug antibodies from patient samples (*37*, *38*). Recently, Biacore SPR has been used to investigate the binding of Fcabs (Fc regions with antigen binding properties) (*39*) to target antigens, Fc receptor γ, and protein A as a function of glycosylation (*40*).

Canziani *et. al.* developed a platform for screening mAb concentration from hybridoma culture supernatants by using an anti-Fc specific antibody which was immobilized to the sensor chip surface. Known concentration of antigen could then be delivered to the captured mAbs and affinity constants determined for each interaction (*41*). This approach permitted the characterization of 100-200 supernatants per day. Further evolutions in Biacore SPR technology, such as the A100 SPR system that utilizes a hydrodynamic addressing flow cell system, have enabled characterization of mAbs from nearly 400 hybridoma samples within a 12 hour period (*42*). Such throughput is achievable through the process of analyzing multiple interactions per flow cell, which is possible by adjusting the relative flow of the two microfluidic inlets of the cell. Platforms such as the Biacore system are enabled by microfluidic technology, which continues to increase the throughput of characterization of mAbs and other biomolecules.

Application of Microfluidics-Based LC/MS to the Analysis of mAbs and Other Complex Biomolecules

The first commercially available microfluidics-based instrument for LC/MS analysis appeared in 2005 (*43*). The HPLC-chip platform, developed by Agilent

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies, was based on polyimide chips which integrated the components necessary to perform nano-LC/MS experiments: an enrichment column, a separation column, and an electrospray tip. These components, as well as other features of the device, are depicted in Figure 1. The incorporation of components onto a single microfabricated device eliminated the need for the fittings and capillaries required to connect these parts together in a typical implementation of nano-LC/MS. This integration represents one of the major advantages of microfluidic devices for LC/MS, and particularly for low flow applications, as a great deal of the user's time can often be invested in making such connections (e.g. cleanly cutting fused silica) and troubleshooting the system (e.g. locating leaks).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

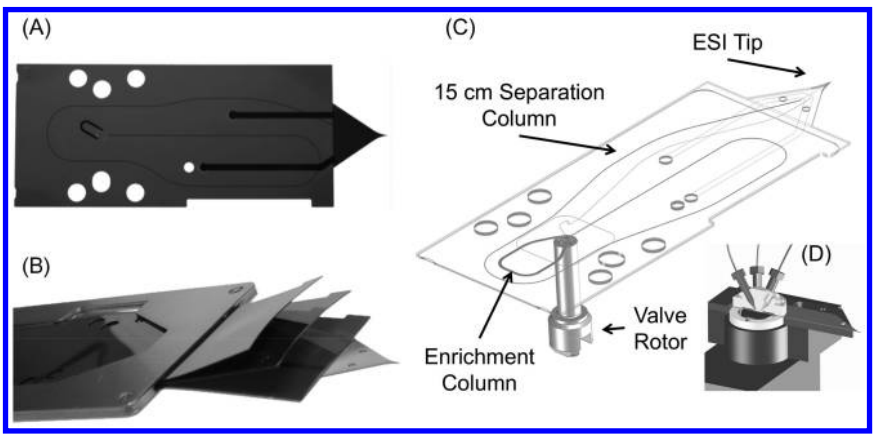

Figure 1. (A) Photograph of a standard HPLC-chip microfluidic device. The enrichment column (150 nL as pictured), analytical column (75 μm x 15 cm as pictured), alignment holes, electrical contacts for ion source ground, and ESI tip are shown. The chip measures ~ 6.5 cm in length. (B) Three HPLC-chip layers (including one spray chip layer) are shown protruding from their metal housing which orients them in the Chip Cube ion source. Chips are face-sealed together to form microfluidic connections between the different layers which can be added or removed to perform various functions. (C) Drawing of an HPLC-chip with reference to the inner 6-port valve rotor which face seals to the device by pushing against a stator (not shown). The drawing is similar in architecture to the photograph in (A) with the exception that the enrichment column is 500 nL and an optional flow path for post-column make-up flow is pictured. (D) Cutaway showing the fluidic connections from the LC to the HPLC-Chip. Capillaries are connected to the stator of the Chip Cube.

In the simplest implementation of the device, sample enrichment and analysis steps are controlled by a 6-port rotary valve which is face-sealed to the chip surface (Figure 1C and 1D). This seal is formed robotically, in a device called the Chip Cube, by sandwiching the chip layers between the rotary valve and the stator.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Fluidic connections from the LC system are then connected to the stator (Figure 1D). A schematic of a typical HPLC-chip is shown in Figure 2. The chip contains a 40 nL enrichment column in addition to a 75 µm x 150 mm separation column. The design allows sample to be loaded over the enrichment column followed by low delay-volume transfer of the sample onto the separation column. An additional 10-port outer rotor (not shown in the figure), which is concentric with the inner 6-port rotor, provides significant flexibility in terms of possible chip-based workflows. The first application demonstrated with the HPLC-chip was peptide separation and identification (*43*). Since its introduction, the HPLC-chip has been used for many proteomics applications, for example in the identification of proteins from epithelial lining fluid of patients with chronic obstructive pulmonary disease (*44*) and to measure proteins from cerebrospinal fluid as a function of delayed storage time (*45*). Such proteomics experiments often require maximum sensitivity, and this is achieved by low flow rate sample introduction into the MS system via the chip device (300 nL min^{-1}).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

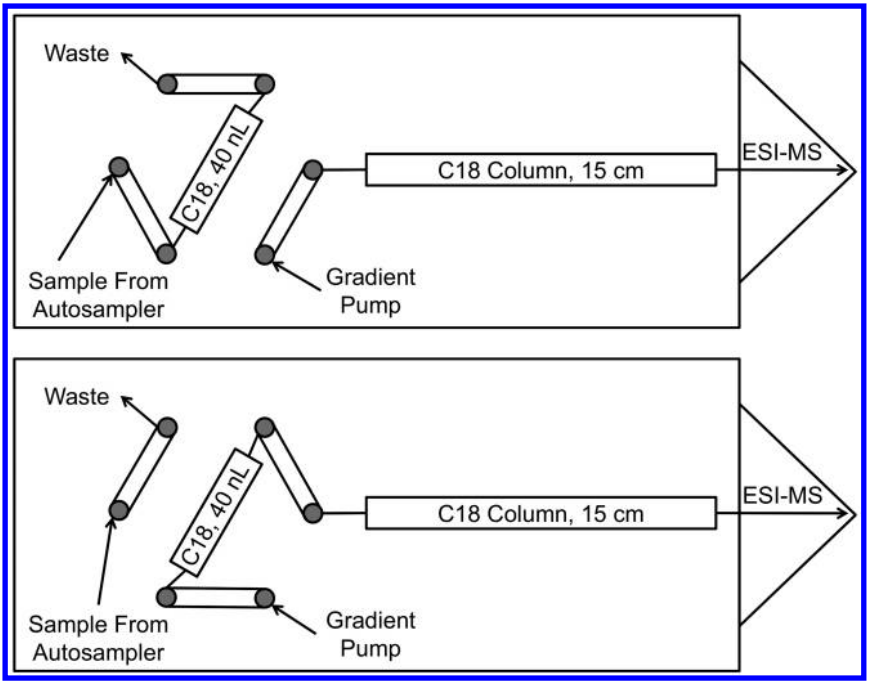

Figure 2. Schematic of a typical HPLC-chip for peptide separations. The inner 6-port rotor of the Chip Cube can be programmatically switched for sample loading and desalting (top schematic) and sample separation (bottom schematic). Samples are loaded onto the HPLC-chip using a low delay volume autosampler.

An example of the capability of HPLC-chips to integrate multiple steps of a workflow is illustrated by the use of these devices for phosphopeptide analysis by Heck's group (*46*). Like other HPLC-chips designed for peptide separation, the

device used in this work contained a 75 μm C18 separation column. However, the enrichment column was a "sandwich" design of the format C18-TiO_2-C18. A protein digest is loaded onto the chip and all peptides bind to the first C18 enrichment column. After desalting, the chip rotor is switched and gradient elution proceeds. Non-phosphorylated peptides are separated and detected during this run, whereas phosphorylated peptides are trapped on the TiO_2 enrichment column. In a subsequent run, a plug of elution buffer is injected over the sandwiched enrichment columns, which elutes the phosphorylated peptides from the TiO_2 and on to the second C18 enrichment column. Gradient elution is then used to separate the phosphopeptides before MS detection. The described design has been used to successfully profile over 1000 phosphopeptides from primary human leukocytes (*47*).

HPLC-chips which contain porous graphitized carbon (PGC) enrichment and separation columns have been used extensively in the investigation of free glycans from complex biological matrices (*48–50*). A recent example is the profiling of ~300 *N*-glycans from mouse serum, including previously undocumented variants of sialic acid containing *N*-glycans modified by lactyl groups (*51*). Particular focus has been placed on glycoprofiling of human milk, and HPLC-chip technology has been instrumental in analyzing milk glycans from various stages of lactation (*52*, *53*) and as a function of blood type and secretor status (*54*). This work has generated new hypotheses about the influence of human milk oligosaccharides on the gut microbiome of developing infants. Recent work has extended the capabilities of the PGC HPLC-chip MS system, in particular the generation of informative fragment ions from CID MS/MS, using ammonium fluoride as a buffer additive (*55*).

HPLC-chips have been used for a variety of other applications. An HPLC-chip capable of passive gradient generation (using only an isocratic pump) was designed and successfully utilized for peptide separations with gradient-delay times even lower than standard HPLC-chips. Such an experiment is feasible because the dead-volume could be reduced to the low nanoliter range using the chip architecture (*56*). LC/MS of glycosaminoglycans (GAGs), highly sulfated and very hydrophilic polymers that have important roles in tissue structure and cell-signaling events, were also analyzed using an HPLC-chip packed with HILIC material (*57*). The chips used in this work enabled routine and robust low-flow analysis of GAGs in the negative-ion mode. Further improvements in the workflow were demonstrated by post-column addition of organic solvent (make-up flow) via an additional microfluidic flow path (*58*). Using this technology, the most highly sulfated domains of heparan sulfates (HS) were profiled. Later, the chip was used to determine that HS from various mammalian sources terminates with long, highly-sulfated domains, a feature that has important implications for the participation of these macromolecules as co-receptors at the cell surface (*59*). HPLC-chip technology drove additional advancements in HS analysis in 2011, when a chip that enabled post-column pulsing of metal cations or supercharging reagents was used to greatly improve tandem MS of HS precursor ions (*60*). Such results were made possible because the chip could be programmed to introduce these reagents only during elution of a peak of interest. These experiments would be very difficult otherwise, as

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the reagents introduced were non-volatile and their continuous infusion would negatively impact the performance of the MS system.

A workflow for the analysis of *N*-glycans from mAbs has also been integrated on an HPLC-chip (*61*). This microfluidic device, called the mAb-Glyco Chip, utilizes both the inner 6-port and the outer 10-port rotors of the Chip Cube. In this workflow, intact mAbs are injected onto the device where they are subjected to deglycosylation in an enzyme reactor packed with immobilized PNGase F. This mode of glycan release is very efficient and proceeds to completion within a period of a few minutes, in contrast to the more typical in-solution release of glycans from mAbs which is allowed to proceed for a period of many hours or even overnight. Subsequently, the *N*-glycans released from the mAb are trapped on a PGC enrichment column, where desalting of the deglycosylation buffer occurs. Finally, the *N*-glycans are separated on a 43 mm PGC column before introduction into the MS. A recent study has demonstrated the utility of the mAb-Glyco Chip platform for measuring differences in glycosylation between an FDA-approved mAb and a biosimilar product (*62*).

Microfluidics-based approaches for LC/MS have led to significant advances in basic research and in the area of biotherapeutic characterization. The HPLC-chip system is capable of delivering high sensitivity sample introduction into the MS system, which is beneficial when sample quantities are limited. In other cases, the HPLC-chip system offers the integration of complex workflows which would otherwise be laborious in terms of set-up or manual execution. In addition to the Agilent HPLC-chip system a variety of other microfluidic devices for LC/MS are now available, such as those from Eksigent (AB SCIEX), Waters, and Advion.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Microfluidic Electrophoresis of mAbs

Microfluidic chip-based protein separation was demonstrated on a planar glass chip in 2001, and was shown to be a faster alternative to traditional SDS-PAGE separations while providing equivalent information (*63*). The technology, developed by Caliper Life Sciences and utilized first in the Agilent 2100 Bioanalyzer (and later in the Experion Automated Electrophoresis System and the Caliper LabChip 90 and LabChip GXII), integrated multiple sample preparation and analysis steps. To begin the workflow, denatured protein sample in complex with SDS is loaded onto the microfluidic device, shown in Figure 3, and mixed in order to achieve non-covalent association with a fluorescent label. Next, the proteins are separated by size in a channel filled with a polydimethylacrylamide-based matrix. En route to the detector, the protein-SDS complexes are electrophoretically diluted via an intersecting microfluidic channel that joins the flow path of the sample. This step eliminates background signal caused by fluorescent dye that is bound to SDS micelles, which do not contain protein. The protein staining and SDS micelle dilution coordinated by the 2100 Bioanalyzer are completed orders of magnitude faster than the destaining step required for SDS-PAGE, accounting for the principal advantage of this chip-based method. The platform also provides tools for quantitative analysis as well as data visualization in the form of digital gel-like images.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

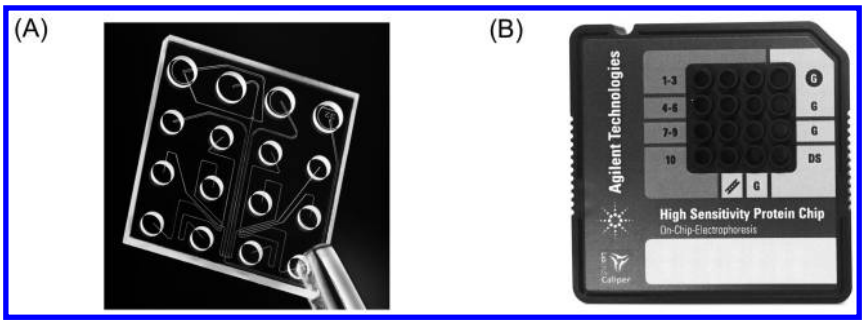

Figure 3. Bioanalyzer chip. (A) Microfluidic flow paths of the chip used in the Agilent Bioanalyzer. A detailed description of the flow paths can be found in (63). (B) Design of consumable chip for microfluidic protein sizing. There are 9 wells for samples in addition to those for the reference ladder, gel (G) and destaining solution (DS).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Since its appearance on the market, microfluidic electrophoresis has been used for various aspects of mAb analysis. In 2002, traditional quantification and size determination of mAbs from hybridoma cell culture using the 2100 Bioanalyzer or traditional SDS-PAGE was compared. Use of the microfluidic platform provided the obvious benefits of speed, but also avoided overestimation of mAb concentration from serum-containing cell culture medium that was observed using traditional SDS-PAGE analysis (*64*). High throughput microfluidic electrophoresis was performed on mAbs treated with Endo H followed by reduction in order to quantify the proportion of hybrid and high mannose-type *N*-glycans (*65*). The proportion of non-glycosylated heavy chains present in mAb formulations has also been measured (*66*). Half-antibodies, which are composed of a heavy chain and a light chain that are not disulfide bonded to another heavy chain:light chain pair, can be produced from cell culture but can also be created as artifacts during analysis steps (*67*). Methods to determine the proportion of half-antibodies in IgG4 samples have been developed using the Bioanalyzer system (*68*, *69*). Bioanalyzer technology is also being applied to the study of ADCs. An example is the characterization of cysteine-treated THIOMABs as a means of assessing the charge heterogeneity inherent to these molecules (*70*).

Platforms like the 2100 Bioanalyzer for microfluidic electrophoresis of proteins such as mAbs continue to gain popularity. The acceptance of this technology has largely been driven by dramatic increases in sample throughput, elimination of laborious steps such as gel pouring and gel destaining, and increased precision and accuracy of quantitation when compared to traditional SDS-PAGE methods.

Capillary Isoelectric Focusing of mAbs

Capillary isoelectric focusing (cIEF) is a microscale separation method useful for the analysis of many biological molecules including DNA, glycans,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

and proteins. cIEF is a variant of capillary electrophoresis, and in the context of mAb analysis, offers a rapid, high resolution method of performing charge variant analysis, which can arise from deamidation, glycation, C-terminal lysine processing, and multiple other sources. Similar to conventional gel-based IEF methods, separation of charge variants using cIEF is achieved by application of an electric field to the protein sample in a pH gradient. However, the separations are performed in narrow-bore capillaries, which are typically 50 μm in diameter and chemically coated to prevent protein absorption and undesirable electroosmotic flow. The advantages of cIEF include high resolving power, quantitative capability, high speed of analysis, and automated sample introduction. Many advances have been made in terms of the performance of cIEF since the first demonstration in 1985, the details of which have been extensively reviewed Kaheki and Shumura. (*71–73*).

cIEF experiments utilize either one or two-step methods. In these approaches, protein charge variants are focused and mobilized toward the detector simultaneously or in sequence. Mobilization toward the detector can be achieved using either hydrodynamic or electrophoretic approaches. As an alternative to the mobilization of separated charge variants toward a single-point detector, methods of whole-column imaging have been demonstrated, and a commercial platform for performing this experiment (*74*) is available from ProteinSimple (formerly Convergent Bioscience).

cIEF has recently been optimized for the separation and quantification of ADCs (*75*), including for the estimation of unconjugated antibody in ADC preparations (*76*). Cysteinylation and glutathionylation, undesirable modifications of engineered cysteines, have also been measured in THIOMAB preparations using cIEF (*70*). Additionally, the technology has been used for the comparison of Trastuzumab and biosimilars thereof (*77*). In high-throughput form, cIEF has also been evaluated in formulation development (*78*). cIEF has also been hyphenated with a variety of other measurement techniques, including MS. Zhu *et. al.* demonstrated a cIEF-MS/MS method for the identification of host-cell proteins from an immunodepleted mAb sample (*79*).

We point out that cIEF experiments are performed in capillaries using capillary electrophoresis platforms, and one can argue whether this qualifies as a microfluidic device or not. Regardless, there has been significant effort made in performing IEF in devices which are unambiguously microfluidic. In a recent example (*80*), chip-based separation of mAb charge variants was performed in under 7 minutes, which was made possible by a short separation channel coupled with whole column imaging. An extensive review of IEF experiments in microfluidic devices can be found in (*81*). Whether the separation occurs in a capillary or in a chip, the benefits over traditional IEF experiments are similar. Besides the fact that gel-based methods are time consuming and lower in resolution, quantification of separated bands is difficult. Comparatively, cIEF represents a faster method that is both quantitative and highly reproducible. It is envisioned that cIEF will continue to play a critical role in stability testing, product characterization, lot release, and other steps in the production of the next generation of antibody therapeutics.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Methods of Analysis for the NISTmAb Reference Material Using HPLC-Chip MS, Microfluidic Electrophoresis, and Capillary Isoelectric Focusing

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Tryptic Digestion

The NISTmAb RM was digested using a protocol optimized for rapid and efficient trypsin digestion (*82*). Briefly, 450 µL of 7.5 M guanidine HCl, 0.25 M Tris was added to a 50 µL aliquot of 10 mg/mL NISTmAb RM. The sample was reduced by addition of 3 µL of 0.5 M dithiothreitol (DTT) for 30 min at room temperature. The sample was then alkylated by addition of 7 µL of 0.5 M iodoacetic acid (IAA) at room temperature in the dark. The alkylation reagent was quenched by addition of 4 µL of 0.5 M DTT. The total volume of sample was then loaded onto a NAP-5 column (GE Healthcare) that was equilibrated with 15 mL of 100 mM ammonium bicarbonate. The sample was eluted with 600 µL of 100 mM ammonium bicarbonate. As a consequence of complete removal of the guanidine via the NAP-5 column, ~40% of the protein was lost (see cited reference). 20 µL of 1 mg/mL sequencing grade modified trypsin (Promega) was added to the desalted solution, resulting in an enzyme:protein ratio of 1:25 (w/w). The digestion reaction was allowed to proceed for 30 min at 37° C before quenching the trypsin by addition of 5 µL of 20% formic acid (FA).

LC/MS Analysis

All LC/MS analyses of the NISTmAb RM described in this work were conducted on a 6550 QTOF LC/MS system (Agilent Technologies) or a 6224 TOF LC/MS system (Agilent Technologies) equipped with an HPLC-Chip Cube ion source. All LC solvents were of the highest purity available. All instrument parameters (e.g. ion optics voltages) were default values unless otherwise noted. Relative quantification of glycans and glycopeptides was performed using the area under the curve of relevant extracted ion chromatograms.

Peptide Analysis

Tryptic peptides were analyzed using a Polaris-HR HPLC-chip (Agilent Technologies) on a 6550 QTOF LC/MS system. Solvent A was 0.1% FA in H_2O and solvent B was 0.1% FA in ACN. Samples (a volume of 0.2 µL prepared as previously described in *Tryptic Digestion*) were loaded onto the enrichment column in 97% A at 2 µL min^{-1} with an injection flush volume of 4 µL to ensure complete enrichment and desalting. Subsequently, the Chip Cube rotor was switched to analysis mode and a gradient from 3-42% B over 30 min was initiated. The separation flow rate was 300 nL min^{-1}. Collision-induced dissociation (CID) was performed with MS and MS/MS acquisition rates of 6 and 3 spectra/sec, respectively. Collision energy was applied in a data-dependent fashion using the following equation: collision energy voltage = ((3.6 * *m/z*) / 100) - 4.8. This setting is recommended for peptide fragmentation on the instrument used. A limit of 10 precursors per cycle was allowed with active exclusion after 1 spectrum

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

per precursor. MS/MS results were interpreted using Morpheus software with a maximum false discovery rate of 1% (*83*).

Glycan Nomenclature

The *N*-glycans measured in this study are described as compositions, which do not provide information on glycosidic linkage. The compositions include the number of hexose (H), *N*-acetylhexosamine (assumed in these experiments to be N-acetylglucosamine, N), deoxyhexose (assumed in these experiments to be fucose, F), and *N*-glycolylneuraminic acid (Sg) residues. We also use the glycan nomenclature utilized by the biopharmaceutical industry for typical mAb glycans (G0F, G1F, etc) when applicable. Using the nomenclature described above, G1F is represented by H4N4F1 (4 hexose, 4 *N*-acetylhexosamine, 1 fucose).

Glycopeptide Analysis

Glycopeptides were analyzed using an HPLC-chip which consisted of a 40 nL HILIC enrichment column (packed with 5 μm Amide-80, Tosoh Biosciences) and a 75 μm x 15 cm separation column packed with 3 μm Polaris C18 particles (Agilent Technologies). The analysis was performed on a 6550 QTOF LC/MS system. In order to prepare the sample for injection, the previously described tryptic digest was diluted 1:20 in 80% ACN. 1 μL of this solution was injected per run. Samples were loaded onto the enrichment column in 80% ACN containing 0.1% TFA (*84*) at a flow rate of 4 μL min^{-1} and an injection flush volume of 4 μL. For the separation, solvent A was 0.1% FA in H_2O and solvent B was 0.1% FA in ACN. The separation flow rate was 300 nL min^{-1}. The HPLC-chip rotor was switched and purified glycopeptides were automatically focused onto the head of the C18 separation column by 97% A. Subsequently, glycopeptides were separated by a 300 nL min^{-1} gradient that reached 42% B in 10 minutes. At the beginning of the gradient, the enrichment column was switched back in line with the sample loading pump in order to equilibrate the enrichment column for the next run.

Released Glycan Analysis

Released glycans were analyzed using a mAb-Glyco Chip (Agilent Technologies) operating on a 6224 TOF LC/MS system. The mAb-Glyco Chip consisted of a 360 nL enzyme reactor packed with immobilized PNGase F beads, a 160 nL PGC enrichment column, and a 43 mm PGC separation column. 2 μL of a 0.25 mg/mL solution of the NISTmAb RM was injected onto the device. The chip was programmed to operate in "heart-cutting" mode where only a portion (150 nL, equivalent in this case to 37.5 ng of mAb) of the injected sample was subjected to deglycosylation in the enzyme reactor. Following a 4 minute incubation period in the reactor, the released glycans were captured on the PGC enrichment column. Subsequently, they were separated on the PGC separation column. For the separation, solvent A was 0.1% FA in H_2O and solvent B was 0.1% FA in ACN. Glycans were separated using a gradient which started at 3% B and reached 32% B in 1.5 min. The %B was then raised to 85% over the course of

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

another minute before column washing and equilibration steps. Due to the speed of the PNGase F release and separation, free glycans were detected primarily in their glycosylamine form (*85*), with only a minor contribution from their free reducing end form. Data analysis was performed automatically using Agilent MassHunter Software.

MS Analysis of Intact mAb before and after PNGase F Treatment

Analysis of deglycosylated (PNGase F treated) NISTmAb RM was performed using a modified mAb-Glyco Chip operating on a 6224 TOF LC/MS system with the fragmentor set to 420 V. The chip consisted of a 360 nL enzyme reactor packed with immobilized PNGase F beads, a 160 nL PLRP-S enrichment column, and a 43 mm PLRP-S separation column. In contrast to the mAb-Glyco Chip described in the previous section, this chip was designed to analyze the protein (rather than glycan) portion of the mAb. To analyze deglycosylated mAb, 2 μL of a 0.125 mg/mL solution of the NISTmAb RM was injected onto the device. The chip was programmed to operate in "heart-cutting" mode where only a portion (150 nL, equivalent in this case to 18.75 ng) of the injected sample was subject to deglycosylation in the enzyme reactor. Following a 4 minute incubation period in the reactor, the deglycosylated mAb was desalted and separated on the PLRP-S enrichment column and separation column, respectively. For the separation, solvent A was 0.1% FA in H_2O and solvent B was 0.1% FA in ACN. The gradient used for the analysis was as follows: t = 0, %B = 3; t = 1, %B = 20; t = 3, %B = 60; t = 5, %B = 85. This gradient was followed by sufficient column washing and equilibration steps for reproducible analysis with minimal carry-over.

To analyze the non-PNGase F treated mAb, the exact same chip and method was used, except that the enzyme reactor of the chip was programmatically bypassed. For injection, the mAb sample was diluted significantly in order to obtain the same amount of protein on column as was used for the analysis of the deglycosylated protein. Summed mass spectra of the intact and deglycosylated mAb were generated using MassHunter software. The spectra were deconvoluted using the maximum entropy algorithm with a mass step of 1 Da. Deconvolution was performed using automatic isotope width and results were calculated using the top 25% of the peak height.

MS Analysis of IdeS-Generated mAb Fragments before and after PNGase F Treatment

The NISTmAb RM was treated with IdeS (also known as FabRICATOR, Genovis Inc.) in order to generate $(Fab')_2$ and Fc/2 fragments. 20 μg of the RM was incubated with 20 U of the enzyme for 1 hr at 37° C. Analysis of the IdeS products was performed using the same modified mAb-Glyco Chip and LC/MS system as was used for the intact mAb analysis except that the fragmentor was set to 175 V for Fc/2 fragments and 350 V for $(Fab')_2$ fragments. The sample was diluted to 0.0625 mg per mL prior to injection, which equates to 9.4 ng of IdeS products that were subject to analysis. The mobile phases and gradient for the separation were also different from those used for the intact mAb analysis:

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

solvent A was 0.5% AcOH, 0.05% TFA in H_2O and solvent B was 0.5% AcOH, 0.05% TFA in 80:10:10 1-propanol: H_2O:ACN. The gradient used for the analysis was as follows: t = 0, %B = 3; t = 11, %B = 100; t = 13, %B = 100; t = 14, %B = 3. This gradient was followed by sufficient column washing and equilibration steps for reproducible analysis with minimal carry-over.

Summed mass spectra of the mAb fragments before and after automatic deglycosylation were generated using MassHunter software. The spectra were deconvoluted using the maximum entropy algorithm with a mass step of 1 Da for $(Fab')_2$ or 0.1 Da for Fc/2. Deconvolution was performed using automatic isotope width and results were calculated using the top 25% of the peak height.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Bioanalyzer Analysis

The NISTmAb RM was analyzed using a 2100 Bioanalyzer (Agilent Technologies) with both the Protein 80 and HSP-250 kits. Samples (starting concentration of 10 mg/mL) and chips were prepared according to the manufacturer's instructions. For the experiments using the Protein 80 kit, mAb samples were reduced using DTT. For the experiments using the HSP-250 kit, mAb samples were covalently modified with a fluorescent dye (again following the manufacturer's instructions) and analyzed without DTT treatment. Data analysis was performed using 2100 Expert Software.

Capillary Isoelectric Focusing

The cIEF method used to analyze the NISTmAb RM was adapted from previously published protocols (*75*, *77*). The separation was conducted using a fluorocarbon-coated 75 μm internal diameter capillary that was cut to a total length of length of 32.5 cm to give an effective length of 24.5 cm. The capillary was installed in an Agilent 7100 Capillary Electrophoresis system that was equipped with a high pass optical filter. Prior to the analyses, the capillary was cleaned with 0.1 M NaOH for 2 minutes, followed by water for 30 minutes. The capillary was then conditioned by flushing with 350 mM acetic acid for 5 minutes followed by water for 2 minutes and then 0.5% methylcellulose (Sigma-Aldrich) for 5 minutes.

A cIEF gel solution was prepared by mixing the following: 30 μL of Pharmalyte 3-10 (ampholyte solution, GE Healthcare), 45 μL of 500 mM L-arginine (Sigma-Aldrich), 50 μL of 200 mM iminodiacetic acid (Sigma-Aldrich) and 1000 μL of an aqueous solution of 0.6% methyl cellulose and 3M urea (Sigma-Aldrich). Sample solutions were prepared by mixing 1 μL of 10 mg/mL NISTmAb RM solution and 1.5 μL of water with 57.5 μL of the cIEF gel solution. Samples were centrifuged at 13,000 g for 1 minute prior to analysis to remove gas bubbles.

The capillary was preconditioned prior to each analysis by flushing with 4.3 M urea at 3.5 bar for 3 minutes and then flushing with water at 3.5 bar for 2 minutes. The capillary was then filled with the sample solution at a pressure of 2 bar for 100 seconds. Focusing was carried out at 12.5 kV in normal polarity mode for 10 minutes using an anolyte of 200 mM phosphoric acid (Sigma-Aldrich) at the inlet

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

and a catholyte of 300 mM NaOH (Sigma-Aldrich) at the outlet. The analytes were mobilized past the detector by switching the outlet solution to 350 mM acetic acid and applying 30 kV in normal polarity mode for 30 minutes. The protein signal was measured during mobilization by a diode array detector at 280 nm (20 nm bandwidth) and a reference at 550 nm (100 nm bandwidth).

Results: Analysis of NISTmAb RM Using Microfluidic Technologies

Three commercially available instrument platforms which make use of microfluidic technology were used for structural characterization of the NISTmAb RM. These include the Agilent HPLC-chip LC/MS system, the Agilent 2100 Bioanalyzer and the Agilent 7100 Capillary Electrophoresis system. Using these platforms, it was possible to generate data that included peptide mapping, glycopeptide mapping, released glycan profiling, MS measurement of intact and deglycosylated intact protein and fragments, electrophoretic protein sizing, and charge variant analysis.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Peptide Mapping Using a Polaris-HR HPLC-Chip

Peptide mapping experiments were performed on the NISTmAb RM using an HPLC-chip designed for high resolution peptide separation. This Polaris-HR chip, which shares the design of the chip depicted in Figure 2, contains a 15 cm separation column packed with 3 μm reversed-phase functionalized particles. Operating at a flow rate of 300 nL min^{-1}, this workflow provides high sensitivity peptide detection. Analysis can routinely be performed on protein digests in the range of 50 fmol on column.

A chromatogram of the Polaris-HR separation of a tryptic digest of the NISTmAb RM is shown in Figure 4. Excellent chromatographic performance was achieved, with average peak widths of ~6 sec (measured at half height). This is despite the fact that the separation was driven by relatively low pressure, which did not exceed 120 bar during the course of the analysis. Tandem MS experiments on the RM tryptic peptides were used to construct a peptide map, shown in Figure 5. Sequence coverage was calculated based upon peptides which had unambiguous determination by MS/MS. Sequence coverage of 86.9% and 89.7% of the heavy chain and light chain was achieved, respectively. In this case, the sequence coverage of the heavy chain is artificially low, since it does not include the heavy chain glycopeptide, which was detected but not confirmed by MS/MS sequencing of the backbone. This is as expected for glycopeptides fragmented by the CID conditions used here. The peptide portion of glycopeptides can however be sequenced using the same instrument with higher collision energy CID (*86*). All of the peptides accounting for gaps in the sequence coverage are short, and many are very hydrophilic. Therefore, they may not have survived the sample enrichment step that was used as part of this workflow. In any case, a strategy to increase confident measurement of the full sequence might rely on the use of complementary proteases, for example treatment with chymotrypsin.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

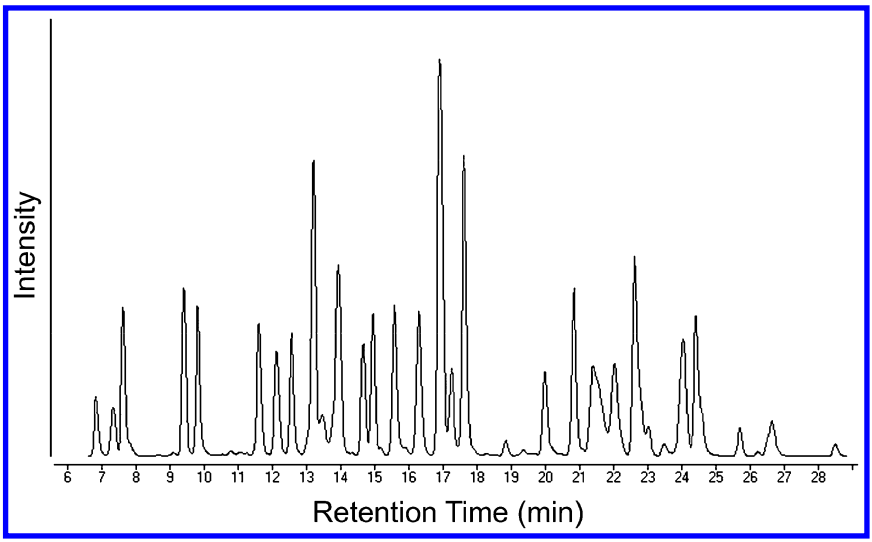

Figure 4. Ion chromatogram of a tryptic digest of the NISTmAb RM produced using a Polaris HR HPLC-chip.

The tryptic digest of the NISTmAb RM was prepared using a rapid protocol for proteolysis developed to limit artifacts often caused by sample preparation. This procedure, which permits tryptic digestion in 30 minutes, limits the amount of deamidation of asparagine residues, for example. Based on the observed results, this protocol appears to be very effective. The peptides detected in the experiment are summarized in Table 1, including any modifications observed. The peptides in the table were identified using a combination of both accurate mass and MS/MS. The data were searched using Morpheus software against the primary sequence of the RM, allowing for variable modifications that included: deamidation of Asn, oxidation of Met and Trp, pyro-Glu, alkylation using iodoacetic acid, and the presence or absence of C-terminal Lys. Separate entries are listed for each peptide and any modifications thereof.

The detected deamidated peptides were typically in low abundance compared to their unmodified counterparts. However, one must use caution in using MS ion signal as a gauge of peptide concentration since the ionization response of peptides with and without deamidation may be substantially different. Most of the deamidated peptides shown in the table co-eluted with their unmodified counterparts, save for some exceptions, such as light chain 126-141. For these peaks, the isotopic distribution of the precursor ion was different than what would be expected for an unmodified peptide (data not shown). This is the result of the isotopic distribution of a normal peptide superimposed with that from a deamidated peptide (0.984 Da higher).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

The C-terminal peptide of the heavy chain was detected with and without a terminal lysine. The N-terminal amino acid of the heavy chain was observed as the cyclization product pyro-glutamic acid. The 30 min chip-based separation proved very useful for confirming the primary sequence of the mAb RM. Detection and MS/MS confirmation of additional peptides and particularly variants thereof would be expected for a longer separation (e.g. 2 hrs), if the goal of the experiment is to comprehensively characterize all peptide peaks in the sample.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

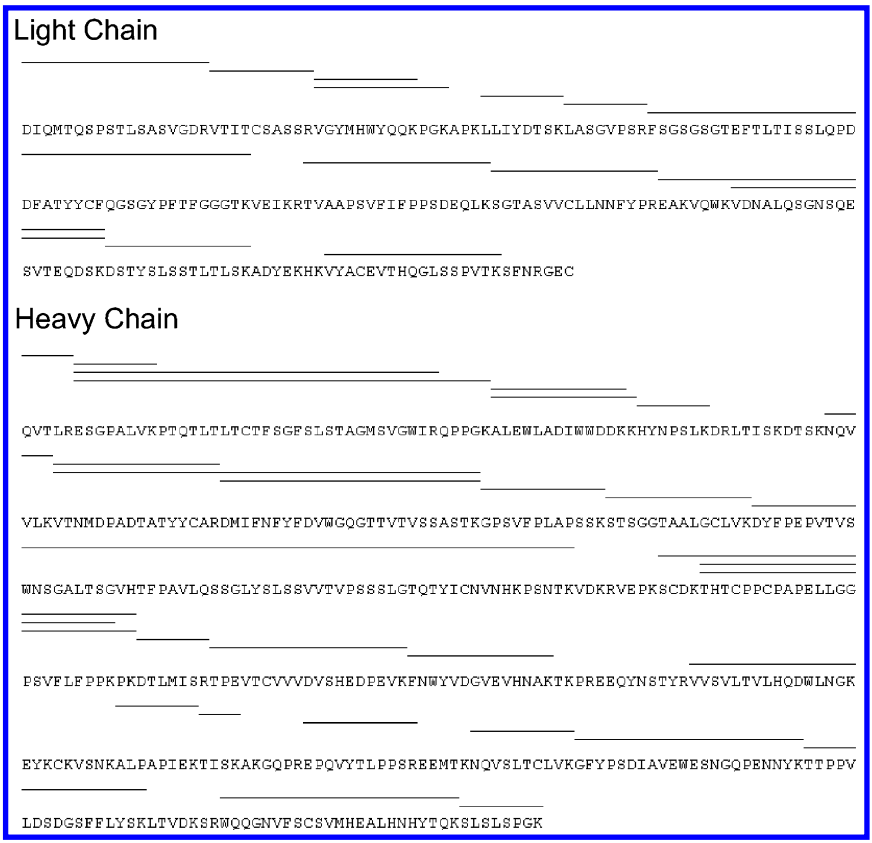

Figure 5. Peptide map of a tryptic digest of the NISTmAb RM produced using a Polaris HR HPLC-chip. The lines above the amino acid sequence correspond to peptides identified in the light chain and heavy chain. Samples were prepared according to the protocol in (82). Peptide ID was confirmed using MS/MS for all listed peptides. Peptide MS/MS data were analyzed using Morpheus (83).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Table 1. Peptides Detected in a Tryptic Digest of the NISTmAb RM Using the HPLC-Chip MS Platform. Alkylation by iodoacetic acid is designated by (IAA). RTs corresponding to the time of the MS/MS spectrum acquired for identification using Morpheus software.

RT	*Z*	*Peptide Sequence*	*Start Res.*	*Stop Res.*	*Missed Cleavages*
Heavy Chain					
11.8	2	-.(pyro-Q)QVTLR.E	1	5	0
9.0	2	R.ESGPALVK.P	6	13	0
25.6	3	R.ESGPALVKPTQTLTLTC(IAA)TF SGFSLSTAGMSVGWIR.Q	6	40	1
24.1	4	R.ESGPALVKPTQTLTLTC(IAA)TFS GFSLSTAGMSVGWIRQPPGK.A	6	45	2
26.6	2	K.ALEWLADIWWDDK.K	46	58	0
24.5	3	K.ALEWLADIWWDDKK.H	46	59	1
6.9	2	K.HYNPSLK.D	60	66	0
7.9	2	K.NQVVLK.V	78	83	0
14.7	2	K.VTNMDPADTATYYC(IAA)AR.D	84	99	0
14.6	3	K.VTNM(oxidation of M)DPAD TATYYC(IAA)AR.D	84	99	0
28.4	4	K.VTNMDPADTATYYC(IAA)ARDMIF NFYFDVWGQGTTVTVSSASTK.G	84	124	1

Continued on next page.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Table 1. (Continued). Peptides Detected in a Tryptic Digest of the NISTmAb RM Using the HPLC-Chip MS Platform. Alkylation by iodoacetic acid is designated by (IAA). RTs corresponding to the time of the MS/MS spectrum acquired for identification using Morpheus software.

RT	*Z*	*Peptide Sequence*	*Start Res.*	*Stop Res.*	*Missed Cleavages*
28.4	4	K.VTNMDPADTATYYC(IAA)A RDMIFN(deamidation of N)FY FDVWGQGTTVTVSSASTK.G	84	124	1
28.3	3	R.DMIFNFYFDVWGQGTTVTVSSASTK.G	100	124	0
16.3	2	K.GPSVFPLAPSSK.S	125	136	0
15.6	2	K.STSGGTAALGC(IAA)LVK.D	137	150	0
25.6	5	K.DYFPEPVTVSWNSGALTS GVHTFPAVLQSSGLYSLSSVVTV PSSSLGTQTYIC(IAA)NVNHKPSNTK.V	151	213	1
25.5	5	K.DYFPEPVTVSWNSGALTSGVHTFPAVLQSS GLYSLSSVVTVPSSSLGTQTYIC(IAA)NV N(deamidation of N)HKPSNTK.V	151	213	1
21.7	5	K.SC(IAA)DKTHTC(IAA)PPC(IAA)P APELLGGPSVFLFPPKPK.D	222	251	2
23.7	3	K.THTC(IAA)PPC(IAA)PAPELL GGPSVFLFPPK.P	226	249	0
22.1	4	K.THTC(IAA)PPC(IAA)PAPELL GGPSVFLFPPKPK.D	226	251	1
11.9	2	K.DTLMISR.T	252	258	0

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

RT	*Z*	*Peptide Sequence*	*Start Res.*	*Stop Res.*	*Missed Cleavages*
11.0	2	K.DTLM(oxidation of M)ISR.T	252	258	0
17.4	3	R.TPEVTC(IAA)VVVDVSHEDPEVK.F	259	277	0
16.9	3	K.FNWYVDGVEVHNAK.T	278	291	0
17.5	3	K.FNWYVDGVEVHN(deamidation of N)AK.T	278	291	0
16.9	3	K.FNW(oxidation of W)YVDGVEVHNAK.T	278	291	0
22.7	3	R.VVSVLTVLHQDWLNGK.E	305	320	0
23.0	3	R.VVSVLTVLHQDWLN(deamidation of N)GK.E	305	320	0
11.7	2	K.ALPAPIEK.T	330	337	0
7.7	1	K.TISK.A	338	341	0
13.4	2	R.EPQVYTLPPSR.E	348	358	0
17.0	2	K.NQVSLTC(IAA)LVK.G	364	373	0
18.2	2	K.N(deamidation of N)QVSLTC(IAA)LVK.G	364	373	0
20.1	3	K.GFYPSDIAVEWESNGQPENNYK.T	374	395	0
20.3	3	K.GFYPSDIAVEWES N(deamidation of N)GQPENNYK.T	374	395	0
20.9	3	K.TTPPVLDSDGSFFLYSK.L	396	412	0
16.9	5	R.WQQGNVFSC(IAA)SVMHEALHNHYTQK.S	420	442	0

Continued on next page.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Table 1. (Continued). Peptides Detected in a Tryptic Digest of the NISTmAb RM Using the HPLC-Chip MS Platform. Alkylation by iodoacetic acid is designated by (IAA). RTs corresponding to the time of the MS/MS spectrum acquired for identification using Morpheus software.

RT	*Z*	*Peptide Sequence*	*Start Res.*	*Stop Res.*	*Missed Cleavages*
16.9	5	R.WQQGNVFSC(IAA)S VM(oxidation of M)HEALHNHYTQK.S	420	442	0
16.8	3	R.WQQGNVFSC(IAA)SVMHEAL HN(deamidation of N)HYTQK.S	420	442	0
13.2	2	K.SLSLSPG.K	443	449	0
10.7	2	K.SLSLSPGK.-	443	450	0
Light Chain					
14.8	2	-.DIQMTQSPSTLSASVGDR.V	1	18	0
14.8	3	-.DIQM(oxidation of M)TQSPSTLSASVGDR.V	1	18	0
9.5	3	R.VTITC(IAA)SASSR.V	19	28	0
13.9	3	R.VGYMHWYQQK.P	29	38	0
13.7	3	R.VGYM(oxidation of M)HWYQQK.P	29	38	0
12.7	4	R.VGYMHWYQQKPGK.A	29	41	1
12.5	4	R.VGYMHW(oxidation of W)YQQKPGK.A	29	41	1
13.9	2	K.LLIYDTSK.L	45	52	0
7.7	2	K.LASGVPSR.F	53	60	0

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

RT	*Z*	*Peptide Sequence*	*Start Res.*	*Stop Res.*	*Missed Cleavages*
26.7	4	R.FSGSGSGTEFTLTISSLQPDDFAT YYC(IAA)FQGSGYPFTFGGGTK.V	61	102	0
21.5	2	R.TVAAPSVFIFPPSDEQLK.S	108	125	0
24.1	2	K.SGTASVVC(IAA)LLNNFYPR.E	126	141	0
21.1	2	K.SGTASVVC(IAA)LL NN(deamidation of N)FYPR.E	126	141	0
21.0	3	K.SGTASVVC(IAA)LL N(deamidation of N)NFYPR.E	126	141	0
15.1	4	R.EAKVQWKVDNALQ SGNSQESVTEQDSK.D	142	168	2
9.8	2	K.VDNALQSGNSQESVTEQDSK.D	149	168	0
17.3	2	K.DSTYSLSSTLTLSK.A	169	182	0
13.0	3	K.VYAC(IAA)EVTHQGLSSPVTK.S	190	206	0

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Glycopeptide Analysis Using a HPLC-Chip Designed for Automated Glycopeptide Enrichment

Automated enrichment of glycopeptides from the NISTmAb RM was accomplished using an HPLC-chip similar to that described for peptide mapping, except that the enrichment column was replaced with one containing 5 μm HILIC particles, as shown in Figure 6. This chip offers the advantage that glycopeptides can be analyzed without ion-suppression from non-glycosylated peptides, which is known to occur in analyses of tryptic digests (*87*). Glycopeptide enrichment serves to increase the signal-to-noise ratio for glycopeptide ions and permits identification of more glycoforms than might be possible in experiments without enrichment. Additionally, the quality of MS/MS is greatly improved as one can obtain a stronger glycopeptide signal without overloading the chromatographic phase with non-glycosylated peptides (*86*).

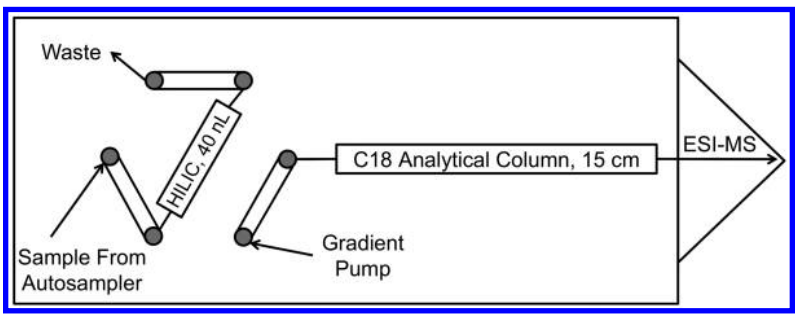

Figure 6. Schematic of an HPLC-chip designed for automated glycopeptide enrichment from tryptic digests of glycoproteins. The chip is similar to that shown in Figure 2 except that the enrichment column is packed with HILIC particles. Using this architecture, glycopeptides were enriched from the NISTmAb RM by the HILIC phase which was equilibrated in 80% ACN containing 0.1% TFA (84). Enriched glycopeptides were eluted from the HILIC enrichment column and focused on the C18 analytical column when the Chip Cube rotor was switched (not shown).

Many approaches for enriching glycopeptides have been demonstrated, and typically make use of hydrophilic solid-phase extraction resins operated in aqueous normal-phase mode. While highly effective, these approaches are often laborious and require extensive manual pipetting. These steps can of course be automated by the use of sample preparation robots, but investment in such solutions can often only be justified in cases where the number of samples that need to be analyzed is very high. The HPLC-chip described in Figure 6 represents a practical compromise, in that no manual manipulation of the sample (beyond diluting the tryptic digest with organic phase) is required for successful glycopeptide enrichment. The throughput of the LC/MS experiment is lowered only slightly as a result of the time needed for glycopeptide enrichment, which is programmed at the beginning of each run (3 minutes per run in these experiments).

For glycopeptide analysis, the previously described tryptic digest of the NISTmAb RM was diluted to a final proportion of 80% ACN and injected

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

onto the HILIC enrichment column of the chip in a running buffer consisting of 80% ACN, 0.1% TFA. These conditions promote non-biased enrichment of glycopeptides. At the same time, non-glycosylated peptides are flushed to waste. After enrichment, the inner rotor of the Chip Cube was switched. As a function of the ratio of enrichment column to separation column volume, as well as the initial buffer composition for separation (97% H_2O, 0.1% FA), glycopeptides are desorbed from the HILIC enrichment column and focused onto the C18 separation column.

Quantification of the glycopeptides enriched from the tryptic digest of the NISTmAb RM using the glycopeptide enrichment HPLC-chip is shown in Figure 7. The data indicate that the RM is modified predominantly by H3N4F1 (G0F), H4N4F1 (G1F), and H5N4F1 (G2F) glycoforms in a ratio typical of many mAbs. Structures which contained additional galactose residues, such as H6N4F1 and H7N4F1, were also present. It is possible that these glycans contain Galα1,3Gal linkages, which are potentially immunogenic. However, additional experiments, for example treatment of the structures using exoglycosidases, are required to confirm this. Another possibility is that these compositions represent hybrid type *N*-glycans. Low levels of *N*-glycolylneuraminic acid containing glycans, such as Sg1H4N3F1, were detected. However, no glycans containing *N*-acetylneuraminic acid were observed. Tandem MS experiments were performed on the enriched glycopeptides, and an example is given in Figure 8. In each case, the presence of oxonium ions can be used to confirm that the peaks are glycopeptide in nature. Furthermore, monosaccharide losses from the glycopeptide precursor ions were consistent with the glycan component of the assigned glycopeptide compositions.

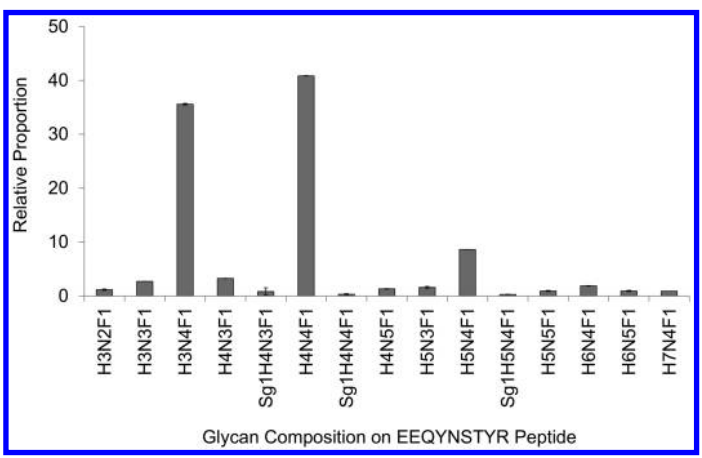

Figure 7. Histogram of N-glycopeptide quantification from the glycopeptide enrichment HPLC-chip analysis of the NISTmAb RM. The masses of the glycopeptides detected were consistent with tryptic peptide sequence EEQYNSTYR. Measurements were made in triplicate and error bars are expressed as standard deviation of the measurement. Assignment of glycopeptide composition was made using accurate mass (<5 ppm). MS/MS spectra were consistent with the given compositions.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

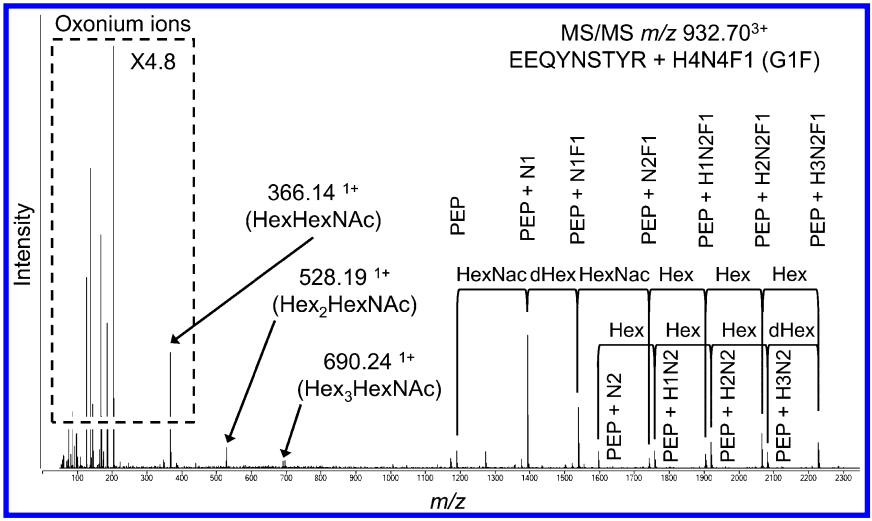

Figure 8. Example of a CID MS/MS spectrum from m/z 932.70^{3+}, which corresponds (with a mass error of 3.1 ppm) to a glycopeptide comprised of the peptide EEQYNSTYR modified by a glycan with the composition H4N4F1 (G1F glycan). Product ions from the glycopeptide precursor were consistent with the composition of the assigned glycan.

N-Glycan Analysis Using mAb-Glyco Chip

The mAb-Glyco Chip is a microfluidic device that integrates the components necessary for release and separation of *N*-glycans from mAbs on-line with MS detection. An example of such a chip is shown in Figure 9A. In Figure 9B, an enlarged view of the area around the rotor interface is shown. Pictured in the diagram are the enzyme reactor, PGC enrichment column, and PGC separation column as well as their arrangement on the inner (6-port) and outer (10-port) rotor. Both of the rotors are automatically turned to control all steps of sample preparation and separation. An optional make-up flow channel is also pictured in the figure, which is useful for work in the negative-ion mode (*58*) or for post-column infusion of chemical reagents. Separation and quantification of *N*-glycans on PGC with MS detection is an established method of antibody glycan characterization (*88*).

The mAb-Glyco Chip workflow steps are depicted in the schematics shown in Figure 10A-D. As shown in Figure 10A, a mAb sample is loaded into the PNGase F enzyme reactor via an autosampler. Once the enzyme reactor is full, the outer rotor switches, and sequesters a ~150 nL volume of mAb solution for a user-defined deglycosylation period (4 min in these experiments) (Figure 10B). Once complete, both the inner and outer rotors are switched, as shown in Figure 10C. The sample loading flow then pushes the released glycans out of the enzyme reactor and across the PGC enrichment column. After sufficient desalting, the inner rotor is switched (Figure 10D) and a separation gradient is initiated. Released glycans are then detected by the MS system.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

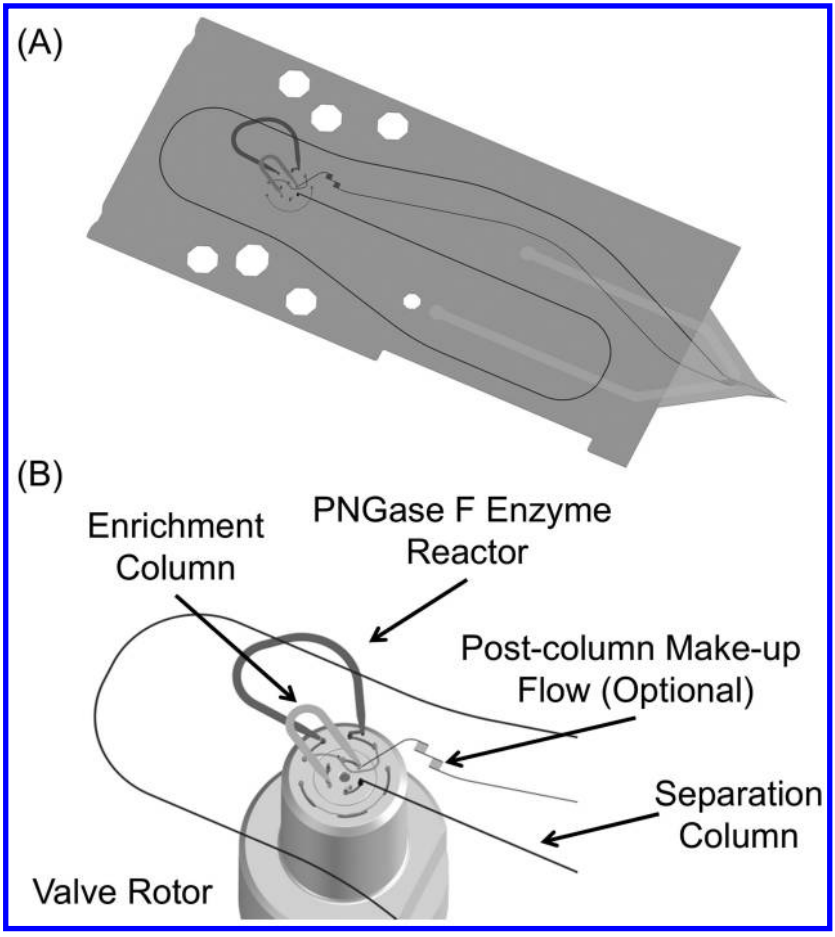

Figure 9. mAb-Glyco Chip for automated preparation and analysis of N-glycans from mAbs. (A) Drawing of a mAb-Glyco Chip. The PNGase F enzyme reactor, PGC enrichment column, and PGC separation column are pictured. (B) Enlarged view of mAb-Glyco Chip flow path connections to the inner 6-port and outer 10-port rotor. This design permits timed incubation (via heart cutting) of samples in the enzyme reactor.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

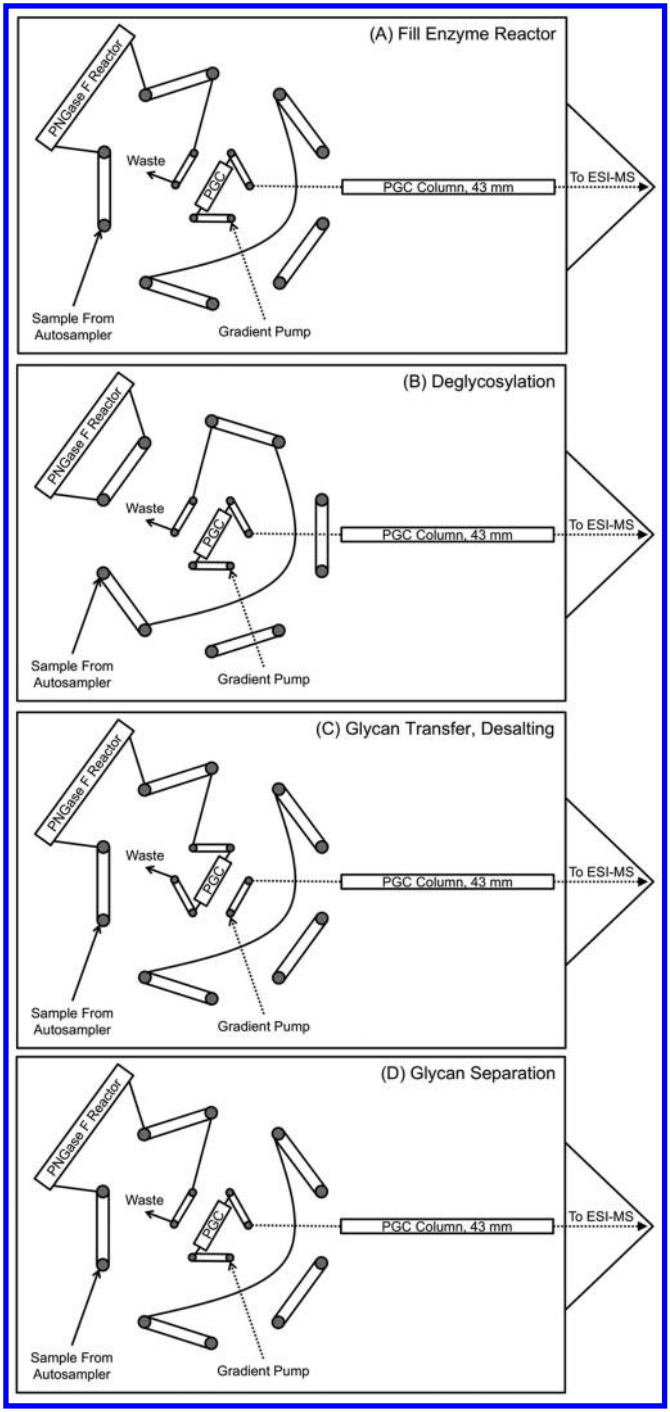

Figure 10. Schematic of mAb-Glyco Chip. The figure depicts all stages of mAb-Glyco Chip operation including filling the enzyme reactor (A), timed

deglycosylation (B), transfer of released N-glycans to the PGC enrichment column (C), and separation of N-glycans on the PGC separation column (D). All steps are coordinated by switching the inner 6-port and outer 10-port rotors of the Chip Cube, which are concentrically oriented with respect to one another.

For analysis using the mAb-Glyco Chip, the NISTmAb RM was diluted to a concentration of 0.25 mg/mL, which equates to about 37.5 ng of mAb that was subject to deglycosylation. A representative ion chromatogram is shown in Figure 11. PNGase F releases *N*-glycans from core proteins in their glycosylamine form, and due to the high speed of analysis using this workflow, these are the primary ions detected by the MS instrument. However, a small portion of the glycans in their free reducing end form is also observed, and these components are generated spontaneously by hydrolysis at the low pH of the separation conditions (0.1% FA).

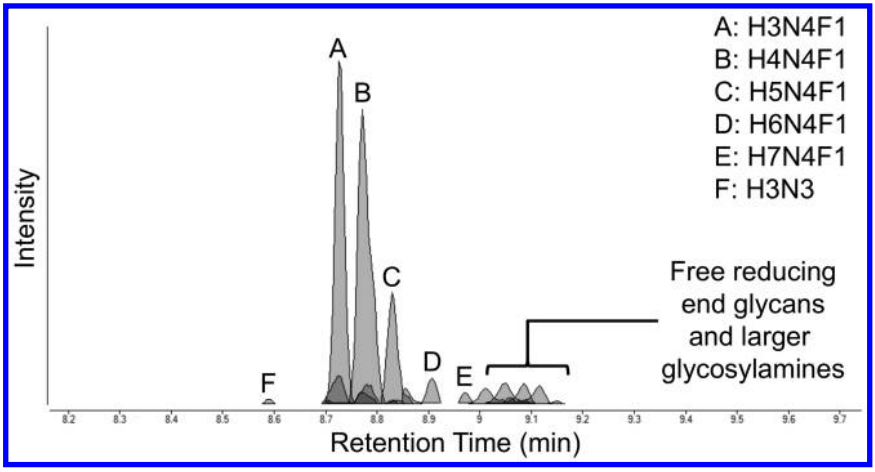

Figure 11. Chromatogram of N-glycans detected from mAb-Glyco Chip analysis of the NISTmAb RM. The major glycoforms are annotated in the chromatogram and chiefly appear in their glycosylamine form. Free reducing end species with later elution times were detected in low abundance.

A histogram of the abundance of the glycans detected from the RM is shown in Figure 12. The abundances shown are based on triplicate analysis, and include signal from all of the charge states and reducing end forms (glycosylamine and free reducing end) of the glycans detected. A total of 24 glycan compositions were detected. Consistent with the glycopeptide dataset, H3N4F1 (G0F), H4N4F1 (G1F), and H5N4F1 (G2F) were observed as the primary glycan modifications of the RM. Also observed were the glycans containing additional hexose residues (H6N4F1 and H7N4F1). It is important to note that many of the measured compositions are comprised of a mixture of structural isomers which are not resolved by the fast chromatographic run used for the separation. However, very efficient separation of glycan isomers can be achieved using PGC separation (*50*) simply by running a longer gradient. When longer gradients are run using the mAb-glyco Chip, the majority of the glycan signal will originate from free

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

reducing end structures (rather than glycosylamines), as there would be sufficient time for glycosylamine hydrolysis during a highly resolving gradient.

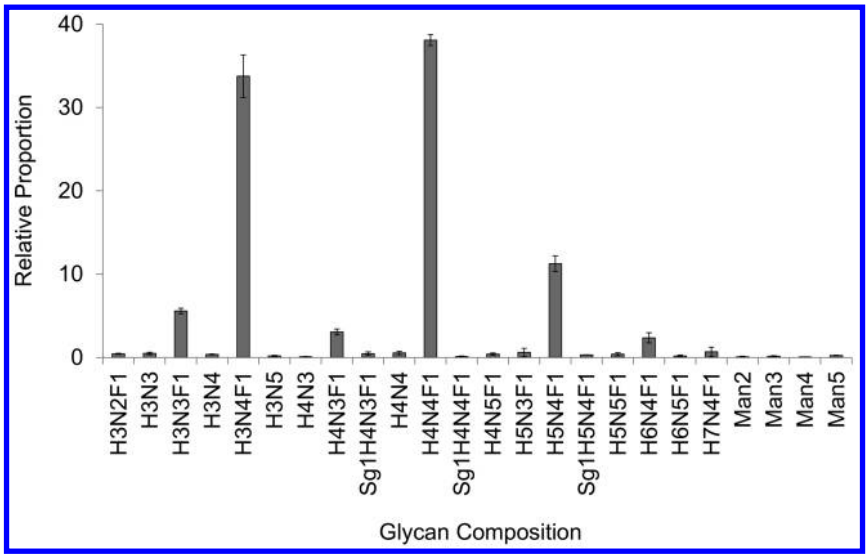

Figure 12. Histogram of N-glycan quantification from mAb-Glyco Chip analysis of the NISTmAb RM. Measurements were made in triplicate and error bars are expressed as the standard deviation of the measurement. Assignment of glycan composition was made using accurate mass (<5 ppm).

MS Analysis of Intact and Deglycosylated mAb Using a Modified mAb-Glyco Chip

A variant of the mAb-Glyco Chip which is designed to capture *N*-deglycosylated mAbs was used to determine the intact mass of the NISTmAb RM. This chip, depicted in Figure 13, has the same architecture as that described for *N*-glycan analysis (Figure 9), except that the PGC enrichment and separation columns are replaced with columns packed with 5 µm, 1000 Å PLRP-S, an inherently hydrophobic polymeric material useful for reversed-phase separations. The RM was diluted to a concentration of 0.125 mg/mL, 2 µL of which was injected onto the HPLC-chip, equating to about 19 ng of mAb subject to deglycosylation and subsequent MS analysis. The resultant, deconvoluted mass spectrum is shown in Figure 14B. The major peak observed had a mass of 145152.7 Da. This is as compared to the calculated theoretical mass of 145148.5 Da for the deglycosylated molecule. The calculated average mass value was determined using atomic weights of elements from organic materials listed in (*89*) and as described in (*90*). Directly before these measurements, the TOF MS used in this study was tuned and achieved a mass resolution of 10,600 at *m/z* 922.01, which is within the expected range of resolution for the instrument.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

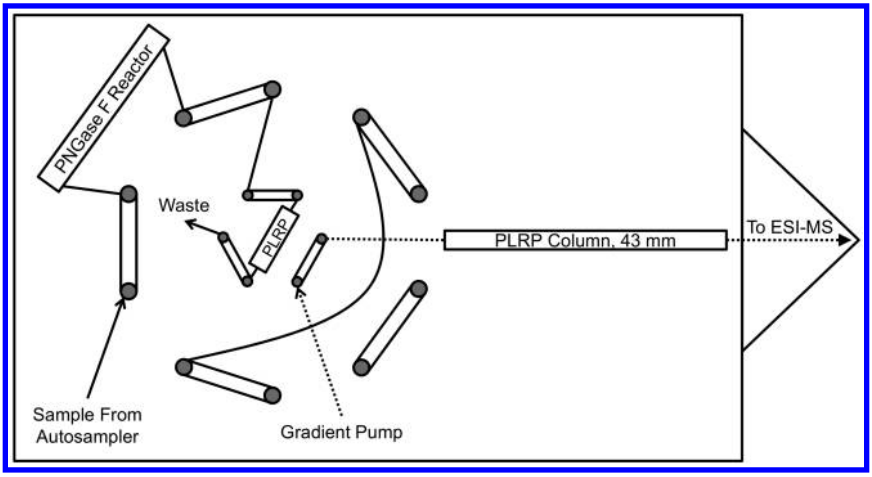

Figure 13. Schematic of a modified intact protein mAb-Glyco Chip designed for analysis of intact, deglycosylated mAb. The operation of this device is similar to that depicted in Figure 10 except that the enrichment and separation columns are packed with PLRP-S rather than PGC.

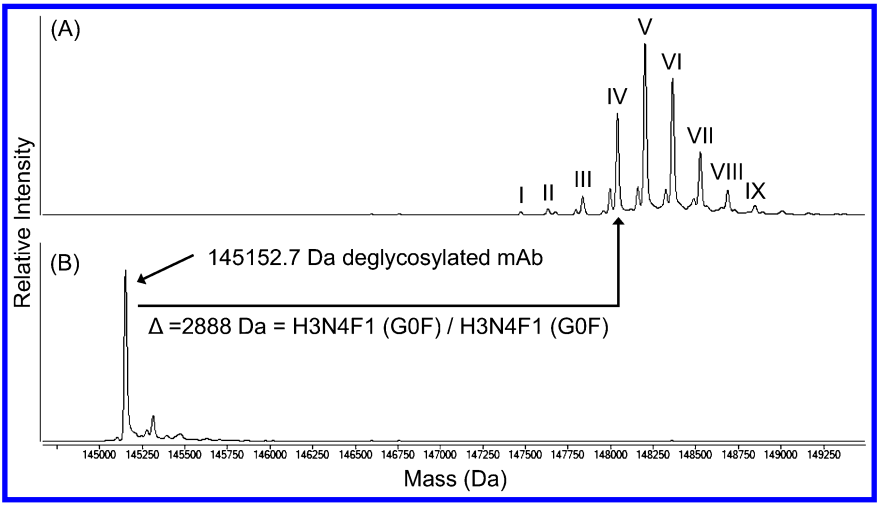

Figure 14. Deconvoluted mass spectra produced from analysis of the NISTmAb RM using the modified mAb-Glyco Chip depicted in Figure 13. (A) Deconvoluted MS of intact NISTmAb RM. The data were produced by bypassing the PNGase F reactor of the chip. Major peaks are labeled in the figure and peak assignments are listed in Table 2. (B) Deconvoluted MS of intact, deglycosylated NISTmAb RM. Automatic PNGase F deglycosylation was coordinated by the chip.

The same HPLC-chip was used to analyze the intact mAb with no PNGase F treatment by programmatically bypassing the enzyme reactor. For the sake of comparison, the RM was further diluted in order to maintain an equivalent amount of protein on column (~20 ng per injection). The intact mass spectrum of the

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

NISTmAb RM is shown in Figure 14A. A variety of glycoforms were detected, the most abundant (peak V) corresponding to a mAb modified by H3N4F1 (G0F) and H4N4F1 (G1F) glycoforms. A peak corresponding to a mAb modified with two H3N4F1 (G0F) glycoforms was observed at 148040.8 Da. The calculated mass for this species is 148037.2 Da. Signals resultant from other combinations of glycan pairs were also observed, corresponding well with the released glycan data from Figure 12 and the glycopeptide data from Figure 7. The major glycoforms are annotated in the figure and the peak assignments are listed in Table 2. A number of lower abundance peaks were also detected (the smaller peaks to the left of those labeled in the figure). The masses of these peaks were on average 41 Da less than the major glycoforms, corresponding to the mass difference between HexNac and Hex. These peaks can thus be assigned to additional glycoforms that would be consistent with the released glycan and glycopeptide data. However, it should be noted that these peaks would overlap with those produced from lysine variants. For example, the peak directly to the left of peak IV would be assigned to H3N4F1 (G0F) / H4N3F1, which would overlap with a variant of peak III which contained Lys.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Table 2. Assignments and Comparison of Observed and Calculated Mass Values for the Peaks in Figure 14A

Peak	*Assignment*	*Obs. Mass (Da)*	*Calc. Mass (Da)*
I	H3N3F1 / H2N3F1	147472.6	147468.6
II	H3N3F1 / H3N3F1	147633.3	147630.8
III	H3N4F1 (G0F) / H3N3F1	147837.3	147834.0
IV	H3N4F1 (G0F) / H3N4F1 (G0F)	148040.8	148037.2
V	H3N4F1 (G0F) / H4N4F1 (G1F)	148202.9	148199.3
VI	H4N4F1 (G1F) / H4N4F1 (G1F)	148365.0	148361.4
VII	H4N4F1 (G1F) / H5N4F1 (G2F)	148526.7	148523.6
VIII	H5N4F1 (G2F) / H5N4F1 (G2F)	148688.1	148685.7

MS Analysis of IdeS-Generated mAb Fragments Using a Modified mAb-Glyco Chip

The modified mAb-glyco Chip described in Figure 13 was also used to analyze fragments of the NISTmAb RM generated by treatment with IdeS, a cysteine protease that cleaves IgG specifically in the hinge region (*91*). Two experiments were performed, similar to those previously described for the intact RM, which included measurement of mAb fragments with and without automated treatment with PNGase F. Before sample injection, IdeS digested RM was diluted to a concentration of 0.0625 mg/mL.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Two well-resolved chromatographic peaks (data not shown) were detected from the IdeS preparation, corresponding to Fc/2 and $(Fab')_2$ subunits. Mass spectra generated from the summed signal under these chromatographic peaks were deconvoluted and the results are shown in Figure 15. The mass spectra of the Fc/2 subunit before and after deglycosylation are shown in parts A and C of the figure, respectively. A number of glycoforms of the Fc/2 fragment were detected, and the peak assignments are summarized in Table 3. Release of the *N*-glycans resulted in a comparatively simple mass spectrum, where the deglycosylated mass of the Fc/2 fragment was observed at 23788.3 Da. The calculated mass for this species was 23787.7 Da. In addition, peaks at +128.1 and +161.8 were observed and were assigned to Fc/2 fragments with lysine or glycation, respectively. It is interesting to note that the lysine variants appear to contribute a very small amount of overall signal to the spectrum of the deglycosylated Fc/2 fragments. This lends additional support to the assignment of the minor peaks in Figure 14A that differ by -41 Da from the major peaks.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

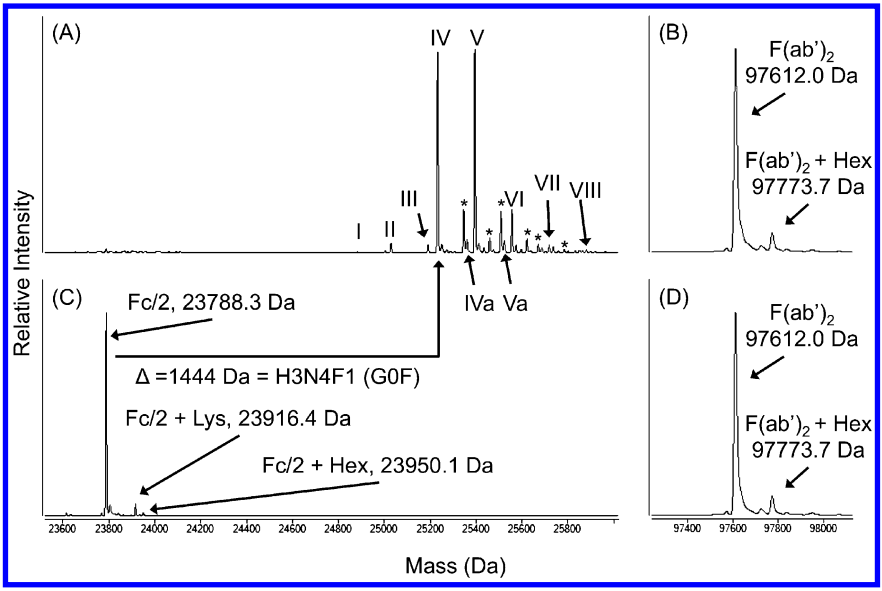

Figure 15. Deconvoluted mass spectra produced from the analysis of IdeS-digested NISTmAb RM using the modified mAb-Glyco Chip depicted in Figure 13. (A) Deconvoluted MS of Fc/2 of NISTmAb RM before deglycosylation. The data were produced by bypassing the PNGase F reactor of the chip. Major peaks are labeled in the figure and peak assignments are listed in Table 3. (B) Deconvoluted MS of F(ab')$_2$ of NISTmAb RM before deglycosylation. (C) Deconvoluted MS of Fc/2 of NISTmAb RM after deglycosylation. (D) Deconvoluted MS of F(ab')$_2$ of NISTmAb RM after deglycosylation.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 3. Assignments and Comparison of Observed and Calculated Mass Values for the Peaks in Figure 15A

Peak	*Assignment*	*Obs. Mass (Da)*	*Calc. Mass (Da)*
I	H3N3	24883.1	24882.7
II	H3N3F1	25029.3	25028.9
III	H4N3F1	25191.3	25191.0
IV	H3N4F1 (G0F)	25232.6	25232.1
IVa	H3N4F1 (G0F) + Lys	25361.1	25360.2
V	H4N4F1 (G1F)	25394.8	25394.2
Va	H4N4F1 (G1F) + Lys	25523.3	25522.4
VI	H5N4F1 (G2F)	25556.6	25556.4
VII	H6N4F1 (G3F)	25718.6	25718.5
VIII	H7N4F1 (G4F)	25880.4	25880.6
*	TFA Adduct	n.a.	n.a.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

The mass spectra of the $(Fab')_2$ subunit before and after deglycosylation are shown in parts B and D of Figure 15, respectively. Owing to the fact that this portion of the molecule is not *N*-glycosylated, no change was observed in the two spectra. The mass of the $(Fab')_2$ subunit was observed at 97612.0 Da, whereas the calculated mass for this fragment is 97609.0 Da. A peak at +161.7 from the main $(Fab')_2$ subunit was observed and is likely the result of glycation on this portion of the molecule.

Preparation of mAbs using IdeS has gained tremendous popularity because of the ease of performing the digestion and the fact that cleavage in the hinge region is fast and typically stoichiometric. In addition, one gains additional spatial information and resolution of glycoforms and other modifications when analyzing mAb subunits as opposed to the intact molecule. In the case of the NISTmAb RM, we have determined that glycation modifications are likely present on both the $(Fab')_2$ and Fc/2 subunits, for example. As demonstrated, initial preparation of the NISTmAb using IdeS is readily compatible (without any sample clean-up) with the described HPLC-chip for automatic deglycosylation and intact protein mass analysis. Although such a chip-based workflow has not been created to date, one can envision a device capable of intact mass analysis of mAbs with and without both IdeS digestion and PNGase F release of *N*-glycans.

As demonstrated, the mAb-Glyco Chip, and the described variant thereof, were very useful for determining the composition of released glycans, a well as the

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

masses of the intact molecule or its subunits before and after *N*-deglycosylation. In the context of these experiments, the microfluidic technology saved significant time. Limited manual sample handling steps were performed. Additional speed was realized from the accelerated glycan release step, which was 4 minutes per sample replicate. This is in contrast to typical in-solution release of glycans from mAbs using PNGase F, which is often performed overnight or at the minimum for a few hours.

mAb Sizing and QC Using the 2100 Bioanalyzer

The NISTmAb RM was analyzed using an Agilent 2100 Bioanalyzer, a system capable of microfluidic electrophoresis of DNA, RNA, and proteins. We chose to analyze the mAb RM using two kits, the protein 80 kit and the HSP-250 kit. These kits differ in a number of ways, including the size ranges they are capable of measuring (5-80 kDa for Protein 80 and 10-250 kDa for HSP-250). However, they also differ in that protein staining is performed on-chip (Protein 80 assay) or off-chip (HSP-250 assay). In the latter case, additional sample preparation steps are required. However, sensitivity is increased (to the pg / μL level) since background levels of fluorescence are eliminated. This level of sensitivity allows one to utilize this workflow to assess very low level protein impurities which may be part of the mAb preparation. Although not used in this study, an additional kit capable of measuring proteins up to 230 kDa and which utilizes on-chip staining is available. This kit allows measurement of intact mAbs without any additional sample handling steps.

For analysis using the Protein 80 kit, the mAb RM sample was reduced using DTT in order to break the disulfide bonds connecting the light chains and heavy chains. The data generated from this analysis (and any analysis made using the Bioanalyzer) take the form of an electropherogram, similar to one that might be generated by capillary electrophoresis. However, these data can be represented as a gel-like image, which is useful for those who wish to view the information as they would a traditional SDS-PAGE separation. Such an image is shown in Figure 16, which shows 5 replicates (lanes 1-5) of the reduced NISTmAb RM. Bands for the heavy chain and light chain are present, in addition to upper and lower marker bands at 95 and 1.6 kDa, respectively. For size comparison, a set of protein markers was run in the lane labeled L. Figure 17 shows the corresponding electropherogram of the reduced mAb sample. Based on the migration time, the light chain was measured at 26.8 kDa and the heavy chain was measured at 57.1 kDa.

A separate analysis was carried out on the NISTmAb RM using the HSP-250 kit. For the analysis using the HSP-250 kit, mAb samples were analyzed without prior DTT treatment in order to measure the intact molecule. An electropherogram of the HSP-250 analysis of the intact NISTmAb RM is shown in Figure 18. Based on these results, the molecular weight of the mAb was determined to be 149 kDa. A small proportion of non-glycosylated mAb was also detected, electrophoresing directly before the intact protein.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

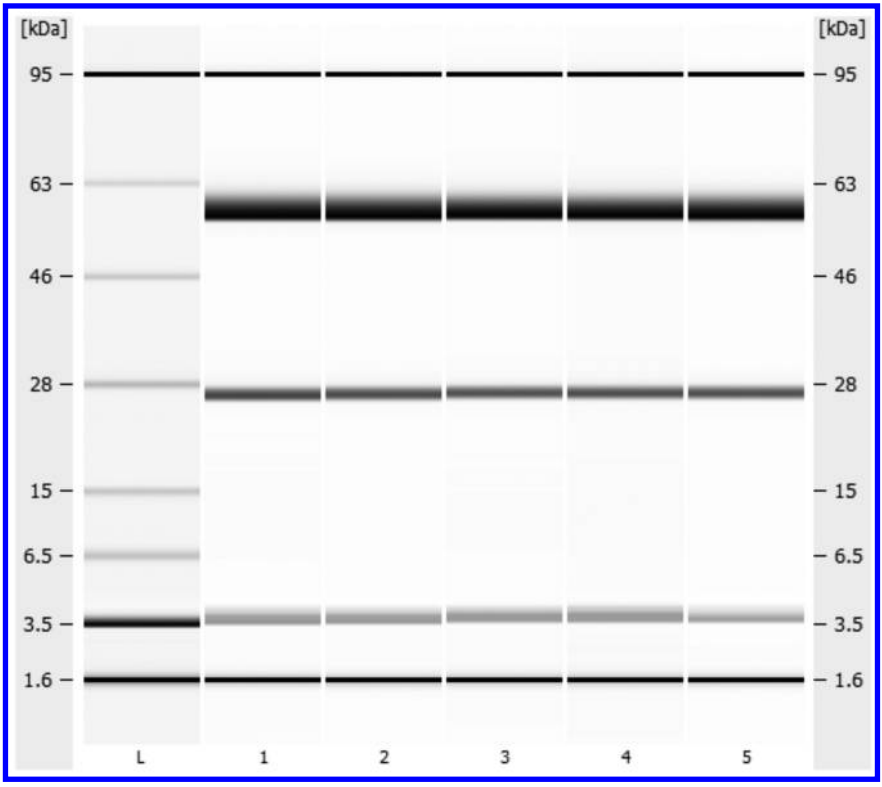

Figure 16. Gel-like image of 2100 Bioanalyzer analysis of NISTmAb RM (lanes 1-5) using the Protein 80 kit. The mAb was reduced using DTT for this experiment.

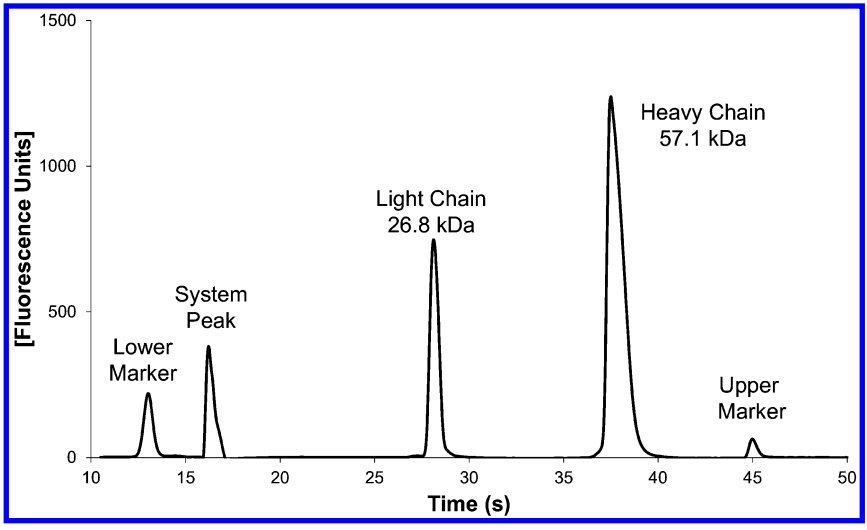

Figure 17. 2100 Bioanalyzer electropherogram of reduced NISTmAb RM using the Protein 80 kit. The mAb was reduced using DTT for this experiment.

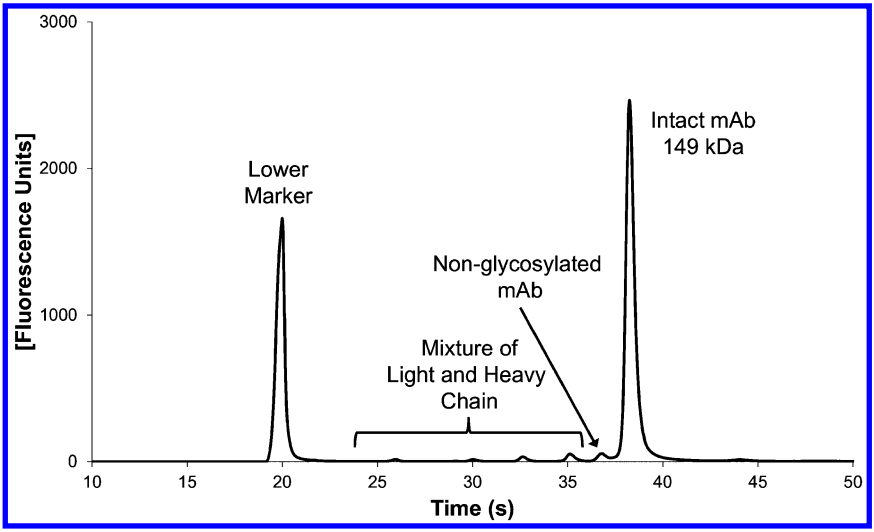

Figure 18. 2100 Bioanalyzer analysis of intact NISTmAb RM using the HSP 250 kit.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Charge Variant Analysis Using Capillary Isoelectric Focusing

cIEF analysis of the NISTmAb RM was performed in order to profile of charge variants present in the preparation. The experiments were performed using a commercial capillary electrophoresis platform in a two-step manner, where the protein isoforms were first focused and then mobilized toward the detector. In addition to cIEF analysis of the untreated mAb RM, samples were analyzed after treatment with carboxypeptidase B. This enzyme specifically cleaves C-terminal basic amino acid residues from peptides and proteins. Thus, it can be used for the determination of lysine variants of mAb preparations. cIEF measurement of the mAb RM was also made following treatment with sialidase A in order to assess the contribution of sialic acid glycoforms to the charge variant profile. This enzyme, which originates from *Arthrobacter ureafaciens*, is capable of cleaving branched and terminal *N*-acetylneuraminic and *N*-glycolylneuraminic acid residues that are α2,3 linked, α2,6 linked, α2,8 linked, or α2,9 linked.

The results of the cIEF analysis are shown in Figure 19. Panel A shows the results for the untreated mAb RM. The main peak is detected at 26 min, with basic variants migrating earlier and acidic variants migrating later. Two main basic variants were detected at migration times of 24 and 25 min, respectively. These peaks were observed to disappear after treatment with carboxypeptidase B, as shown in panel B. Thus, they are likely to represent the lysine variants of the mAb that were detected in the peptide mapping and subunit analysis experiments. It is inferred that the peak at 24 min represents the population of mAb with two unprocessed lysines while the peak at 25 min represents the population of mAb with one unprocessed lysine.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

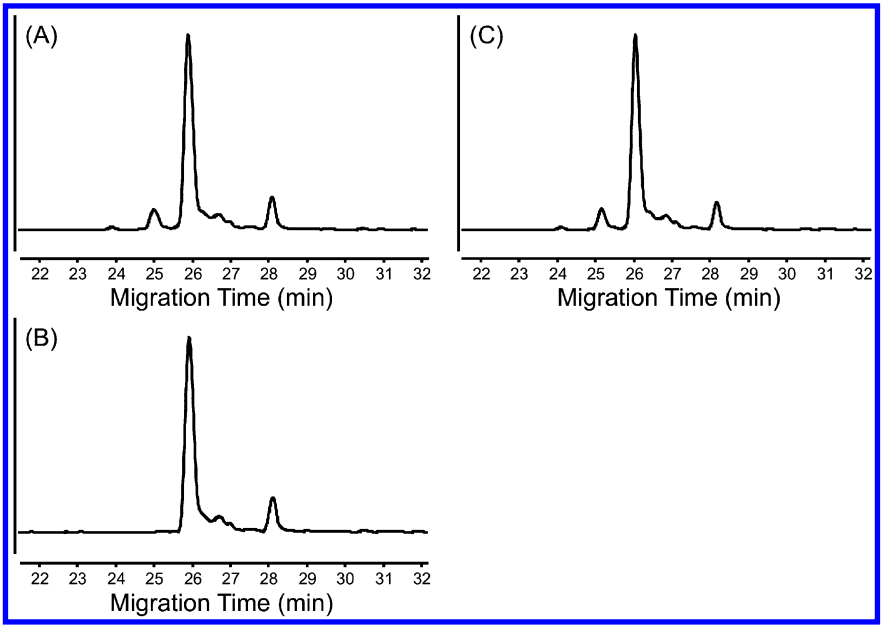

Figure 19. cIEF analysis of NISTmAb RM. (A) mAb RM with no treatment (B) mAb RM after treatment with carboxypeptidase B (C) mAb RM after treatment with Sialidase A.

A number of acidic variants were detected which migrated later than the main peak. To determine whether these peaks were the result of sialylated glycans, the mAb RM was treated with Sialidase A. The resultant electropherogram is shown in panel C. As can be seen, there was no appreciable difference in the results as compared to the untreated sample. This is consistent with the very low amount of *N*-glycolylneuraminic acid modified glycans that were detected by the previously described analysis of released glycans (Figure 12) and glycopeptides (Figure 7). Identification of the acidic variants thus requires additional experiments, which were not undertaken during this study. However, it is likely that some of the acidic variants can be explained by glycation and deamidation, which were detected at the intact protein level and peptide level, respectively. Though cIEF offers the highest available resolution of charge variants (compared to cation exchange chromatography or gel IEF), one disadvantage is that peaks cannot be collected for orthogonal analyses (*e.g.* peptide mapping). This is an inherent property of the extremely small sample quantity used for cIEF separations. One alternative is to utilize preparative scale, high resolution IEF techniques in order to prepare samples for orthogonal analysis. This has been achieved using OFFGEL electrophoresis, for example (*92*). This strategy is advantageous because the fractions prepared for orthogonal analysis are separated using the same principle as cIEF. Each peak in the cIEF profile can then be assigned using orthogonal methods (or even cIEF). Once the charge variant profile is well-known, cIEF can

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

be used as a rapid way of assessing any changes that may occur during production or storage. Thus, cIEF can be used for lot release analysis, where the advantages of reproducibility, speed, and in particular high resolution, can be realized.

Conclusions and Outlook

There is a tremendous investment in the research and production of mAb therapeutics by the pharmaceutical industry. It is of utmost importance to understand the critical quality attributes of these molecules in order to quickly and cost-effectively produce the safest and most efficacious therapeutic agents. Indeed, with the implementation of quality by design, there has never been a greater demand placed on analytical technology, for example in quickly determining how changes in process influence the structure of a biomolecule. To this end, the application of microfluidic technologies has been greatly beneficial in the characterization of mAbs. We have illustrated how microfluidic technologies can provide a broad survey of mAb structure while minimizing manual effort. Quick time to measurement combined with ease of use has popularized instrument platforms which incorporate microfluidic components. It is likely that such instruments will continue to appear in our analytical toolkit for use in the characterization of mAbs, other emerging therapeutic proteins, and throughout the field of bioanalytical science with the end goal of increasing the efficiency of our work.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

References

1. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol.* **2010**, *10*, 345–352.
2. Scott, A. M.; Allison, J. P.; Wolchok, J. D. *Cancer Immun.* **2012**, *12*, 14.
3. Nadler, L. M.; Stashenko, P.; Hardy, R.; Kaplan, W. D.; Button, L. N.; Kufe, D. W.; Antman, K. H.; Schlossman, S. F. *Cancer Res.* **1980**, *40*, 3147–3154.
4. Dotan, E.; Aggarwal, C.; Smith, M. R. *P & T* **2010**, *35*, 148–57.
5. Dillman, R. O.; Beauregard, J. C.; Halpern, S. E.; Clutter, M. *J. Biol. Response Mod.* **1986**, *5*, 73–84.
6. Scott, A. M.; Wolchok, J. D.; Old, L. J. *Nat. Rev. Cancer* **2012**, *12*, 278–287.
7. Li, J.; Zhu, Z. *Acta Pharmacol. Sin.* **2010**, *31*, 1198–1207.
8. Berger, M. S.; Leopold, L. H.; Dowell, J. A.; Korth-Bradley, J. M.; Sherman, M. L. *Invest. New Drugs* **2002**, *20*, 395–406.
9. Chames, P.; Baty, D. *mAbs* **2009**, *1*, 539–547.
10. Carter, P. J. *Exp. Cell Res.* **2011**, *317*, 1261–1269.
11. Nelson, A. L.; Reichert, J. M. *Nat. Biotechnol.* **2009**, *27*, 331–337.
12. Nelson, A. L. *mAbs* **2010**, *2*, 77–83.
13. Dalziel, M.; Crispin, M.; Scanlan, C. N.; Zitzmann, N.; Dwek, R. A. *Science* **2014**, *343*, 1235681.
14. Weiner, L. M.; Surana, R.; Wang, S. *Nat. Rev. Immunol.* **2010**, *10*, 317–327.
15. Carter, P. J. *Nat. Rev. Immunol.* **2006**, *6*, 343–357.

16. Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. *J. Pharm. Sci. Technol.* **2007**, *96*, 1–26.
17. Seidel, U. J.; Schlegel, P.; Lang, P. *Front. Immunol.* **2013**, *4*, 76.
18. Chan, A. C.; Carter, P. J. *Nat. Rev. Immunol.* **2010**, *10*, 301–316.
19. Zeck, A.; Regula, J. T.; Larraillet, V.; Mautz, B.; Popp, O.; Gopfert, U.; Wiegeshoff, F.; Vollertsen, U. E.; Gorr, I. H.; Koll, H.; Papadimitriou, A. *PLoS One* **2012**, *7*, e40328.
20. Liu, H.; Gaza-Bulseco, G.; Faldu, D.; Chumsae, C.; Sun, J. *J. Pharm. Sci. Technol.* **2008**, *97*, 2426–2447.
21. Kanda, Y.; Yamada, T.; Mori, K.; Okazaki, A.; Inoue, M.; Kitajima-Miyama, K.; Kuni-Kamochi, R.; Nakano, R.; Yano, K.; Kakita, S.; Shitara, K.; Satoh, M. *Glycobiology* **2007**, *17*, 104–118.
22. Jefferis, R. *Biotechnol. Progr.* **2005**, *21*, 11–16.
23. Nose, M.; Wigzell, H. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6632–6636.
24. Goetze, A. M.; Liu, Y. D.; Zhang, Z.; Shah, B.; Lee, E.; Bondarenko, P. V.; Flynn, G. C. *Glycobiology* **2011**, *21*, 949–959.
25. Umana, P.; Jean-Mairet, J.; Moudry, R.; Amstutz, H.; Bailey, J. E. *Nat. Biotechnol.* **1999**, *17*, 176–180.
26. Shields, R. L.; Lai, J.; Keck, R.; O'Connell, L. Y.; Hong, K.; Meng, Y. G.; Weikert, S. H.; Presta, L. G. *J. Biol. Chem.* **2002**, *277*, 26733–26740.
27. Restelli, V.; Butler, M. In *Cell Engineering*; Al-Rubeai, M., Ed.; Kluwer Academic Publishers: The Netherlands, 2002; pp 61-92.
28. Sheeley, D. M.; Merrill, B. M.; Taylor, L. C. *Anal. Biochem.* **1997**, *247*, 102–110.
29. Borys, M. C.; Dalal, N. G.; Abu-Absi, N. R.; Khattak, S. F.; Jing, Y.; Xing, Z.; Li, Z. J. *Biotechnol. Bioeng.* **2010**, *105*, 1048–1057.
30. Tangvoranuntakul, P.; Gagneux, P.; Diaz, S.; Bardor, M.; Varki, N.; Varki, A.; Muchmore, E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12045–2050.
31. Ghaderi, D.; Taylor, R. E.; Padler-Karavani, V.; Diaz, S.; Varki, A. *Nat. Biotechnol.* **2010**, *28*, 863–867.
32. Malhotra, R.; Patel, V.; Chikkaveeraiah, B. V.; Munge, B. S.; Cheong, S. C.; Zain, R. B.; Abraham, M. T.; Dey, D. K.; Gutkind, J. S.; Rusling, J. F. *Anal. Chem.* **2012**, *84*, 6249–6255.
33. Karlsson, R.; Fagerstam, L.; Nilshans, H.; Persson, B. *J. Immunol. Methods* **1993**, *166*, 75–84.
34. Karlsson, R.; Michaelsson, A.; Mattsson, L. *J. Immunol. Methods* **1991**, *145*, 229–240.
35. Fagerstam, L. G.; Frostell, A.; Karlsson, R.; Kullman, M.; Larsson, A.; Malmqvist, M.; Butt, H. *J. Mol. Recognit.* **1990**, *3*, 208–214.
36. Yan, B. X.; Steen, S.; Hambly, D.; Valliere-Douglass, J.; Bos, T. V.; Smallwood, S.; Yates, Z.; Arroll, T.; Han, Y. H.; Gadgil, H.; Latypov, R. F.; Wallace, A.; Lim, A.; Kleemann, G. R.; Wang, W. C.; Balland, A. *J. Pharm. Sci.* **2009**, *98*, 3509–3521.
37. Swanson, S. J. *Dev. Biol. (Basel)* **2003**, *112*, 127–133.
38. Lofgren, J. A.; Dhandapani, S.; Pennucci, J. J.; Abbott, C. M.; Mytych, D. T.; Kaliyaperumal, A.; Swanson, S. J.; Mullenix, M. C. *J. Immunol.* **2007**, *178*, 7467–7472.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

39. Wozniak-Knopp, G.; Bartl, S.; Bauer, A.; Mostageer, M.; Woisetschlager, M.; Antes, B.; Ettl, K.; Kainer, M.; Weberhofer, G.; Wiederkum, S.; Himmler, G.; Mudde, G. C.; Ruker, F. *Protein Eng. Des. Sel.* **2010**, *23*, 289–297.
40. Jez, J.; Antes, B.; Castilho, A.; Kainer, M.; Wiederkum, S.; Grass, J.; Ruker, F.; Woisetschlager, M.; Steinkellner, H. *J. Biol. Chem.* **2012**, *287*, 24313–24319.
41. Canziani, G. A.; Klakamp, S.; Myszka, D. G. *Anal. Biochem.* **2004**, *325*, 301–307.
42. Safsten, P.; Klakamp, S. L.; Drake, A. W.; Karlsson, R.; Myszka, D. G. *Anal. Biochem.* **2006**, *353*, 181–190.
43. Yin, H.; Killeen, K.; Brennen, R.; Sobek, D.; Werlich, M.; van de Goor, T. *Anal. Chem.* **2005**, *77*, 527–533.
44. Franciosi, L.; Govorukhina, N.; Fusetti, F.; Poolman, B.; Lodewijk, M. E.; Timens, W.; Postma, D.; ten Hacken, N.; Bischoff, R. *Electrophoresis* **2013**, *34*, 2683–2694.
45. Rosenling, T.; Stoop, M. P.; Smolinska, A.; Muilwijk, B.; Coulier, L.; Shi, S.; Dane, A.; Christin, C.; Suits, F.; Horvatovich, P. L.; Wijmenga, S. S.; Buydens, L. M.; Vreeken, R.; Hankemeier, T.; van Gool, A. J.; Luider, T. M.; Bischoff, R. *Clin. Chem.* **2011**, *57*, 1703–711.
46. Mohammed, S.; Kraiczek, K.; Pinkse, M. W.; Lemeer, S.; Benschop, J. J.; Heck, A. J. *J. Proteome Res.* **2008**, *7*, 1565–71.
47. Raijmakers, R.; Kraiczek, K.; de Jong, A. P.; Mohammed, S.; Heck, A. J. *Anal. Chem.* **2010**, *82*, 824–832.
48. Hua, S.; Lebrilla, C.; An, H. J. *Bioanalysis* **2011**, *3*, 2573–2585.
49. Ruhaak, L. R.; Taylor, S. L.; Miyamoto, S.; Kelly, K.; Leiserowitz, G. S.; Gandara, D.; Lebrilla, C. B.; Kim, K. *Anal. Bioanal. Chem.* **2013**, *405*, 4953–4958.
50. Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B. *Proteomics* **2009**, *9*, 1939–1951.
51. Hua, S.; Jeong, H. N.; Dimapasoc, L. M.; Kang, I.; Han, C.; Choi, J. S.; Lebrilla, C. B.; An, H. J. *Anal. Chem.* **2013**, *85*, 4636–4643.
52. Barboza, M.; Pinzon, J.; Wickramasinghe, S.; Froehlich, J. W.; Moeller, I.; Smilowitz, J. T.; Ruhaak, L. R.; Huang, J.; Lonnerdal, B.; German, J. B.; Medrano, J. F.; Weimer, B. C.; Lebrilla, C. B. *Mol. Cell. Proteomics* **2012**, *11*, M111 015248.
53. Zivkovic, A. M.; German, J. B.; Lebrilla, C. B.; Mills, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4653–4658.
54. Totten, S. M.; Zivkovic, A. M.; Wu, S.; Ngyuen, U.; Freeman, S. L.; Ruhaak, L. R.; Darboe, M. K.; German, J. B.; Prentice, A. M.; Lebrilla, C. B. *J. Proteome Res.* **2012**, *11*, 6124–6133.
55. Ni, W.; Bones, J.; Karger, B. L. *Anal. Chem.* **2013**, *85*, 3127–3135.
56. Brennen, R. A.; Yin, H.; Killeen, K. P. *Anal. Chem.* **2007**, *79*, 9302–9309.
57. Staples, G. O.; Bowman, M. J.; Costello, C. E.; Hitchcock, A. M.; Lau, J. M.; Leymarie, N.; Miller, C.; Naimy, H.; Shi, X.; Zaia, J. *Proteomics* **2009**, *9*, 686–695.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

58. Staples, G. O.; Naimy, H.; Yin, H.; Kileen, K.; Kraiczek, K.; Costello, C. E.; Zaia, J. *Anal. Chem.* **2010**, *82*, 516–522.
59. Staples, G. O.; Shi, X.; Zaia, J. *J. Biol. Chem.* **2010**, *285*, 18336–18343.
60. Huang, Y.; Shi, X.; Yu, X.; Leymarie, N.; Staples, G. O.; Yin, H.; Killeen, K.; Zaia, J. *Anal. Chem.* **2011**, *83*, 8222–8229.
61. Bynum, M. A.; Yin, H.; Felts, K.; Lee, Y. M.; Monell, C. R.; Killeen, K. *Anal. Chem.* **2009**, *81*, 8818–8825.
62. Wang, B.; Gucinski, A. C.; Keire, D. A.; Buhse, L. F.; Boyne, M. T., 2nd *Analyst* **2013**, *138*, 3058–3065.
63. Bousse, L.; Mouradian, S.; Minalla, A.; Yee, H.; Williams, K.; Dubrow, R. *Anal. Chem.* **2001**, *73*, 1207–1212.
64. Ohashi, R.; Otero, J. M.; Chwistek, A.; Hamel, J. F. *Electrophoresis* **2002**, *23*, 3623–3629.
65. Chen, X.; Tang, K.; Lee, M.; Flynn, G. C. *Electrophoresis* **2008**, *29*, 4993–5002.
66. Wong, R. B.; Zeng, M.; Lee, A.-H.; Raju, T. S.; Cheng, K.-C. *Open Pharmacol. J.* **2012**, *6*.
67. Taylor, F. R.; Prentice, H. L.; Garber, E. A.; Fajardo, H. A.; Vasilyeva, E.; Blake Pepinsky, R. *Anal. Biochem.* **2006**, *353*, 204–208.
68. Forrer, K.; Hammer, S.; Helk, B. *Anal. Biochem.* **2004**, *334*, 81–88.
69. Vasilyeva, E.; Woodard, J.; Taylor, F. R.; Kretschmer, M.; Fajardo, H.; Lyubarskaya, Y.; Kobayashi, K.; Dingley, A.; Mhatre, R. *Electrophoresis* **2004**, *25*, 3890–3896.
70. Chen, X. N.; Nguyen, M.; Jacobson, F.; Ouyang, J. *mAbs* **2009**, *1*, 563–571.
71. Silvertand, L. H.; Torano, J. S.; van Bennekom, W. P.; de Jong, G. J. *J. Chromatogr. A* **2008**, *1204*, 157–170.
72. Shimura, K. *Electrophoresis* **2009**, *30*, 11–28.
73. Kilar, F. *Electrophoresis* **2003**, *24*, 3908–3916.
74. He, X. Z.; Que, A. H.; Mo, J. J. *Electrophoresis* **2009**, *30*, 714–722.
75. Maeda, E.; Urakami, K.; Shimura, K.; Kinoshita, M.; Kakehi, K. *J. Chromatogr. A* **2010**, *1217*, 7164–7171.
76. Lin, J.; Lazar, A. C. *Methods Mol. Biol.* **2013**, *1045*, 295–302.
77. Cao, J.; Sun, W.; Gong, F.; Liu, W. *Electrophoresis* **2014**, *35*, 1461–1468.
78. Li, N.; Kessler, K.; Bass, L.; Zeng, D. *J. Pharm. Biomed. Anal.* **2007**, *43*, 963–972.
79. Zhu, G.; Sun, L.; Wojcik, R.; Kernaghan, D.; McGivney, J. B. t.; Dovichi, N. J. *Talanta* **2012**, *98*, 253–256.
80. Kinoshita, M.; Nakatsuji, Y.; Suzuki, S.; Hayakawa, T.; Kakehi, K. *J. Chromatogr. A* **2013**, *1309*, 76–83.
81. Sommer, G. J.; Hatch, A. V. *Electrophoresis* **2009**, *30*, 742–757.
82. Ren, D.; Pipes, G. D.; Liu, D.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. *Anal. Biochem.* **2009**, *392*, 12–21.
83. Wenger, C. D.; Coon, J. J. *J. Proteome Res.* **2013**, *12*, 1377–1386.
84. Mysling, S.; Palmisano, G.; Hojrup, P.; Thaysen-Andersen, M. *Anal. Chem.* **2010**, *82*, 5598–5609.
85. Rasmussen, J. R.; Davis, J.; Risley, J. M.; Van Etten, R. L. *J. Am. Chem. Soc.* **1992**, *114*, 1124–1126.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

86. Khatri, K.; Staples, G. O.; Leymarie, N.; Leon, D. R.; Turiak, L.; Huang, Y.; Yip, S.; Hu, H.; Heckendorf, C. F.; Zaia, J. *J. Proteome Res.* **2014**.
87. Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 161–170.
88. Stadlmann, J.; Pabst, M.; Kolarich, D.; Kunert, R.; Altmann, F. *Proteomics* **2008**, *8*, 2858–2871.
89. Coplen, T. B.; Bohlke, J. K.; De Bievre, P.; Ding, T.; Holden, N. E.; Hopple, J. A.; Krouse, H. R.; Lamberty, A.; Peiser, H. S.; Revesz, K.; Rieder, S. E.; Rosman, K. J. R.; Roth, E.; Taylor, P. D. P.; Vocke, R. D.; Xiao, Y. K. *Pure Appl. Chem.* **2002**, *74*, 1987–2017.
90. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
91. Chevreux, G.; Tilly, N.; Bihoreau, N. *Anal. Biochem.* **2011**, *415*, 212–214.
92. Meert, C. D.; Brady, L. J.; Guo, A.; Balland, A. *Anal. Chem.* **2010**, *82*, 3510–3518.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch009

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Chapter 10

Intact Antibody Characterization Using Orbitrap Mass Spectrometry

Zhiqi Hao,[*,1] Terry Zhang,[1] Yue Xuan,[2] Hongxia Wang,[1] Jie Qian,[1] Shanhua Lin,[1] Jenny Chen,[1] David M. Horn,[1] Dayana Argoti,[1] Alain Beck,[3] Sarah Cianférani,[4] Patrick Bennett,[5] Ken Miller,[1] and Alexander Makarov[2]

[1]Thermo Fisher Scientific, San Jose, California 95134, United States
[2]Thermo Fisher Scientific, 28199 Bremen, Germany
[3]Centre d'Immunologie Pierre Fabre (CIPF), 74160 Saint-Julien-En-Genevois, France
[4]Institut Pluridisciplinaire Hubert Curien (IPHC), University of Strasbourg, 67087 Strasbourg, France
[5]PPD Bioanalytical Lab, Richmond, Virginia 23230, United States
[*]E-mail: zqhao11@hotmail.com

Recent innovations and improvement in Orbitrap technology have enabled routine Orbitrap MS-based analysis of intact monoclonal antibody and related products. In this chapter, LC-MS and LC-MS/MS solutions that include sample preparation, Orbitrap instrumentation and data processing are introduced for routine intact antibody mass measurement as well as subunit sequencing and fragment structure analysis. Measurement of antibody molecular mass under both denatured and native conditions are described with recommendations for best practice and key Orbitrap MS parameters. The results demonstrate accurate and reproducible measurement of molecular mass and relative abundance of glycoforms. The intact subunit sequencing and middle-down approaches efficiently dissociate the light chain and the heavy chain, leading to 67% backbone fragmentation for the intact light chain, and 52% and 32% backbone fragmentation for the single chain Fc (scFc) and Fd′ fragments (Fab-domains of heavy chain), respectively. The use of complementary dissociation methods, ETD and HCD, not only improves sequence coverage,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

but also allows the confident identification and localization of sequence modifications including glycosylation. The improved resolution and scan speed in the new generation Orbitrap mass spectrometers enable the practical utility of online LC in conjunction with subunit or middle-down MS/MS in a high-throughput format. The methods presented here can be used routinely in biopharmaceutical applications.

Introduction

Recombinant monoclonal antibodies (mAb) and their derivatives have become the fastest growing class of human therapeutics and drug candidates because of their high selectivity in binding to an antigen, long circulation lifetime, and the ability to invoke immune cell effector responses. There has been a dramatic increase of mAb-based drug development in the past two decades since their introduction into the clinic in the late 1980s. Today, more than 40 products are currently approved and more than 300 mAb-based therapeutics are in clinical trials (*1*). Meanwhile, most of the first generation mAb therapeutics such as rituximab, trastuzumab, infliximab, etc, will soon be off patent (*2*), providing the opportunity for biosimilar mAbs to enter the biotherapeutics market. The high demand for mAb-based therapeutics has resulted in an increased need for rapid and structurally informative characterization techniques. Recent advancements in mass spectrometry (MS) for proteomics applications over the past two decades have enabled scientists to address the biophysical and biochemical properties of proteins using this technology with higher throughput. Today, MS has become one of the key techniques in a biopharmaceutical laboratory for confirmation of mAb primary sequence, identification and localization of post-translational modifications (PTMs), and characterization of higher order structure throughout all stages of mAbs development and production (*3*, *4*).

The first step in structural characterization is to confirm the molecular mass of a therapeutic mAb and its glycoforms, which provides information about protein heterogeneity. Commonly, intact protein mass measurement is performed under denaturing conditions, where the protein is ionized in acidified solvent used as the liquid chromatography MS (LC-MS) mobile phase (*4*, *5*). For therapeutic mAb, current commercial mass analyzers provide sufficient resolving power to resolve glycoforms at the intact protein level. mAbs can also be analyzed using native MS, where the protein is ionized in an aqueous buffer with pH close to neutral (*6*). Under native conditions, proteins and protein complexes retain their folded tertiary and quaternary structure. Native MS, therefore, may provide unique structural information not available using any other approach. Reported applications for native MS include probing antibody-antigen binding, structural features, and dynamics (*7*, *8*). Recently, native MS has gained interest in the analysis of bispecific mAbs and mAb mixtures due to the simplified, easy to interpret MS spectrum obtained under native conditions (*9*).

Further characterization is typically performed using a ‘bottom-up’ approach, in which the protein is digested using an enzyme into peptides and analyzed by

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

liquid chromatography tandem spectrometry (LC-MS/MS) (*10*, *11*). The bottom up approach provides molecular mass, amino acid sequence, type and location of PTMs for these peptides as well as impurities and degradants in protein drugs (*12*, *13*). However, the bottom-up approach is labor intensive and time consuming. Also, artifacts from enzymatic digestion and sample handling can interfere with this type of analysis (*14*, *15*). Furthermore, the information of protein isoform specific modifications is often lost in bottom-up approach where all the protein isoforms are subjected to enzymatic digestion before peptide mapping. The limitations of the bottom-up approach have become the driving force for the continued development of the top-down approaches.

The alternative and complementary protein characterization method is the top-down approach in which detailed structural information is obtained by dissociating the intact protein into smaller fragment ions in the mass spectrometer (*16*). The ideal scenario would be to separate the intact protein isoforms by liquid chromatography (LC) and then perform on-line mass and top-down analyses to determine the sequence and the sites of modifications. However, LC-MS of intact proteins is much more challenging than for peptides. This is mainly because of poorer solubility and separation of proteins under LC-MS conditions, as well as the lower sensitivity of mass spectrometers for intact proteins (*17*). Moreover, as protein size increases, the tertiary structure of proteins becomes more difficult to disrupt, thereby limiting the tandem mass spectrometry (MS/MS) fragmentation efficiency of intact proteins. Thus, most of the published top-down applications focused on proteins smaller than 50 kDa, although there are reported top-down of larger proteins (*16*, *18–21*). The development of more advanced, robust and easy to operate high resolution MS instruments, especially the new generation Orbitrap-based instruments, has made top-down analysis an increasingly practical approach for rapid protein characterization (*22*). The fast scan speed, ultra high resolution as well as the multiple fragmentation mechanisms offered by the high field Orbitrap instruments are advantageous for top-down protein analysis. Top-down analysis was reported to be successful in identifying site-specific modifications and amino acid changes in small proteins (*23–26*).

For large proteins, a ‘middle-down’ approach was recommended by cleaving proteins into a few large pieces before MS/MS analysis (*5*). The combination of intact protein mass measurement and top-down or middle-down analysis has been proven to provide accurate and detailed structural information such as molecular mass of isoforms, amino acid sequence and PTM site locations (*5*, *27*). For mAb characterization, the subunit sequencing approach on the intact light chain or heavy chain, or the middle-down approach on large fragments could potentially provide high resolution, in depth structural information.

In this chapter, we describe the current development of LC-MS and LC-MS/MS based intact mAb characterization methods to support the discovery and development of protein therapeutics in the pharmaceutical industry. Most of the data presented in this chapter was generated using a standard reference mAb, NISTmAb, supplied by the National Institute of Standards and Technology. For some of the experiments, samples other than NISTmAb were used as specified. Characterization under both denatured and native conditions are demonstrated. Solutions that include sample preparation, instrumentation and

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

data processing for routine intact mAb mass measurement, subunit sequencing, and middle-down structure analysis are introduced. Throughout this chapter, the use of state-of-the-art and emerging MS technologies, such as ultra high resolution measurement, multiple fragmentation mechanisms and native MS are presented and discussed. Recommendations for best practice are provided with special attention paid to key mass spectrometry parameters. The approaches described here are representative of those that can also be used for characterization of other therapeutic proteins.

Intact Antibody Mass Measurement

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Mass measurement is important in comfirming the heterogeneity of therapeutic mAbs including glycoforms and and other post-translational modifications as well as appropriate conjugation of payload molecules as necessary for antibody-drug conjugate (ADC) products. Molecular mass of intact monoclonal IgG antibodies, their subunits and large fragments has been measured using matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI) sources coupled with almost any type of mass analyzers. The first successfully used instruments were MALDI time-of-flight (MALDI TOF) (*28–30*) and electrospray ionization (ESI) quadrupole (*31–33*) mass spectrometers. The ESI TOF or ESI quadrupole time-of-flight (ESI qTOF) MS systems became a popular choice for intact mass measurement not long after due to their wide mass range and higher resolution to resolve individual glycoforms (*34–37*).

Since its launch in 2005, Orbitrap MS has become established as one of the most powerful tools for protein identification and characterization due to its advanced capabilities, including high sensitivity, high resolution, high mass accuracy, and wide dynamic range (*38*, *39*). The first report of the determination of intact molecular mass of antibodies using an Orbitrap instrument was by Zhang *et al.* at the ASMS Conference (2007), followed by another report at WCBP Conference (2008) from the same group. In these studies, the intact molecular mass of mAb was determined using a Thermo Scientific LTQ Orbitrap instrument with off-line infusion after desalting with a NAP-5 column (*40*, *41*). In 2009, Bondarenko *et al.* reported the first intact mAb mass analyses on an Orbitrap analyzer by on-line LC-MS using a similar approach (*3*). The analysis of large intact proteins by Orbitrap MS is challenging due to the fast decay of signal caused by loss of intact protein ions from fragmenting collisions and dephasing of ion packets (*42*). For best results, the detection needs to start as early as possible after ion injection from the C-trap into the Orbitrap analyzer. In the case of early generation Orbitrap systems, the detection is not switched on until 10-20 milliseconds (ms) after the ion injection into Orbitrap to allow voltage stabilization before detection. To improve the signal to noise ratio (S/N) for intact protein mass measurement, Bondarenko *et al.* pioneered the use of reduced FT detection delay in 2008 (*43*). However, the Orbitrap analyzer did not become a routine choice for intact mAb mass measurement until recent innovations of hardware and software significantly improved its performance. One of the most important innovations is the modifications to the preamplifier and Orbitrap

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

analyzer design which improved the stabilization of voltages, reducing the delay in the switching-on of detection from 10-20 ms to less than 0.6 ms (*3*, *44*, *45*). All the results presented in this chapter were obtained using the new generation Orbitrap instruments in which these innovations have been implemented.

The molecular mass of an intact protein is measured either as an average mass when each charge state signal of a target protein is detected as a single peak, or as a monoisotopic mass when the various isotopomers of the intact protein are well resolved. The instrument resolving power required for measuring intact protein monoisotopic mass is much higher than that needed for measuring average mass. For Orbitrap instruments, the numerical value of instrument resolution parameter setting represents the resolution at 200 *m/z* (for Exactive or Q Exactive instruments and for Orbitrap Fusion instrument) or at 400 *m/z* (for Orbitrap Elite instrument). For Orbitrap MS the resolution decreases with square root of mass. For example, the resolution of 140 K at 200 *m/z* becomes 100K at 400 *m/z*. Using the instrument resolution setting of 240 K or higher, a protein as big as 50 KDa, such as an antibody heavy chain, can be isotopically resolved with less than 1 minute (min) of spectrum averaging. Therefore, measurement of the isotopic mass of heavy chain with a mass accuracy of less than 5 ppm error can now be achieved on an LC time scale. For proteins as large as an intact mAb (150KDa), isotope resolution cannot be achieved using even the highest resolution setting offered by current commercially available MS instruments. However, isotopic resolution of a intact mAb was recently achieved in a modified Thermo Scientific Orbitrap Elite instrument using helium as the bath gas with reduced C-trap and HCD cell gas pressure, while trapping ions in the HCD cell prior to mass analysis. These modifications greatly increased sensitivity and reduced signal decay for large protein ions (*46*).

In the case of intact mAb, the average molecular mass is routinely measured using the lower resolution settings of an Orbitrap instrument, which reduces signal loss due to decay of large, intact protein associated with longer transient acquisition. The instrument resolution setting of 15K to 30K is recommended for measuring average mass of intact mAbs using a commercial Orbitrap instrument. Higher resolution (60K or higher) may be required for more complex or heterogeneous samples such as intact antibody mixtures or for heavily glycosylated intact proteins where more resolving power is requied to distinguish species in the sample. Besides resolution, several other MS parameters are also important for the successful measurement of an intact protein mass using the Orbitrap mass analyzer. For analysis over a broad mass range, the Automated Gain Control (AGC) setting, which controls the number of charges sent into the C-trap for subsequent transfer into the Orbitrap analyzer, is usually set to 1e6 for Q Exactive or 1e5 for Orbitrap Fusion instrument. The maximum injection time (IT) is the maximum allowed time to inject analyte ions into the trap. To ensure that enough protein ions are accumulated for analysis even when ion flux is low, the maximum IT is usually set to at least 100-200 ms. The actual IT for ion injection, which is determined based on the AGC setting and the ion flux, is much shorter when the ion flux is strong. To facilitate desolvation of intact protein ions, some in-source collision induced dissociation (In-source CID or SID) energy is usually applied. Under neutral pH, more energy is needed for

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the desolvation of intact protein. Thus, for native MS, besides higher In-source CID energy, additional energy deposition in the collision cell may be required for desolvation (section 3). To improve the spectrum quality, multiple micro scans are recommended for intact protein measurement although in most cases, a single micro scan can produce high quality spectrum. Beside the MS parameters, ionization conditions can also significantly affect the spectrum quality. The temperature of the Heated Electrospray Ionization (HESI) source probe should be lower for an intact protein than for peptide and small molecule analytes. Other settings, such as gas flow and spray voltage, are optimized during tuning of the instrument on the spray stability and signal intensity.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Measurement of Average Molecular Mass

Figure 1 shows a typical MS spectrum of the NISTmAb that can be routinely obtained using any new generation Orbitrap instrument. In this study, the NISTmAb was analyzed using a bench-top quadrupole Orbitrap instrument, the Thermo Scientific Q Exactive, coupled with online High-performance liquid chromatography (HPLC). The HPLC mobile phase used was water with 0.1% formic acid (FA) (solvent A) and acetonitrile with 0.1% FA (solvent B). 40 ng of this mAb was loaded onto a ProSwift RP-10R monolithic column (1 x 50mm, Thermo Scientific) and eluted using a 7 min gradient at a flow rate of 80 µl/min. The HESI source probe temperature was set to 100 °C and the capillary temperature at 320 °C. The spray voltage was set to 3.8 KV with a sheath gas flow rate at 20 and an auxiliary gas flow rate of 5. Data was acquired using MS instrument settings of 17.5K resolution, AGC of 1e6, In-source CID of 45eV, S-lens of 50, maximum IT of 200 ms and 10 microscans. The raw MS spectrum averaged over the 30 second elution peak shows the distribution of the charge states of this mAb under the acidic pH of the HPLC mobile phase (Figure 1A). For each charge state, several peaks representing glycoforms are baseline resolved (Figure 1B). This raw spectrum was then deconvoluted using the ReSpect algorithm in Thermo Scientific Protein Deconvolution 2.0 software. The deconvolution of a intact protein mass spectrum transforms a raw spectrum into a spectrum of molecular mass. All the charge states for a given component are collapsed into a single peak at the mass value of that protein in the spectrum. The resulting deconvoluted mass spectrum of NISTmAb shows the measured average molecular mass of five major glycoforms with an error of less than 15 ppm from the theoretical mass value that were calculated using ProteinProspector (Figure 1C).

To evaluate the reproducibility of mAb mass measurement using the Q Exactive instrument, the sample was analyzed six times under the same experimental conditions. The average ppm error is 5.0 +/- 10.9 for a total of 30 measurements for five different glycoforms across the six experiments (data not shown). For the top 5 glycoforms, the CV for relative intensity reproducibility is within 6.7 percent (data not shown).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

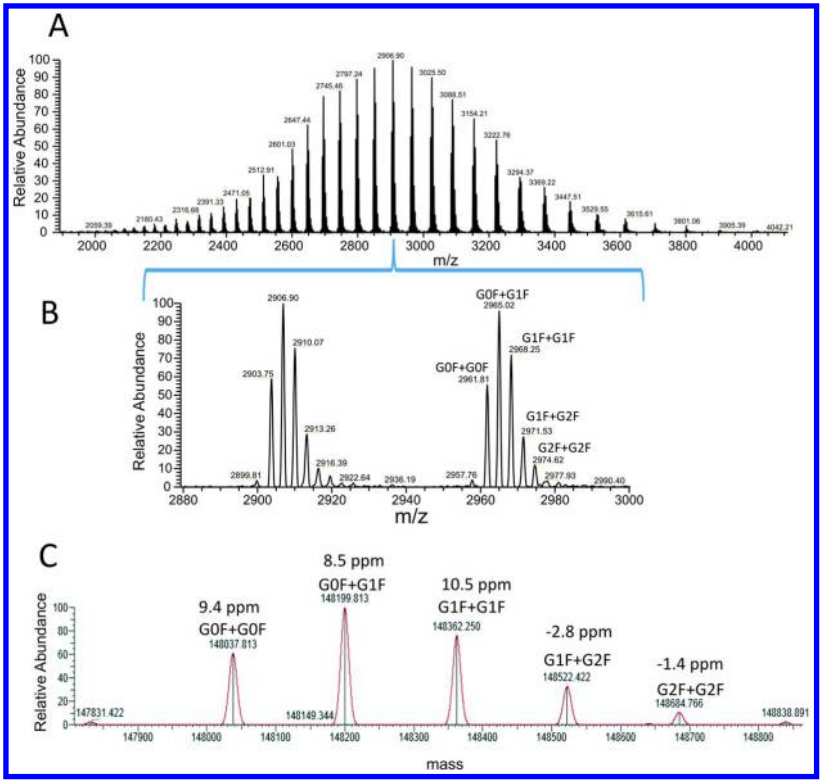

Figure 1. Molecular mass measurement of NISTmAb using Q Exactive Orbitrap LC-MS. (A) Full MS spectrum of intact NISTmAb showing charge distribution of this protein. (B) Detailed look of two consecutive charge states from A. Major glycoforms are baseline resolved and labeled. (C) Deconvoluted spectrum showing distribution and molecular mass of glycoforms. Major glycoforms and their calculated experimental errors in ppm are labeled.

Measurement of Monoisotopic Molecular Mass

Characterization of the mAb light and heavy chains (subunit analysis) not only confirms molecular mass of the two subunits, but also provides component specific information such as modifications specific to the light or heavy chain and a higher resolution analysis of the glycoform distribution on the heavy chain. The light and heavy chain of the NISTmAb were produced by denaturing and reducing the intact antibody. The two chains were separated on a PLRP-S 1000A column (2.1x50mm, 5µm Agilent Technologies) using a 9 min gradient and then directly introduced to a Thermo Scientific Q Exactive Plus mass spectrometer. This instrument provides an optional resolution setting up to 280K and a software controlled protein mode that has regulated gas pressure to facilitate high resolution intact protein analysis. A resolution setting of 140K and standard operation mode is enough to resolve the isotopic peaks of light chain, while a 280K instrument resolution setting in conjunction with the protein mode was necessary to produce

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

isotopically-resolved peaks of the heavy chain. The results were analyzed using the Xtract deconvolution algorithm in the Protein Deconvolution software. The Xtract algorithm matches a theoretical isotopic distribution to the experimental spectrum to generate monoisotopic mass.

In the full MS spectrum of NISTmAb light chain, isotopic peaks of each charge state are baseline resolved (Figure 2A). Deconvolution of this spectrum generated monoisotopic mass of 23113.3196 Da, 0.7 ppm off the expected value (Figure 2B). To achieve the best S/N for an isotopic resolved heavy chain spectrum, the Selected Ion Monitoring (SIM) scan mode with isolation window of 20 *m/z* was used to monitor only a single charge state with AGC of 2e5 and maximum IT set to 800 ms. With spectrum averaging over the two-min LC elution peak, the isotopic peaks of all the three major glycoforms of the heavy chain are baseline resolved (Figure 2C). The apparent noisiness of the spectrum was due to the lower intensity signal as a result of glycoform heterogeneity of the heavy chain. The Xtract deconvolution with relatively conservative parameters of 20 for S/N and 92% for the Fit Factor produced the 3 peaks for heavy chain corresponding to the three major glycoforms (Figure 2D) while lowering some thresholds would produce more deconvoluted peaks down in the noise. Measuring monoisopotic mass of the heavy chain on an LC time scale is challenging. Reproducible measurement with less than 5 ppm mass error can be achieved when the specrum has a good S/N of above 3. The measured monoisotopic masses for the two major glycoforms of NISTmAb heavy chain, G0F and G1F, were within 3 ppm of the theoretical value. The measurement error for the third, G2F, was 20 ppm due to low S/N (Figure 2D).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Native MS of Antibody and Related Products

An alternative strategy to investigate intact protein structure using MS, the native MS approach, is to introduce proteins or protein complexes into the mass spectrometer in or near their native conformation. Conformational variations within proteins or protein complexes can be monitored based on their charge-state distribution in native MS (*47*). Within the last ten years, native MS has emerged as a valuable technique for intact, noncovalent protein complex characterization, reaching a high level of reliability (*6*, *48*, *49*). To perform a native MS experiment, an ESI compatible volatile solution with near neutral pH is used to prepare protein samples (*50*)**.** Under these conditions, the ionized proteins carry fewer charges than those produced by ESI in acidic, denatured condition, leading to their detection at a significantly higher *m/z* range.

Native MS analysis provides fast and reliable structure information of antibody aggregates, antibody-antigen complexes, and antibody-drug conjugates (*51*, *52*). Because of the reduced spectrum complexity compared to classical denaturing MS, native MS has also gained interest for characterization of bispecific mAbs and antibody mixtures (personal communication). This approach is now ready for the routine characterization of heterogeneous therapeutic mAbs.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

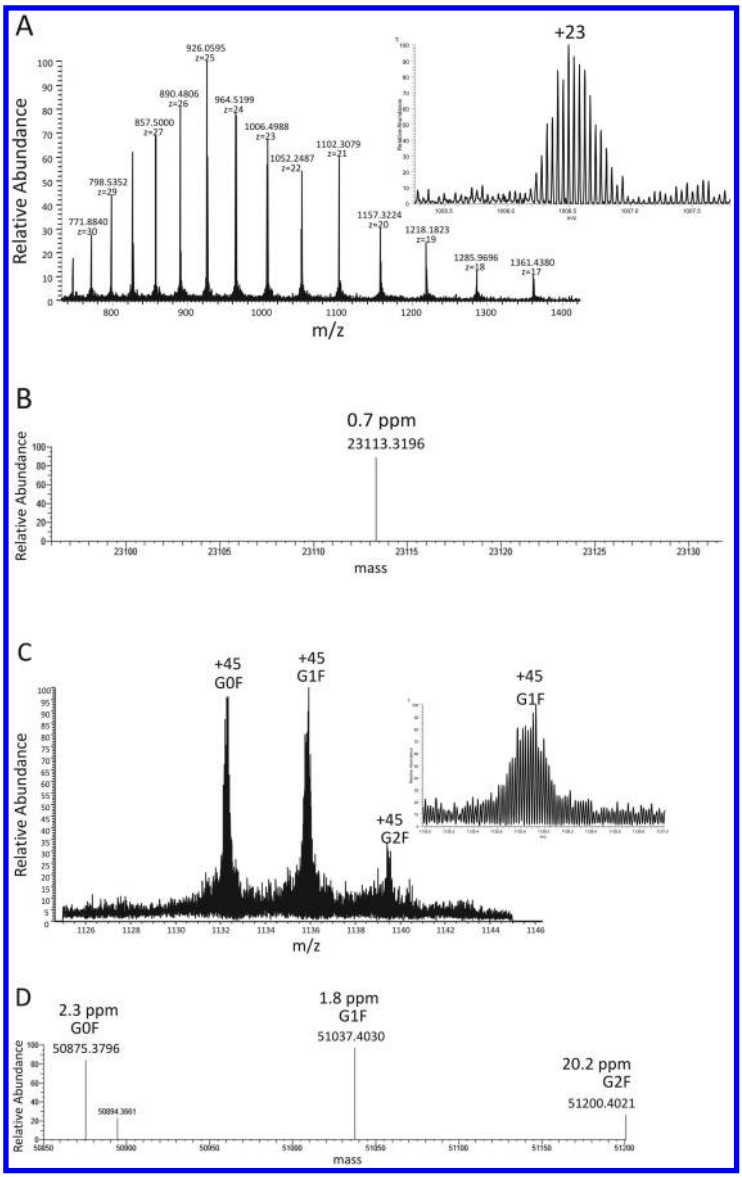

Figure 2. Monoisotopic mass measurement of the NISTmAb light chain and heavy chain using bench-top Q Exactive Plus Orbitrap LC-MS. (A) Full MS spectrum showing charge distribution of NISTmAb light chain. Data was collected over a half-a-min elution peak. Insert is the +23 charged light chain. (B) Deconvoluted spectrum showing a monoisotopic mass of NISTmAb light chain. Calculated experimental error in ppm is labeled. (C) MS spectrum of +45 charged heavy chain. Insert shows baseline resolved isotopic peaks for G1F form of heavy chain. (D) Deconvoluted spectrum showing monoisotopic mass of three glycoforms of NISTmAb heavy chain. Calculated experimental errors in ppm for each glycoforms are labeled.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Sara Rosati *et al.* first reported native MS analysis of intact mAbs using high-resolution Orbitrap mass spectrometry on a modified Exactive Plus instrument (*53*). The routine operation of Orbitrap instrument for native MS of mAbs was not possible until the commercialization of the first native MS Orbitrap system, the Thermo Scientific Exactive Plus EMR MS, which offers an extended mass range up to 20,000 *m/z* and improved detection of high-mass ions. Using this system, native MS can be performed using direct sample infusion or coupled with size exclusion LC for front end sample desalting or separation. With direct infusion, molecular mass measurements of mAbs and related products can be accomplished in a very rapid analysis with low ppm measurement error.

Native MS Analysis of Antibody-Antigen Complex

The sample used in this section, J10.4, is a commercial mouse monoclonal IgG1 raised against recombinant JAM fusion protein of human origin for detection of JAM-A. JAM-A is over-expressed in many tumor tissues and therefore is of prime interest as a target in oncology. Two JAM-A molecules are expected to bind to one J10.4 mAb. The native mass spectrum of this antibody-antigen (mAb-Ag) complex, is shown in Figure 3A. The J10.4 mAb was mixed with JAM-A antigen after they were buffer exchanged against 150 mM ammonium acetate. The native MS spectrum was recorded at a resolution setting of 35K with an In-source CID voltage of 150 eV to facilitate the desolvation of intact protein ions under neutral pH. Other MS instrument parameters used were 3e6 AGC target value, 10 microscans, 300 ms maximum IT and 100% S-lens level. With spectrum averaging over 1.2 min, a S/N of greater than 100 was achieved. The raw spectrum, showing the entire charge state distribution in native condition, was deconvoluted using Protein Deconvolution software. When a 4-fold excess of JAM-A (20 μM) was added to J10.4 mAb (5 μM), three species were detected: the intact free J10.4 mAb, 1:1 and 1:2 J10.4 mAb:JAM-A complexes (Figure 3B). The measured molecular masses correspond to the main G0F/G0F glycoforms of this antibody. Using the relative abundances estimated from MS peak intensities, the proportions of free mAb and mAb:Ag complexes at 1:1 and 1:2 stoichiometries were calculated to be 33%, 37% and 30% respectively.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

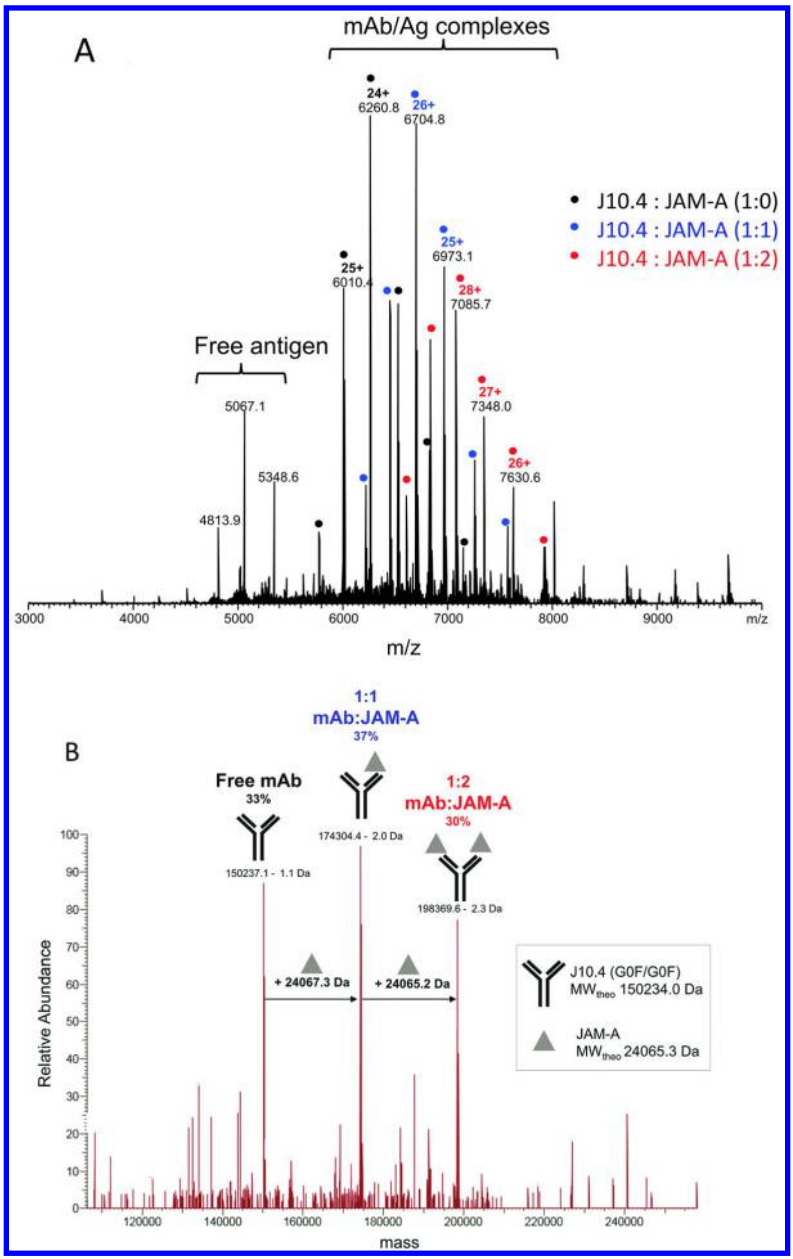

Figure 3. Native MS analysis of mAb-Ag complex. (A) Full MS spectrum showing the charge state distribution of J10.4-JAM-A complex when 4 fold JAM-A (20 μM) was added to J10.4 mAb (5 μM). (B) Deconvoluted spectrum showing the intact free J10.4 mAb (MW 150237.1 ± 1.1 Da, black), 1:1 (174304.4 ± 2.0 Da, blue) and 1:2 (198369.6 ± 2.3 Da, red) J10.4 mAb:JAM-A complexes. (see color insert)

Native MS Analysis of a Mixture of Eleven N-Deglycosylated Humanized Antibodies

Cocktails of mAbs with additive or synergistic effects are increasingly foreseen as potential new therapeutic entities. An example of using native MS to analyze an antibody mixture is shown in Figure 4. This mixture includes 1 μM of 2 marketed therapeutic mAbs, rituximab and trastuzumab, and 9 point mutation variants of the Hz6F4-2 mAb (*7*, *54*). Figure 4A presents a full native mass spectrum of the mixture of eleven distinct deglycosylated mAbs. To resolve this complex mixture, a 140K instrument resolution setting was used with 100eV in-source CID energy. The data was collected over 3.9 min of spectrum averaging. Deconvolution of the well resolved spectrum resulted in the unambiguous assignment of ten out of the eleven compounds (Figure 4B). Trastuzumab and Hz6F4-2v6 cannot be differentiated due to the very small difference in their molecular mass of only 2 Da. Peaks corresponding to Hz6F4-2 and Hz6F4-2v3, which differ by only 21 Da in molecular mass, are clearly distinguished on the mass spectrum. These two species are not baseline resolved due to the low S/N, resulting in relatively high measurement error of 37.3 ppm for Hz6F4-2. However, in the case of Hz6F4-2v9 and 6F4-2v10 when S/N is better, low ppm measurement error is achieved for both species even though peaks are not baseline resolved. The measured and theoretical masses for the mixture of eleven N-deglycoslated humanized antibodies are listed in Table 1. The result indicates that native MS using the Exactive Plus EMR system enables high throughput screening of mAb mixtures with an excellent mass accuracy for each individual mAb.

Sequencng of Intact Subunit and Large Fragment of mAb

Top-down, intact subunit sequencing and middle-down approaches overcome some disadvantages of bottom-up approach in which the digestion of proteins leads to the loss of information about intact protein isoforms. These methods include separation of intact proteins, their subnunits or large fragments, ionization, and fragmentation followed by mapping of the fragment ion mass spectral data to the target sequence using database searching software. General discussion of protein top-down and middle-down approaches, including strengths and weaknesses compared to bottom-up approach and current applications, is out of the scope of this chapter, the readers are therefore referred to other review articles (*16*, *25*, *55*, *56*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

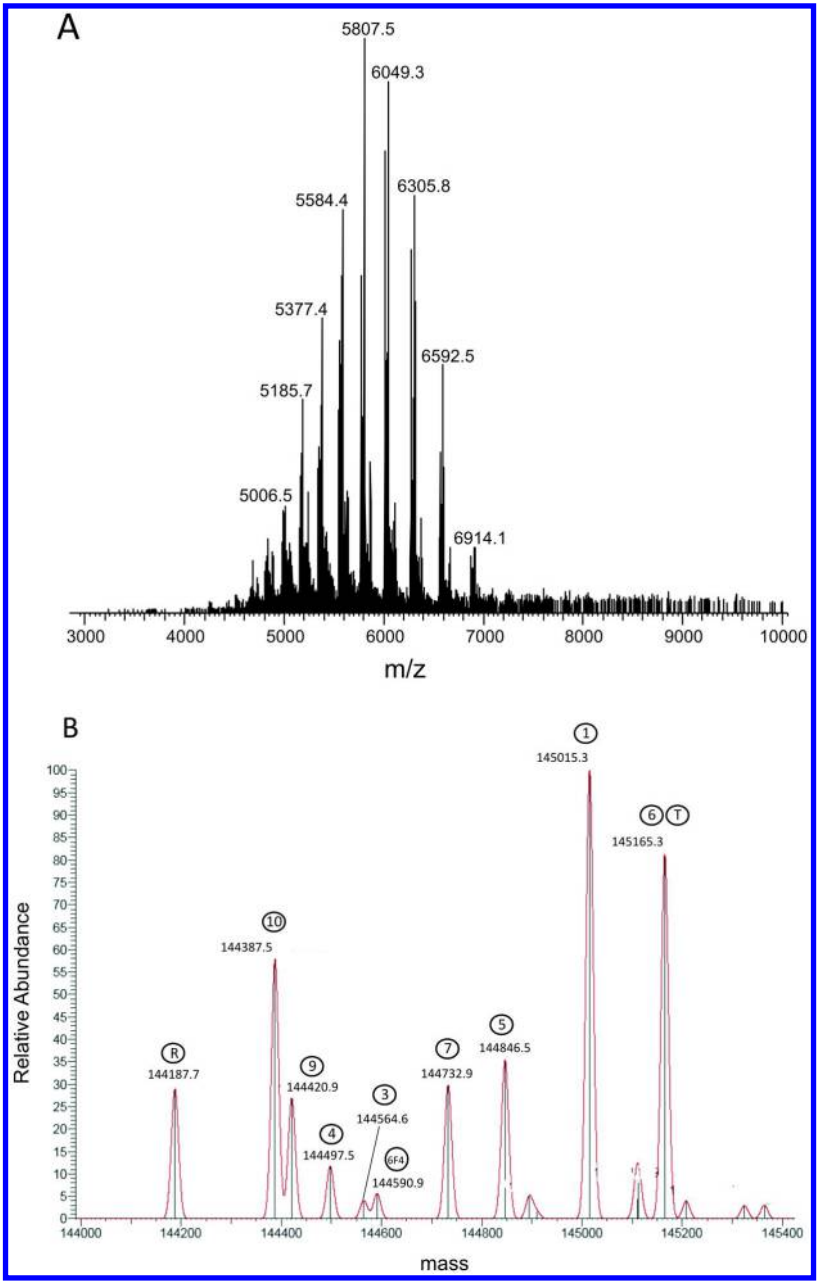

Figure 4. Native MS analysis of a mixture of eleven N-deglycosylated humanized antibodies. (A) Full MS spectrum of eleven, N-deglycosylated mAb mixture. (B) Deconvoluted spectrum of the mAb mixture.

Table 1. The Measured and Theoretical Masses for the Mixture of Eleven N-Deglycoslated Humanized Antibodies

	Species	*Theoretical masses (Da)*	*Measured masses (Da)*	*Mass accuracy (ppm)*
R	Rituximab	144186.3	144187.7	9.7
10	6F4-2 v10	144388.3	144387.5	5.5
9	6F4-2 v9	144420.5	144420.9	2.8
4	6F4-2 v 4	144498.4	144497.5	6.2
3	6F4-2 v3	144564.4	144564.6	1.4
6F4	6F4-2	144585.5	144590.9	37.3
7	6F4-2 v7	144732.5	144732.9	2.8
5	6F4-2 v5	144846.9	144846.5	2.8
1	6F4-2 v1	145015.3	145015.3	0
6	6F4-2 v6	145163.3	N.D	N.D
T	Trastuzumab	145165.5	145165.3	1.4

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Top-down fragmentation of intact antibodies and antibody subunits using Orbitrap mass spectrometers has been reported using In-source CID (*26*) or using electron transfer dissociation (ETD) (*57*). ETD has also been utilitzed on a qTOF mass spectrometer to fragment intact antibodies (*58*). Top-down MS/MS of antibodies challenges a mass spectrometer to the extreme as the spectra produced are extremely complex and contain vast populations of overlapping, multiply charged product ions. Therefore, high performance on top-down sequencing can only be achieved by MS instruments with high resolving power and high sensitivity. The multiple fragmentation mechanisms offered by some MS instruments, such as CID, ETD and higher energy collision dissociation (HCD) in some of the Orbitrap instruments, produce complementary product ions leading to improved sequence coverage. Recent innovations in new generation Orbitrap instruments, especially the high-field Orbitrap mass analyzer and the enhanced Fourier transform algorithm (eFT), further improved the resolving power and scan speed. Top-down sequencing using Orbitrap instrument has become a practical approach suitable for LC-MS/MS (*45*).

The main limitation of top-down LC-MS/MS of large proteins is the relatively low S/N due to large number of potential fragmentation channels from a protein precursor ion. To generate a spectrum with a good S/N, averaging as many scans as possible during the elution of a protein in a LC-MS/MS experiment is the common practice. For analysis using Orbitrap MS, in addition to the multiple micro scans set in the instrument method, post acquisition spectrum averaging over the LC elution peak is also performed to generate an averaged MS/MS spectrum for sequence assignment with improved S/N. To further improve S/N of a top-

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

down MS/MS spectrum, the AGC setting, which controls the number of charges sent into the trap, and the maximum IT can be set to a higher level than that for peptide MS/MS analysis. However, the number of charges should be still kept below the onset of space charge effects to avoid deterioration of spectrum quality. Usually, one of the relatively high abundant charge states is selected as precursor for top-down MS/MS. When the precursor ion intensity is low, a wider isolation width of 100 *m/z* or more can be used so that multiple charge states are selected for fragmentation. In some of the Orbitrap instruments, such as the Q Exactive series, a unique spectrum multiplexing mode allows isolation of multiple charge states for top-down sequencing by performing multiple, consecutive isolations of single charge state using a narrow isolation window. Ions from each charge state precursor are fragmented followed by Orbitrap detection of all the fragment ions from the multiple precursors.

Another challenge in top-down sequencing of large protein is the limitation of fragmentation efficiency by higher order structure of proteins. With increasing protein size, higher order structure requires more energy to disrupt. Due to such limitations of top-down sequencing of large intact proteins, the subunit sequencing and middle-down approaches have emerged as potentially more practical. Subunit sequencing of mAbs involves reduction of interchain disculfide bonds to produce disociated heavy and light chain subunits for analysis. Middle-down protein mass spectrometry requires enzymatic proteolysis of proteins into long (3–20 kDa) peptides and efficient separation of such mixtures prior to MS/MS. A previous study has shown that the middle-down approach could achieve similar sensitivity as bottom-up approach and could be performed under the same chromatographic conditions as well (*59*), indicating its significant advantage over a conventional top-down approach. Meanwhile sequencing larger peptides enables identification of combinatorial PTM patterns that may be missed in a standard bottom-up experiment (*60*, *61*) and also may produce more sequence coverage for the target protein.

mAbs can be enzymatically cleaved at the hinge region to produce Fab and Fc fragments. Reduction of disulfide linkage further produces a variety of fragments for analysis including the light chain subunit, the N-terminal half of the heavy chain (Fd′) from the Fab fragment, and the C-terminal half of the heavy chain (single chain Fc, or scFc) from the Fc fragment. The three large fragments produced from mAb by this process are 20-25 kDa, a size that can be routinely analyzed with high-resolution MS instruments. Assembly of sequencing results from the three large pieces of a mAb substructure can potentially produce significantly improved structure information compared to top-down sequencing of intact antibody or of intact heavy chain.

Sequencing of Intact mAb Subunit

Light and heavy chains of NISTmAb were produced by denaturation and reduction and then separated on a 500 μm X 5 cm PepSwift monolithic column (Thermo Fisher Scientific). The online intact subunit MS/MS was performed using an Orbitrap Elite instrument with both ETD and HCD (Figure 5A). Although alternating HCD and ETD can be performed in a single LC-MS/MS

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

run, in this study the two methods were performed in separate runs to achieve maximum combined sequence coverage. For ETD MS/MS, activation time was 5 to 50 ms while for HCD MS/MS, normalized collision energy level was 10-25 %. The resulting averaged spectra were deconvoluted using the Xtract function of Xcalibur software and analyzed using ProSightPC 3.0 software with a fragment ion tolerance of 10 ppm.

ETD or HCD of the intact light chain generated spectra containing numerous multiply charged product ions which are well resolved using the instrument resolution setting of 120K (Figure 5B). From the ETD spectrum generated using an activation time of 15 ms, 111 ***c*** and ***z*** ions that were identified corresponding to 52% backbone fragmentation coverage. For HCD, the maximum backbone fragmentation of 39% was achieved using 18% normalized collision energy, producing 85 ***b*** and ***y*** ions. The combined backbone fragmentation of the light chain from both methods reached 67% using a single ETD activation time and a single HCD collision energy (Figure 6A).

ETD of intact subunits requires much shorter reaction time than that for peptides because the number of charges carried by a subunit is much higher. In an Orbitrap Elite instrument, using AGC setting 5e5 for both ETD reagent ion and the precursor ion, reaction time as short as 8 ms was enough to generate extensive fragmentation of the NISTmAb light chain. The reaction time that generates the highest sequence coverage is different for precursor ions that carry different number of charges. For +25 charged precursor at 926.06 *m/z*, 15 ms generated the most sequence coverage. With the increase of reaction time, more terminal c and z ions were identified while the overall sequence coverage dropped (data not shown), presumably due to secondary fragmentation that breaks the longer fragment ions to produce shorter, terminal ones. This correlation of ETD reaction time and terminal sequence coverage was also observed previously using standard intact proteins such as carbonic anhydrase and enolase (data not shown). Thus, longer ETD reaction time is recommended for achieving maximum terminal sequence coverage. Table 2 lists the number of ***c*** and ***z*** ions identified at different reaction times for precursor ion at 926.06 *m/z*.

With increasing protein size, the higher order structure that limits MS/MS fragmentation efficiency becomes more difficult to disrupt. Further, larger proteins have more potential fragmentation channels due to the increased length of the peptide backbone, leading to decreased signal for any given fragment and increased complexity in the MS/MS spectrum overall. Subunit sequencing of the intact heavy chain is therefore more challenging than that of light chain. In this study, a subunit analysis of the intact NISTmAb heavy chain over a one-min LC elution peak produced combined backbone fragmentation of 21% from ETD and HCD. The backbone fragmentation from each individual method was 11% from 10 ms of ETD reaction and 12% from HCD with 20% normalized collision energy, respectively (Figure 6B). The largest fragment ions from ETD, $\boldsymbol{c}_{215}$ and $\boldsymbol{z}_{117}$, cover 74% of the heavy chain sequence. High resolution and accurate mass allowed the identification of large fragment ions with molecular mass over 10 kDa. The N-terminal modified amino acid, pyroglutamate, and the C-terminal lysine loss were identified using either the ETD or HCD approaches (Figure 6B).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

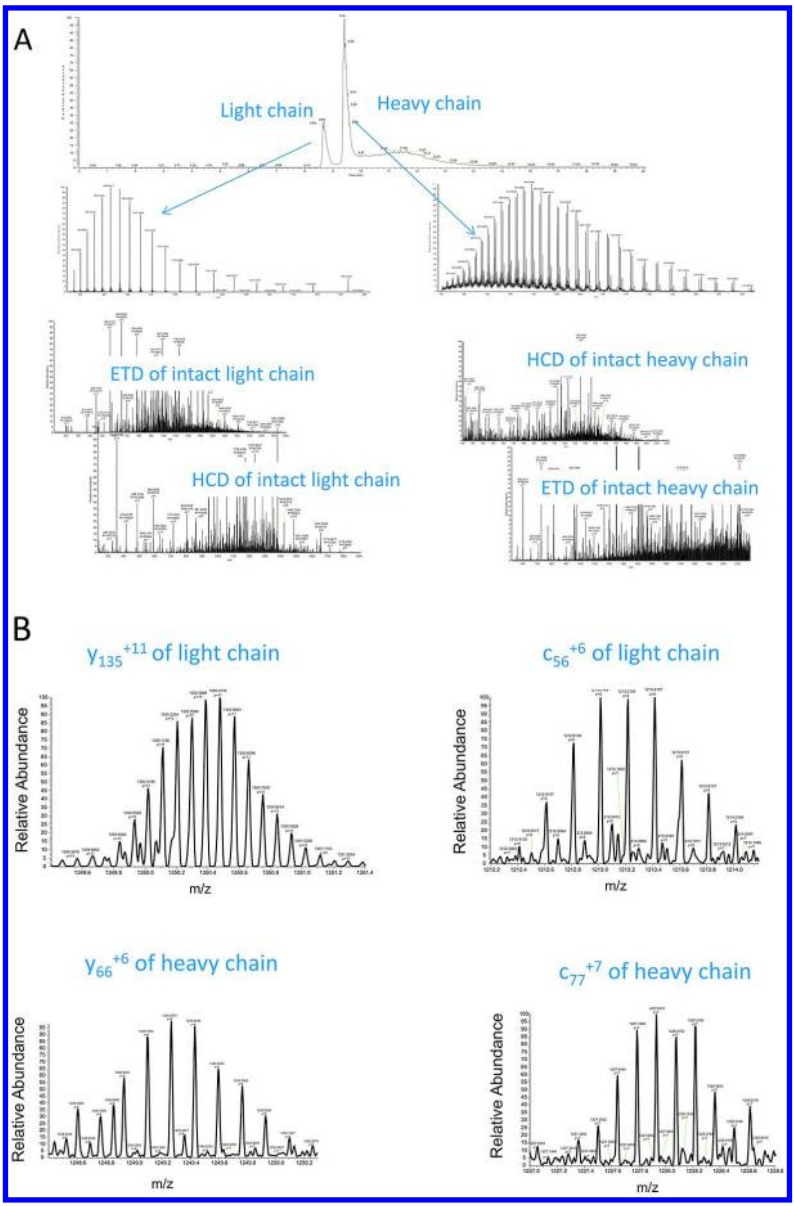

Figure 5. LC-MS/MS of NISTmAb intact light and heavy chains. (A) The online MS/MS of NISTmAb intact light and heavy chains using both ETD and HCD. For light chain, +25 charged precursor at 926 m/z and +18 charged precursor at 1285 m/z were selected for ETD and HCD, respectively. For heavy chain, +56 charged precursor at 912 m/z and +48 charged precursor at 1064 m/z were selected for ETD and HCD rewspectively. (B) Examples of large, multiply charged fragment ions identified from heavy chain and light chain using ETD and HCD of intact subunit.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Besides primary sequence and PTM information, results from intact subunit MS/MS can provide additional structural confirmation. Comparing subunit HCD sequence coverage of partially and completely reduced mAb light chain can reveal the position of disulfide linkages. In one of our previous studies of partially reduced light chain of another mAb (not NISTmAb), subunit HCD identified only those ***b*** and ***y*** ions that are not bridged by the two internal disulfide linkages. For the completely reduced light chain, significant number of b and y ions beyond the disulfide linkage-forming cysteine residues were also identified (*62*).

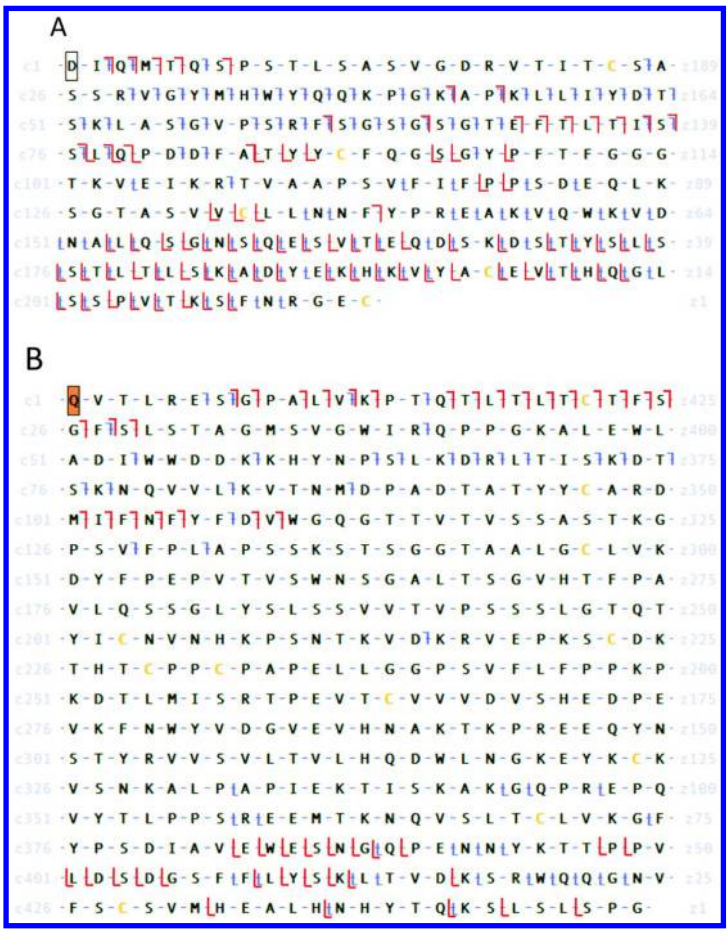

Figure 6. Protein backbone fragmentation and sequence coverage for NISTmAb light and heavy chains from intact subunit MS/MS. (A) Combined backbone fragmentation and sequence coverage for the intact light chain from ETD (blue) and HCD (red). +25 charged precursor at 926 m/z and +18 charged precursor at 1285 m/z were selected for ETD and HCD, respectively. (B) Combined backbone fragmentation and sequence coverage for the intact heavy chain from ETD (blue) and HCD (red). +56 charged precursor at 912 m/z and +48 charged precursor at 1064 m/z were selected for ETD and HCD rewspectively. (see color insert)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Table 2. Number of *c* and *z* Ions Identified from Intact NISTmAb Light Chain Using Different ETD Reaction Time

ETD activation time (ms)	*Number of c ions*	*Number of z ions*
4	5	4
8	15	21
15	63	48
20	42	25
50	22	21
100	14	12

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Middle-Down Sequencing of mAb

To obtain in depth structural information for heavy chain, a middle-down approach was used. The NISTmAb was subjected to digestion in the hinge region using FabRICATOR (Genovis, Sweden), producing F(ab′)2 and Fc fragments. The disulfide linkage in the resulting F(ab′)2 and Fc fragments were further reduced with 30mM DTT to produce the Fab portion of the heavy chain fragment (Fd′), the single chain Fc fragment (scFc) and the light chain subunit. Each of these antibody substructures is about 25 kDa and, therefore, is significantly easier to fragment by MS/MS than the 150 kDa intact mAb. The above products of digestion and reduction were separated by HPLC using a PepSwift monolithic column (200 μm x 25 cm, Thermo Fisher Scientific) at a flow rate of 0.8 μl/min. The mobile phases used were water with 0.1% FA (solvent A) and acetonitrile with 0.1% FA (solvent B). A linear gradient of increasing solvent B was utilized for the separation of light chain, Fd′, and scFc fragments. Each of the three parts of this mAb was analyzed online with both HCD and ETD using the Orbitrap Fusion mass spectrometer. The resulting spectra were deconvoluted using the Xtract function of Xcalibur software and subsequently analyzed using ProSightPC 3.0 with a fragment ion tolerance of 10 ppm.

Precursor ions at 964.65 *m/z* from the light chain, 1117.87 *m/z* from the Fd′ and 902.20 *m/z* from the G0F glycoform of scFc fragment were selected for ETD and HCD with an instrument resolution setting of 120K. The ETD activation time was 3-5 ms and the normalized collision energy for HCD was 15-25%. The high resolution ETD and HCD spectra generated contain numerous well resolved, multiply charged fragment ions (Figure 7A). Matching these spectra to the sequence of this mAb identified complementary fragment ions that were generated from backbone fragmentation of 50% for light chain (data not shown) and 52% and 32% for scFc and Fd′, respectively, when combining results from ETD and HCD (Figure 7B, 7C). Fragment ions from both ETD and HCD of Fd′ confirmed N-terminal modification of pyroglutamate and those of scFc confirmed the Lys loss at the C-terminus. Furthermore, the extensive backbone fragmentation on scFc, especially the 80% backbone cleavage up to Asn61 as well as the multiple identified fragments between Asn61 and Asn79, unambiguously locates the glycosylation site to Asn61 on the scFc (Asn300 of heavy chain).

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

ETD is known to be advantageous in locating the sites of labile PTMs which remain attached to fragments ions during ETD fragmentation. In this experiment, a few fragment ions containing the glycan, G0F, were identified from ETD spectrum of scFc fragment, providing further evidence of the type of glycoform and glycosylation site. The high resolution and accurate mass from Orbitrap MS enabled the confident identification of low abundance glycan-containing fragment ions in the presence of interference. An example of this case is the identification of the +9 charged $\boldsymbol{c}_{61}$ ion with G0F (Figure 7A insert). Although not all the isotopic peaks of this ***c*** ion were observed due to background interference, at least seven of them were identified within 3 ppm mass error. Figure 7A inserts show two of the fragment ions that are critical in identification and localization of G0F on Asn61 residue.

Results from this study indicated that extensive backbone fragmentation, as well as the fragmentation mechanism that retains labile PTMs allowed the identification and localization of PTMs including glycosylation. Middle-down analysis produced information deep into the middle of the heavy chain, the part of the molecule that was not accessible using intact subunit sequencing approach Figures 6B, 7B, 7C.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Automation of LC-MS Based Intact Therapeutic mAb Analysis for Harvest Cell Culture Screening

During development of recombinant mAbs, a large of number of harvest cell culture (HCC) samples must be screened for IgG titer, aggregation, and charge variants. Affinity chromatography is often used first to purify mAbs, with typical yields of more than 95%. Size-exclusion chromatography (SEC) is used to identify and quantify mAb aggregations. Finally, ion-exchange chromatography (IEC) is used to characterize charge variants.

A complete workflow solution has been developed that includes two dimensional chromatography purification followed by Orbitrap characterization for screening of HCC. The 2D-HPLC workflow uses an integrated HPLC system consists of a dual-gradient pump, a UV/VIS detector, a column oven, and an autosampler capable of both sample injection and fraction collection (Figure 8). In the first dimention of affinity purification, unpurified harvest cell culture sample is loaded onto a Protein A column. Elution from protein A column is automatically collected into a 96 well plate in the autosampler. The IgG fractions are then injected separately onto Thermo Scientific MAbPac SEC-1 (4 × 300 mm) and MAbPac SCX-10 (3 μm, 4 × 50 mm) columns for further analysis (Figure 9A). Furthermore, the fractions collected off the MAbPac SCX-10 column are analyzed by an Orbitrap mass spectrometer (Figure 9B). Using this workflow HCC was characterized by affinity purification, followed by SEC and charge variant analysis. The combination of chromatography separation and Orbitrap LC-MS detection provides an efficient way to obtain structural information of mAb variants.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

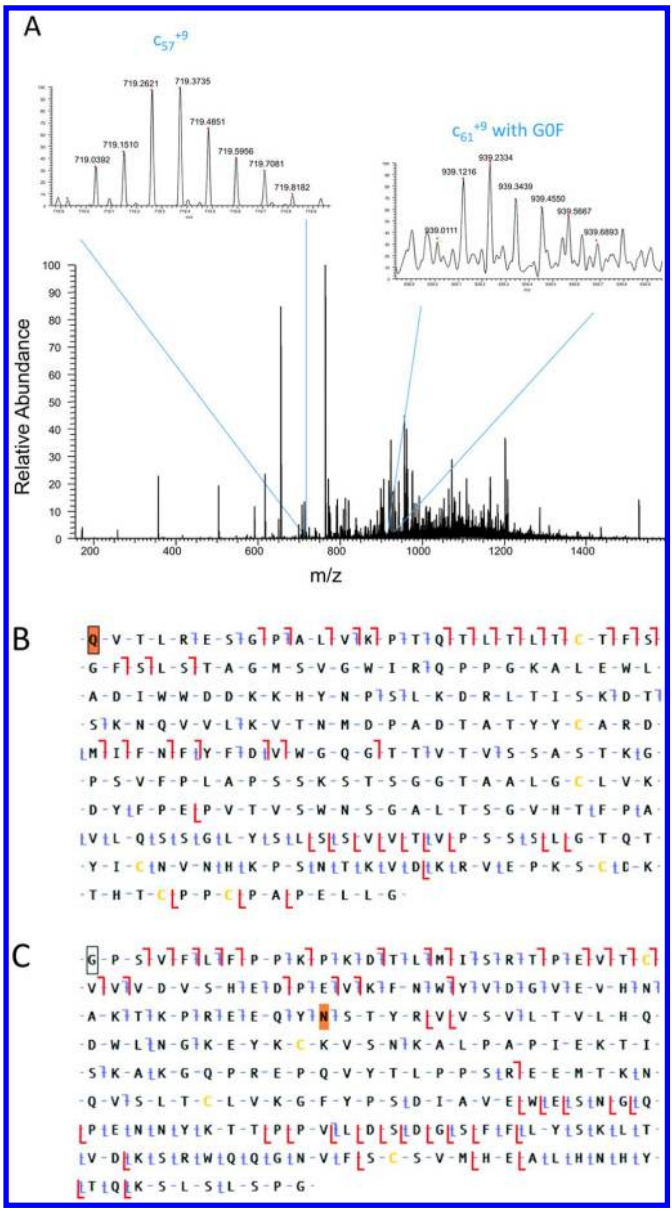

Figure 7. Middle-down LC-MS/MS of NISTmAb. (A) High resolution ETD spectrum of scFc G0F. The inserts show two fragment ions, $\mathbf{c}_{61}$ *with G0F and* $\mathbf{c}_{57}$*, which are among several fragment ions critical in identification and localization of G0F to Asn61 in Fc. (B) Combined backbone fragmentation and sequence coverage for Fd' from ETD (blue) and HCD (red) of precursor ion at 1117.87 m/z. (C) Combined backbone fragmentation and sequence coverage for scFc G0F from ETD (blue) and HCD (red) of precursor ion at 902.20 m/z. The Asn residue highlighted in brown is the glycosylation site. (see color insert)*

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

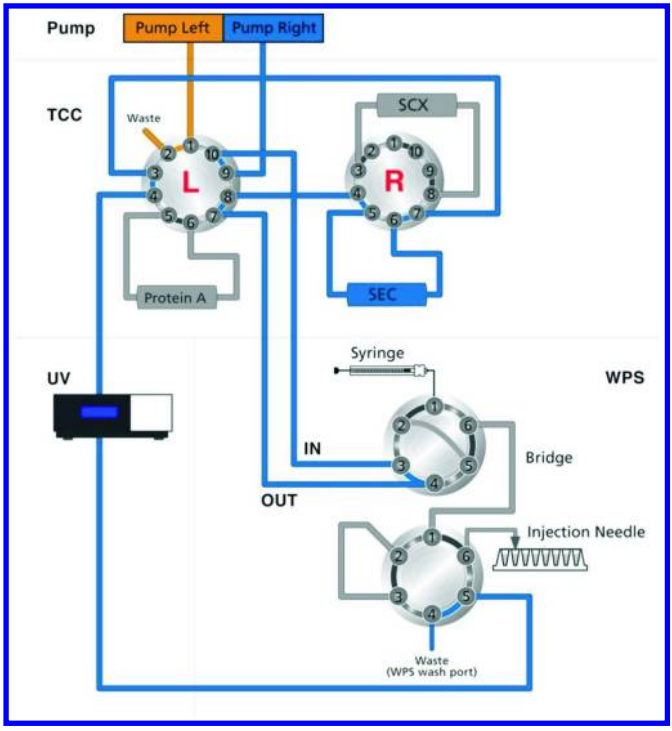

Figure 8. Fluidic configuration of the automated 2D-LC system using the well plate bio-inert autosampler. HCC purified by the first dimension Protein A affinity column are collected by the autosampler into IgG fractions which are then injected separately onto MAbPac SEC-1 and MAbPac SCX-10 column for further analysis. (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

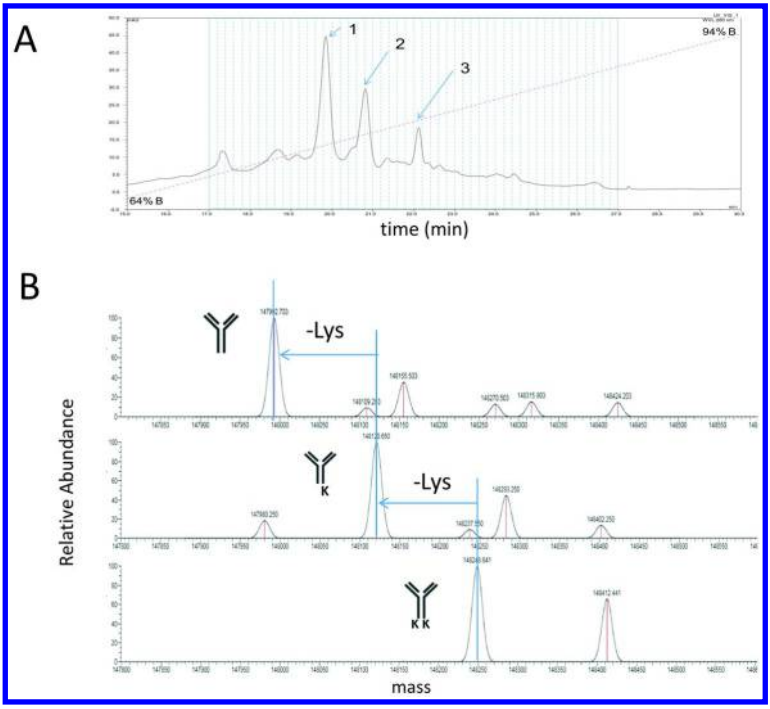

Figure 9. Characterization of mAb (not NISTmAb) charge variants using pH-gradient separation and Orbitrap LC-MS. (A) Elution peaks of a mAb variants from MAbPac SCX-10 column. Three major peaks were collected for MS analysis. (B) Deconvoluted mass spectrum for the three peaks from A. The mass difference between the mAb charge variants confirms C-terminal loss of one, or two Lys residues.

Conclusions and Perspectives

Recent innovations and improvement of the Orbitrap technology has made this type of high resolution MS instrument a routine tool for intact mAb analysis. The analysis of intact mAb provides molecular mass and relative abundance of different glycoforms with excellent reproducibility across experiments and even instruments (*62*). The ultra high resolving power of Orbitrap instruments enables the monoisotopic mass measurement of not only the light chain, but also the heavy chain. The recently developed Orbitrap instrument with an extended mass range and improved ion transmission of high mass range, the Exactive Plus EMR, is an outstanding tool for sensitive and accurate measurement at neutral pH of whole protein complexes in the 100 KDa to 1 MDa range (*63*). Native MS can be used not only to monitor protein complexes such as antibody-antigen complex and antibody aggregates, but also protein mixtures or heterogeneous intact proteins such as antibody cocktails and glycoproteins. The Orbitrap-based intact

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

subunit sequencing and middle-down approach efficiently dissociates the light chain and the heavy chain of mAb, producing in-depth structural information. The improved resolution and scan speed enables online LC in conjunction with intact subunit or middle-down MS/MS, giving these approaches practical utility in a high-throughput format. The use of complementary dissociation methods not only improves sequence coverage, but also allows the confident identification and localization of labile PTMs. With the continued development of technology on sensitivity, resolution and dynamic range, more information can be obtained by analyzing intact proteins with low abundance modifications, isoforms and impurities. The robust, high performing methods presented here can be used routinely in biopharmaceutical applications.

The high performance of Orbitrap instrument on subunit or middle-down sequencing has allowed new approaches to be developed for MS-based protein bioanalysis. Besides comprehensive characterization, bioanalysis, the concentration determination of mAb drug substances and their related compounds in biological matrixes is also a critical part of drug discovery and development. Bioanalysis of protein therapeutics is conventionally performed through monitoring the representative peptide(s) resulting from analyte protein digestion, a process that requires time consuming sample preparation. Furthermore, the peptide level approach cannot differentiate a protein analyte and its products derived from in vivo degradation processes if unchanged peptide segments are monitored. These disadvantages can be overcome by performing quantitation on intact protein or subunit analyte using a high resolution MS instrument that is capable of resolving isotopic peaks of intact subunit analyte or its large, multiply charged fragment ions. Ruan *et al.* reported protein bioanalysis using LC-MS of intact proteins on a first generation LTQ –Orbitrap Classic instrument and achieved good selectivity (*64*). Enhanced sensitivity and selectivity is expected with a more advanced Orbitrap mass spectrometer featuring higher resolution and faster scan speed.

The Orbitrap mass analyzer has become a powerful addition to the arsenal of high resolution LC-MS as one of the powerful tools used in protein drug discovery and development. Its analytical performance can support a wide range of applications from research to routine analysis, and from characterization to bioanalysis. The Orbitrap technology will continue to evolve not only toward increased acquisition speed, higher resolving power, mass accuracy and sensitivity, but also toward enhanced robustness and ease of operation. This evolution will undoubtedly allow the Orbitrap instruments penetrating into new areas of applications in research, development as well as in regulatory quality control environment.

Acknowledgments

The authors thank Andreas Huhmer, Rosa Viner and Eugen Damoc at Thermo Fisher Scientific, and Michael Boyne at US Food and Drug Administration for their critical review of this manuscript.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

References

1. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol* . **2010**, *10*, 345–352.
2. Beck, A. *mAbs* **2011**, *3*, 107–110.
3. Bondarenko, P. V.; Second, T. P.; Zabrouskov, V.; Makarov, A. A.; Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1415–1424.
4. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
5. Zhang, J.; Liu, H.; Katta, V. *J. Mass Spectrom.* **2010**, *45*, 112–120.
6. Heck, A. J. *Nat. Methods* **2008**, *5*, 927–933.
7. Atmanene, C.; Wagner-Rousset, E.; Malissard, M.; Chol, B.; Robert, A.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2009**, *81*, 6364–6373.
8. Rose, R. J.; Labrijn, A. F.; van den Bremer, E. T.; Loverix, S.; Lasters, I.; van Berkel, P. H.; van de Winkel, J. G.; Schuurman, J.; Parren, P. W.; Heck, A. J. *Structure* **2011**, *19*, 1274–1282.
9. Rosati, S.; Thompson, N. J.; Barendregt, A.; Hendriks, L. J.; Bakker, A. B.; de Kruif, J.; Throsby, M.; van Duijn, E.; Heck, A. J. *Anal. Chem.* **2012**, *84*, 7227–7232.
10. Chelius, D.; Xiao, G.; Nichols, A. C.; Vizel, A.; He, B.; Dillon, T. M.; Rehder, D. S.; Pipes, G. D.; Kraft, E.; Oroska, A.; Treuheit, M. J.; Bondarenko, P. V. *J. Pharm. Biomed. Anal.* **2008**, *47*, 285–294.
11. Lewis, D. A.; Guzzetta, A. W.; Hancock, W. S.; Costello, M. *Anal. Chem.* **1994**, *66*, 585–595.
12. Tao, L.; Ackerman, M.; Wu, W.; Liu, P.; Russell, R. Characterization of Impurities and Degradants in Protein Therapeutics by Mass Spectrometry. In *Characterization of Impurities and Degradants Using Mass Spectrometry*, John Wiley & Sons, Inc.: Hoboken, NJ, 2011; pp 391−426.
13. Chen, G.; Warrack, B. M.; Goodenough, A. K.; Wei, H.; Wang-Iverson, D. B.; Tymiak, A. A. *Drug Discovery today* **2011**, *16*, 58–64.
14. Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S. *J. Pharm. Biomed. Analysis* **2000**, *21*, 1099–1128.
15. Chelius, D.; Rehder, D. S.; Bondarenko, P. V. *Anal. Chem.* **2005**, *77*, 6004–6011.
16. McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; Kong, X.; Begley, T. P. *FEBS J.* **2007**, *274*, 6256–6268.
17. Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699–711.
18. Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109–112.
19. Ge, Y.; El-Naggar, M.; Sze, S. K.; Oh, H. B.; Begley, T. P.; McLafferty, F. W.; Boshoff, H.; Barry, C. E., 3rd. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 253–261.
20. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W. *J. Amer. Chem. Soc.* **2002**, *124*, 672–678.
21. Zhang, H.; Ge, Y. *Circ.: Cardiovasc. Genet.* **2011**, *4*, 711.
22. Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. *Biochem. Biophys. Res. Comm.* **2014**.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

23. Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 197A–203A.
24. Meng, F.; Forbes, A. J.; Miller, L. M.; Kelleher, N. L. *Mass Spectrom. Rev.* **2005**, *24*, 126–134.
25. Siuti, N.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 817–821.
26. Zhang, Z.; Shah, B. *Anal. Chem.* **2007**, *79*, 5723–5729.
27. Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A. *mAbs* **2013**, *5*, 699–710.
28. Ashton, D. S.; Beddell, C. R.; Cooper, D. J.; Craig, S. J.; Lines, A. C.; Oliver, R. W.; Smith, M. A. *Anal. Chem.* **1995**, *67*, 835–842.
29. Roberts, G. D.; Johnson, W. P.; Burman, S.; Anumula, K. R.; Carr, S. A. *Anal. Chem.* **1995**, *67*, 3613–3625.
30. Siegel, M. M.; Hollander, I. J.; Hamann, P. R.; James, J. P.; Hinman, L.; Smith, B. J.; Farnsworth, A. P.; Phipps, A.; King, D. J.; Karas, M.; Ingendoh, A. P. H.; Hillenkamp, F. *Anal. Chem.* **1991**, *63*, 2470–2481.
31. Bennett, K. L.; Smith, S. V.; Lambrecht, R. M.; Truscott, R. J.; Sheil, M. M. *Bioconj. Chem.* **1996**, *7*, 16–22.
32. Feng, R.; Konishi, Y. *Anal. Chem.* **1992**, *64*, 2090–2095.
33. Tito, M. A.; Miller, J.; Walker, N.; Griffin, K. F.; Williamson, E. D.; Despeyroux-Hill, D.; Titball, R. W.; Robinson, C. V. *Biophys. J.* **2001**, *81*, 3503–3509.
34. Dillon, T. M.; Bondarenko, P. V.; Speed Ricci, M. *J. Chromatogr. A* **2004**, *1053*, 299–305.
35. Gadgil, H. S.; Pipes, G. D.; Dillon, T. M.; Treuheit, M. J.; Bondarenko, P. V. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 867–872.
36. Wang, L.; Amphlett, G.; Lambert, J. M.; Blattler, W.; Zhang, W. *Pharm. Res.* **2005**, *22*, 1338–1349.
37. Guan, S.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 64–71.
38. Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. *J. Proteome Res.* **2007**, *6*, 114–123.
39. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4*, 2010–2021.
40. Zhang, J., Katta, V. In *Top-Down Characterization of Therapeutic Antibody by High Resolution LTQ Orbitrap Mass Spectrometry*, Proceedings of the 55th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN, June 3−7, 2007.
41. Zhang, J., Katta, V. In *High Resolution Orbitrap Mass Spectrometry: New Applications for Process Research and Development*, 12th Symposium on the Interface of Regulatory and Analytical Sciences for Biotechnology Health Products, Washington, DC, 2008.
42. Makarov, A.; Denisov, E. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1486–1495.
43. Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., Zhang, Z. In *Intact Mass and Top-Down HPLC/MS Analysis of Monoclonal IgG Antibodies on Orbitrap*, 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, 2008.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

44. Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. *Mol. Cell. Proteomics* **2011**, *10*, M111 011015.
45. Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Muller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J. J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. *Molecular & cellular proteomics : MCP* **2012**, *11*, O111 013698.
46. Shaw, J. B.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85*, 8313–8318.
47. Kaltashov, I. A.; Abzalimov, R. R. *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1239–1246.
48. Thompson, N. J.; Rosati, S.; Heck, A. J. *Methods* **2014**, *65*, 11–17.
49. Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. *Angew. Chem.* **2013**, *52*, 4020–4023.
50. Benesch, J. L.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544–3567.
51. Beck, A.; Reichert, J. M. *mAbs* **2013**, *5*, 621–623.
52. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2013**, *85*, 715–736.
53. Rosati, S.; van den Bremer, E. T.; Schuurman, J.; Parren, P. W.; Kamerling, J. P.; Heck, A. J. *mAbs* **2013**, *5*.
54. Debaene, F.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. *Anal. Chem.* **2013**, *85*, 9785–9792.
55. Garcia, B. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 193–202.
56. Benesch, J. L.; Robinson, C. V. *Curr. Opin. Struct. Biol.* **2006**, *16*, 245–251.
57. Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. *Mol. Cell. Proteomics* **2012**, *11*, 1758–1767.
58. Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. *Anal. Chem.* **2011**, *83*, 8919–8927.
59. Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B. *J. Proteome Res.* **2005**, *4*, 1155–1170.
60. Zhang, J.; Wu, S. L.; Kim, J.; Karger, B. L. *J. Chromatogr. A* **2007**, *1154*, 295–307.
61. Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Mol. Cell. Proteomics* **2007**, *6*, 1942–1951.
62. Hao, Z., Horn, D., Wu, S., Jiang, X., Huhmer, A. F., Bennett, P. K.; *A High Resolution Bench-Top Orbitrap LC-MS Workflow Solution for Comprehensive Intact Monoclonal Antibody Characterization*. In WCBP 2013: 17th Symposium on the Interface of Regulatory and Analytical Sciences for Biotechnology Health Products, Washington, DC, 2013.
63. Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. *Nat. Methods* **2012**, *9*, 1084–1086.
64. Ruan, Q.; Ji, Q. C.; Arnold, M. E.; Humphreys, W. G.; Zhu, M. *Anal. Chem.* **2011**, *83*, 8937–8944.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch010

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Chapter 11

Rapid Middle-Down Sequence Determination of Antibodies by MALDI In-Source Decay MS

Sergei Dikler* **and Amanda L. Bulman**

Bruker Daltonics, Inc., Billerica, Massachusetts 01821, United States
***E-mail: sergei.dikler@bruker.com**

This chapter describes the analysis of an intact antibody using two middle-down sequencing approaches based on MALDI in-source decay. Simple disulfide bond reduction permitted the rapid confirmation of 47% of the antibody sequence and the detection of two terminal sequence modifications in the heavy chain subunit. A second approach utilizing a hinge region specific enzyme followed by disulfide bond reduction permitted confirmation of 72% of the antibody sequence. MALDI in-source decay can provide a time-efficient and robust way to confirm both N and C terminal sequences in the emerging field of biosimilar antibody characterization.

Introduction

MALDI-TOF MS has proven to be valuable for characterization of peptide and protein targets. Intact mass measurements may be quickly and reliably made via MALDI-TOF with minimal sample requirements and preparation steps, and the analysis of in-source decay (ISD) fragmentation patterns has proven to well-complement extensive digest-based TOF and TOF/TOF bottom up workflows (*1–5*). The determination of peptide amino acid sequences by MALDI-TOF/TOF relies on the generation of predominantly a, b, and y fragment ions resulting from metastable ion decay (collision or laser induced) occurring after ions exit the ionization source. In contrast, the ISD top-down approach to protein sequencing is based on the simultaneous creation of singly charged fragment ions from the protein termini of intact targets within the MALDI ion source. Modern MALDI sources employ a precisely tunable extraction delay, that is, a separation in time between the ionization event and ion acceleration,

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

which may be optimized to provide improved resolution for a given mass range in standard reflector MS and MS/MS acquisition modes. In-source decay fragmentation occurs within the ion source between the target surface and extraction plate, and the extraction delay time can be increased to improve the detection and resolution of ISD ions beyond the normal range used for peptide analyses. Upon ion extraction, sequence tags may be observed from approximately the 10th residue from the C- and/or N terminus. This uniform fragmentation along the protein backbone is evidenced by predominantly c, y, and (z+2) fragment ions, with preservation of post-translational modifications, including phosphorylation, methylation, and glycosylation (*6*, *7*). N-terminal blockage or modifications, such as acetyl or pyroglutamyl groups, do not limit in-source decay, and both N and C terminal fragments may be observed, which gives the method significant advantages over traditional Edman Sequencing. The technique is amenable to both linear and reflector modes of collection, however, the improvement in mass accuracy in reflector mode is critical for many sequencing applications. ISD is also valuable for the study of heterogeneous biomolecules, such as recombinant fusion proteins and PEGylated products. In the latter case, site(s) of PEGylation may be determined/confirmed without any sample pretreatment, and oligomers may also be characterized within the same preparation (*8*).

This top-down sequencing approach has been generally applied to pure or nearly pure proteins. As there is no selection of precursor ions in ISD, the top-down analysis of complex mixtures could result in very complex spectra due to the simultaneous creation of multiple terminal residue strings in the same mass range. Protein samples with several components, while complex, are interpretable if the expected sequences are known. In protein QC applications, database-driven, automated comparisons to expected protein sequences can be used for high throughput screening (*9*). Sequencing via in-source decay has also been applied to peptides (*1*), tissue (*10*, *11*) and oligonucleotides (*12*, *13*).

The protocols for sample preparation for in-source decay are straightforward, and informative data can often be collected in only a few seconds. The choice of matrix and matrix-to-sample ratios are key elements, as the proposed mechanism of fragmentation stems from hydrogen radical transfer from the matrix (*14–16*). 2,5-Dihydroxybenzoic acid (DHB) and 90% 2,5-dihydroxybenzoic acid, 10% 2-hydroxy-5-methoxybenzoic acid (sDHB) have been routinely used for ISD, but utility has been shown for other matrices as well, including 5-aminosalicylic acid (5-ASA), 5-formylsalicylic acid (5-FSA), 5-nitrosalicylic acid (5-NSA), 2-aminobenzamide (2-AB), and 2-aminobenzoic acid (2-AA) (*17–19*). 1,5-Diaminonaphthalene (DAN) has been shown to be valuable for direct reduction of disulfide bonds in peptide and protein targets (*20*). The benefit of additives has been shown, including picolinic acid (*15*) and ammonium persulfate (*21*).

Regardless of the choice of matrix, detection of the extreme terminal residues is often challenging due to prominent overlapping matrix signals, up to approximately 800 Da. If a sequence is known, as is the case in many product characterization applications, indirect assignments of extreme residues may be reported. Deviations from expected sequences (via truncations or modifications)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

result in reproducible mass shifts, which may also be indirectly assigned. For unambiguous sequence determination, single ISD fragment ions may be isolated using a timed ion selector, and further fragmented via (subsequent) metastable ion decay. The detection and utilization of these granddaughter ions (typically a, b, and y ions) is known as pseudo-MS^3 or T^3-Sequencing (*2*, *22*, *23*). Sequence coverage may be limited compared to a direct MALDI-TOF/TOF analysis of peptides, however.

The analysis of protein therapeutics, including monoclonal antibodies (mAbs), has proven to be a good match for the ISD sequencing approach. Given their size and complex structure, often a middle-down approach is advantageous, defined here as MS analysis following cleavage into a few large fragments by reducing disulfide bonds or by limited proteolysis (*24*). Many published works have also utilized LC-purification of chain subunits, as extensive target characterization efforts frequently benefit from such an approach (*25*, *26*). An immunoglobulin–degrading enzyme of *Streptococcus pyogenes* (IdeS), which specifically cleaves in the IgG hinge region, has recently demonstrated to be very useful in the middle down MS analyses of antibody forms on many MS platforms (*27–30*).

As both primary sequence elements and post-translational modifications can dramatically alter the behavior of therapeutic antibodies, confirmation of desired structure(s) or elucidation of variants is of great value. Protein sequencing via MALDI in-source decay can provide a rapid and robust means to characterize targets in detail and may be amenable to many proteomic workflows. This chapter focuses on rapid middle-down analyses of a purified IgG1 mAb via in-source decay.

Materials and Methods

Sample Preparation

NISTmAb sample (purified IgG1 monoclonal antibody) was provided for this study. The sample concentration was 10 mg/mL, in 12.5 mM L-His and 12.5 mM L-His HCl (pH 6.0).

2,5-Dihydroxyacetophenone (DHAP, Bruker Daltonics) was used for intact subunit and intact fragment mass measurements. The DHAP solution was prepared from 7.6 mg of DHAP in 375 µl ethanol and 125 µl of an 18 mg/ml diammonium hydrogen citrate solution. The diammonium hydrogen citrate was prepared fresh from solid (Sigma Aldrich). The samples were mixed with 2% TFA solution and DHAP solution in a 1:1:1 ratio. The mixture was pipetted up and down until it turned opaque, at which time it was spotted onto the stainless steel target. Protein Calibration Standard II (Bruker Daltonics) was used for mass calibrations.

To reduce disulfide bonds, dithiothreitol (DTT, Sigma-Aldrich) was added to 25 mM and incubated with the mAb sample at 50°C for 30 minutes. The reduced sample was subjected to a buffer exchange into 0.1% TFA via C4 ZipTip™ (Millipore). The sample was applied to a stainless steel target as a dried droplet in a 2:1 ratio (matrix to sample). Samples were prepared in the matrices 90% 2,5-

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

dihydroxybenzoic acid, 10% 2-hydroxy-5-methoxybenzoic acid (sDHB, Bruker Daltonics) for ISD sequencing and 1,5-diaminonaphthalene (DAN, Alfa Aesar) for T^3-sequencing. Both matrices were prepared in 50% acetonitrile, 0.1% TFA. sDHB was prepared at 50 mg/mL, and the DAN was saturated. Bovine Serum Albumin (BSA, Calbiochem) was prepared at 100 pmol/μL and applied in sDHB for calibration of ISD spectra.

The mAb solution (10 μL aliquot) was added to a sealed microcentrifuge tube containing 100 units of lyophilized modified IdeS protease (FabRICATOR; Genovis, Lund, Sweden) and incubated for 30 minutes at 37 °C. A portion of the cleaved sample was reserved for analysis, with the remainder being subjected to disulfide bond reduction as described above. IdeS cleaved and cleaved/reduced samples were enriched with C4 ZipTips then prepared in DHAP for intact mass measurement and sDHB for ISD as described above.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Data Collection

All MALDI mass spectra were acquired using an ultrafleXtreme MALDI-TOF/TOF mass spectrometer equipped with smartbeam II laser (Bruker Daltonics), using a 1 kHz acquisition speed in positive ion mode. Compass 1.4 software (flexControl 3.4, flexAnalysis 3.4) was used for data acquisition and processing. Intact mass determinations were made in linear mode, with approximately 1000-2000 laser shots collected per spectrum.

In-source decay data was collected in reflector mode, with approximately 15000-30000 laser shots per spectrum. Instrument settings were as follows: matrix suppression, 900 Da; pulsed ion extraction delay, 170 ns; acceleration voltage (IS1), 25 kV; extraction voltage (IS2), 22.4 kV; lens voltage, 7.75 kV; reflector voltage, 26.45 kV; reflector 2 voltage 13.45 kV. The Sophisticated Numerical Annotation Procedure (SNAP) algorithm was used for monoisotopic peak annotation with signal-to-noise threshold set to 2 (*31*). BioTools 3.2 SR4 software (Bruker Daltonics) was used for subunit and middle-down sequence analysis. Sequence tags representing the amino acid sequence of the protein in the ISD spectra were created within BioTools. The Sequence Editor functionality was used to fit the data to known sequences and evaluate differences between the expected and observed masses. Matching a, c, y and (z+2) ions were automatically assigned within the ISD spectra. The average mass error in the ISD spectra was 18 ppm (0.08 Da), calculated for all assigned fragment ions in BioTools software.

The T^3-sequencing (pseudo-MS^3) spectra were acquired in MS/MS mode by selecting abundant ISD fragment ions as parent ions and collecting 15000 laser shots. Following smoothing and baseline subtraction, the spectra were matched to the presumed sequences using Sequence Editor.

Results and Discussion

In this work we have used two middle-down sequencing approaches in order to obtain maximum sequence coverage of the monoclonal antibody. The first approach was focused on MALDI-ISD analysis of heavy chain and light chain

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

after the antibody was reduced, without any prior separation. This approach will be referred to as subunit analysis (Figure 1A).

In the second approach the antibody was digested with IdeS protease, producing $F(ab')_2$ fragment and two single chain Fc fragments (scFc), followed by reduction. This treatment results in single chain Fc fragment, Fab heavy chain fragment (Fd) and light chain. This second approach was focused on the heavy chain domains with no prior LC separation and will be referred to as middle-down heavy chain domain analysis (Figure 1B).

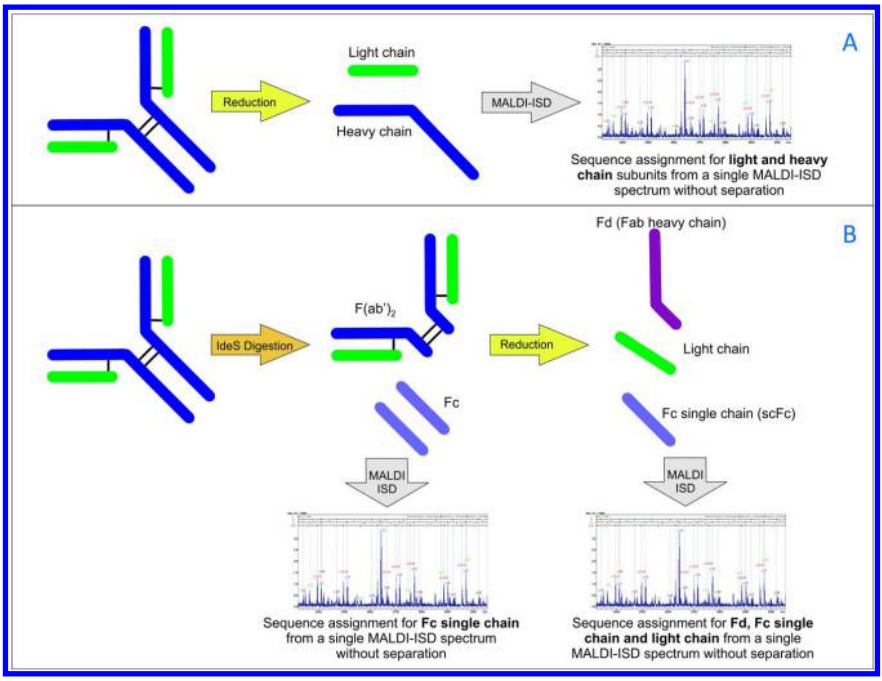

Figure 1. The subunit analysis workflow (reduction only) (A) and the middle-down heavy chain domain analysis workflow (digestion with IdeS and reduction) (B). (see color insert)

Intact Subunit and Intact Fragment Mass Analysis

Sample preparation steps for both subunit and middle-down sequencing approaches were monitored by MALDI-TOF in linear mode. The mass range of particular interest in these measurements was 22000-28000 Da since it encompasses the reaction products including the light chain, Fc single chain fragment (scFc), and Fd fragment (Fab heavy chain), as shown in Figure 2. This analysis simultaneously provides information on the intact fragment masses and therefore whether they are consistent with the expected primary sequence.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

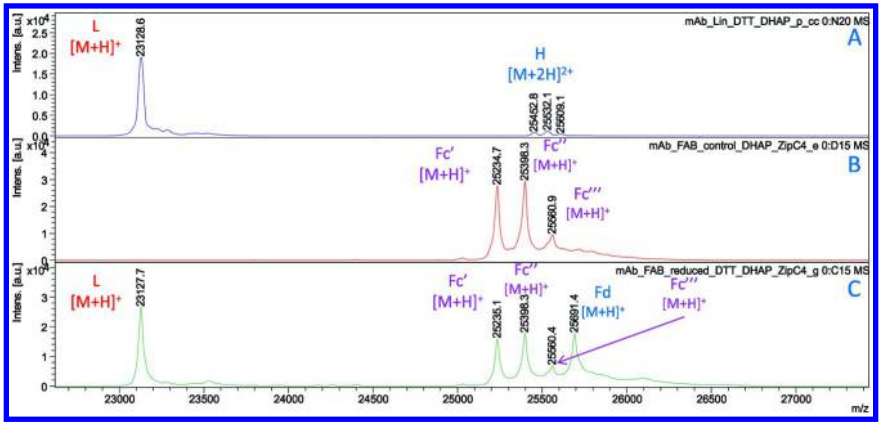

Figure 2. MALDI-TOF spectra in linear mode of mAb reduced with DTT (A), mAb digested with IdeS before reduction (B) and mAb digested with IdeS and reduced with DTT (C). Simple reduction produces light chain (L) and heavy chain (H) subunits (A). Digestion with IdeS protease produces three glycoforms of Fc single chain fragment (Fc′, Fc″ and Fc‴)(B). Digestion with IdeS protease followed by reduction generates L and Fd fragment (Fab heavy chain) in addition to the glycoforms of Fc fragment. (see color insert)

In the subunit analysis, the reduction with DTT resulted in the quantitative conversion to light chain and heavy chain subunits with three glycoforms of doubly charged heavy chain (Figure 2A). The measured mass of the light chain subunit at *m/z* 23128.6 matched exactly the calculated average mass of singly charged protonated ion based on the presumed sequence at *m/z* 23128.6.

The middle down heavy chain domain analysis had two steps. The first step was digestion with a highly specific IdeS protease, which cleaves only between Gly239 and Gly240 residues in the hinge region of the heavy chain. This cleavage site is below the locations of the two interchain heavy-heavy disulfide bonds (Cys229 and Cys232) and one interchain heavy-light disulfide bond (Cys223). Thus, this cleavage created an $F(ab')_2$ fragment connected via interchain disulfide bonds and two Fc single chain fragments. This treatment allowed exclusive sequencing of the Fc single chain fragment, since $F(ab')_2$ with multiple interchain disulfide bonds is not prone to ISD fragmentation. Effectively, this eliminated the need for LC separation in order to isolate the Fc single chain fragment. Figure 2B shows three glycoforms of single chain Fc fragment (Fc′, Fc″ and Fc‴), with the 162 Da mass differences between the forms corresponding to hexose residues. The measured average masses of these glycoforms at *m/z* 25234.7 (Fc′), 25398.3 (Fc″) and 25560.9 (Fc‴) were all in good agreement with the calculated average masses for the NISTmAb with glycan compositions of G0F (calculated *m/z* 25237.0; mass error -2.3 Da), G1F (calculated *m/z* 25399.2; mass error -0.9 Da), and G2F (calculated *m/z* 25561.3; mass error -0.4 Da) determined to be the dominant glycoforms in the Glycosylation chapter/Volume 2, Chapter 4.

The second step in the middle down heavy chain domain analysis was reduction with DTT, which added the light chain subunit and Fd fragment (Fab

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

heavy chain) to the three glycoforms of Fc. The measured mass of the light chain in this step was at *m/z* 23127.7 (mass error -0.9 Da) and Fd fragment measured mass was at *m/z* 25691.4 (calculated *m/z* 25689.9; mass error 1.5 Da) (Figure 2C). The measured mass of Fd fragment corresponds to the conversion of heavy chain Gln1 to pGlu1 discussed in detail in the "Sequence modifications of the heavy chain" section. The average mass error in linear mode measurements of intact subunits and fragments in the 22000-28000 Da mass range was 1.4 Da.

Subunit Sequencing

Basic Sequence Assignment Based on the Presumed Sequence of the Light Chain

The interpretation of the MALDI-ISD spectrum of reduced antibody began with the presumed sequence of the light chain. This sequence was an excellent match for the data, allowing assignment of 79 residues from the N-terminus (using mainly c-type fragment ions) and 86 residues from the C-terminus (using equally y-type and (z+2)-type fragment ions) (Table 1 and Figure 3). There was very good sequence coverage by c, y and (z+2) fragment ions over the mass range 1000-7000 Da, with a few short gaps largely corresponding to the location of proline residues. The cyclic structure of proline prevents formation of c and (z+2) fragment ions and gaps often correspond to a XP sequence motif (*2*).

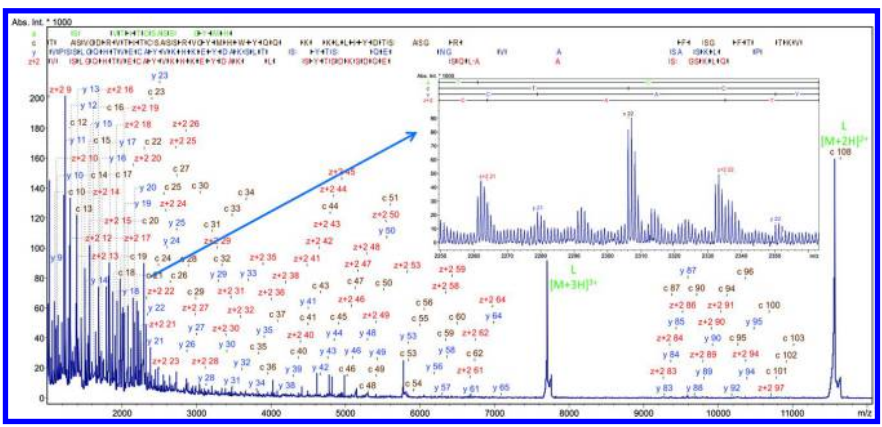

Figure 3. MALDI-ISD spectrum of mAb (reduced with DTT and prepared with sDHB matrix) with the sequence assignment for the light chain based on the presumed sequence. Inset shows c_{22} ISD fragment ion later selected for T^3 measurement. (see color insert)

The ISD spectrum had a large gap between 7000 and 9000 Da followed by more lower intensity c, y and (z+2) fragment ions in the 9000-11000 Da range (Figure 3). One possible explanation is that the very abundant triply charged and doubly charged parent ions of the light chain at *m/z* 7710 and 11564 suppress fragment ion intensities in the immediate vicinity.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Table 1. MALDI-ISD Sequence Coverage

	Sample	*Chain*	*N-terminal residues*	*C-terminal residues*	*Total number of confirmed residues*	*Sequence coverage %*[a]
1	mAb reduced with DTT	Light chain	79	86	165	77.5
		Heavy chain	71	76	147	32.7
2	mAb digested with IdeS and reduced	Light chain	76	91	167	78.4
		Fd	64	81	145	60.7
		Fc	60	87	147	70.0
3	mAb digested with IdeS (no reduction)	Fc	60	106	166	79.0
Total 1	**Subunit analysis (reduction only)**	**H+L**	**150**	**162**	**312**	**47.1**
Total 2	**Middle-down heavy chain domain analysis (IdeS + reduction)**	**L+Fd+Fc**	**200**	**278**	**478**	**72.2**

[a] MALDI-ISD sequence coverage includes sequence coverage obtained from T^3-sequencing.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Sequence Modifications of the Heavy Chain

The next logical step in the interpretation of the MALDI-ISD spectrum of the reduced antibody from the subunit analysis approach was to match the presumed heavy chain sequence. The presumed heavy chain sequence provided a very poor match, indicating modification(s) and/or amino acid residue deletion/substitution on both N and C termini (Figure 4A). The N-terminus of the heavy chain was established by using sequence tag analysis within the BioTools software. This software tool allowed the exclusion of previously assigned fragment ions corresponding to the light chain and considered any proline gaps. The sequence tag KPTQT matched the presumed sequence residues 13-17, but the starting mass was 17 Da lower than in the presumed sequence (Figure 4A inset). The 17 Da mass difference and the fact that the first residue in the heavy chain is glutamine indicated a transformation to pyroglutamic acid. The modified heavy chain with pGlu1 matched N-terminal fragmentation in the ISD spectrum (Figure 4B).

The last two residues of the presumed sequence of the heavy chain, Gly449 and Lys450, correspond to the secretory tail. It was logical to assume that the most likely modification of the C-terminus would be deletion of one or more residues. Matching the shortened sequence [1-449] with deletion of Lys450 provided an excellent fit to the ISD data, while both full sequence and sequence [1-448] did not match (Figure 4B).

It was reasonable to expect some low degree of overlap in fragment ion assignments due to isobaric signals. The MALDI-ISD spectrum of reduced antibody had slightly more than 600 labeled signals with 341 signals assigned to the sequences of light and heavy chain subunits. The detailed analysis of the ISD spectrum showed 6 isobaric fragment ions (3 pairs) in the light chain sequence in the mass range 1000-12000 Da, which comprised 3.2% of all 188 light chain sequence assignments. One example of such an isobaric fragment ion pair was the signal at *m/z* 2047.044, assigned as light chain y_{19} (mass error 20 ppm) and a_{20} (mass error 16 ppm) (Figure 5A). There were 8 isobaric fragment ion assignments (4 pairs) accounting for 5.2% of all 153 heavy chain sequence assignments. The number of isobaric fragment ion assignments with one of the assignments corresponding to the light chain and the other to the heavy chain was 12 (6 pairs) or 3.5% of 341 signals assigned to both light and heavy chain sequences. This analysis allowed us to conclude that the number of isobaric fragment ion assignments was very low and did not interfere with the sequence determination goals (Figure 5). Moreover, the percent of isobaric fragment ion assignments spread between the light and heavy chain sequences is very similar to the percent of isobaric fragment ion assignments within the light or the heavy chain sequences.

The actual sequence of the heavy chain differs from the presumed sequence by pGlu1 rather than Gln1 at the N-terminus and deletion of Lys450 at the C-terminus (Figure 6). Both conversion of N-terminal glutamine to pyroglutamic acid and deletion of the C-terminal lysine are fairly common in the heavy chains of antibodies (*28*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

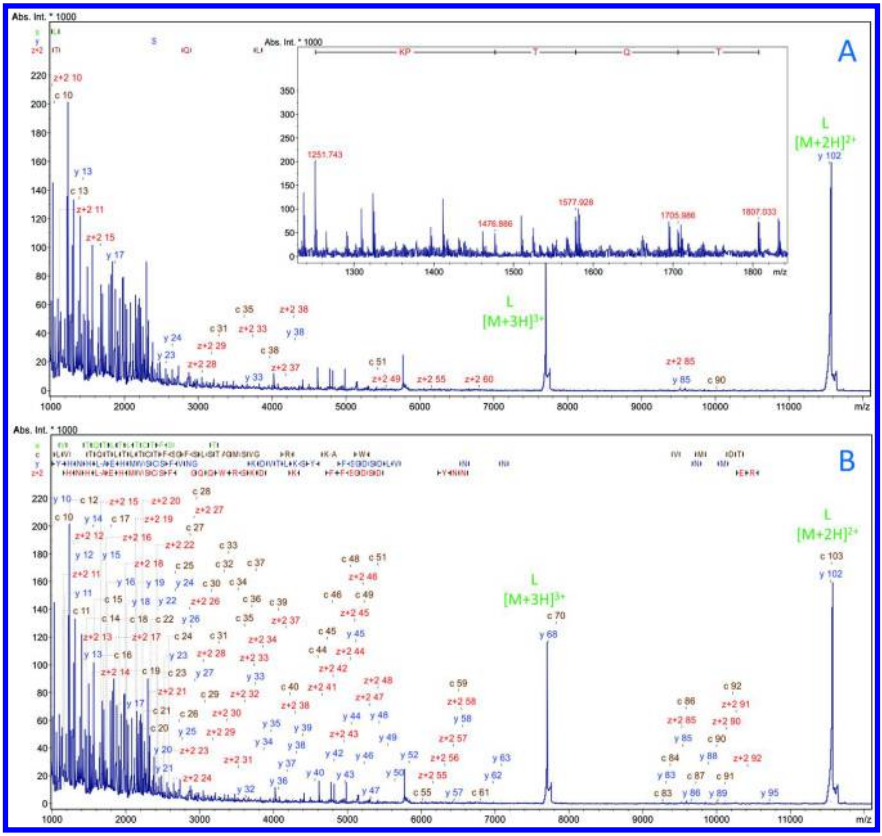

Figure 4. MALDI-ISD spectrum of mAb with the assignment for the heavy chain based on the presumed sequence (A) and the actual sequence (B). Poor match of the presumed sequence indicates modifications of both N- and C-termini. The inset (A) shows sequence tag KPTQT that helped in finding the mass shift corresponding to transformation of Gln1 residue to pGlu. The spectrum analysis also revealed deletion of Lys450 residue. (see color insert)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

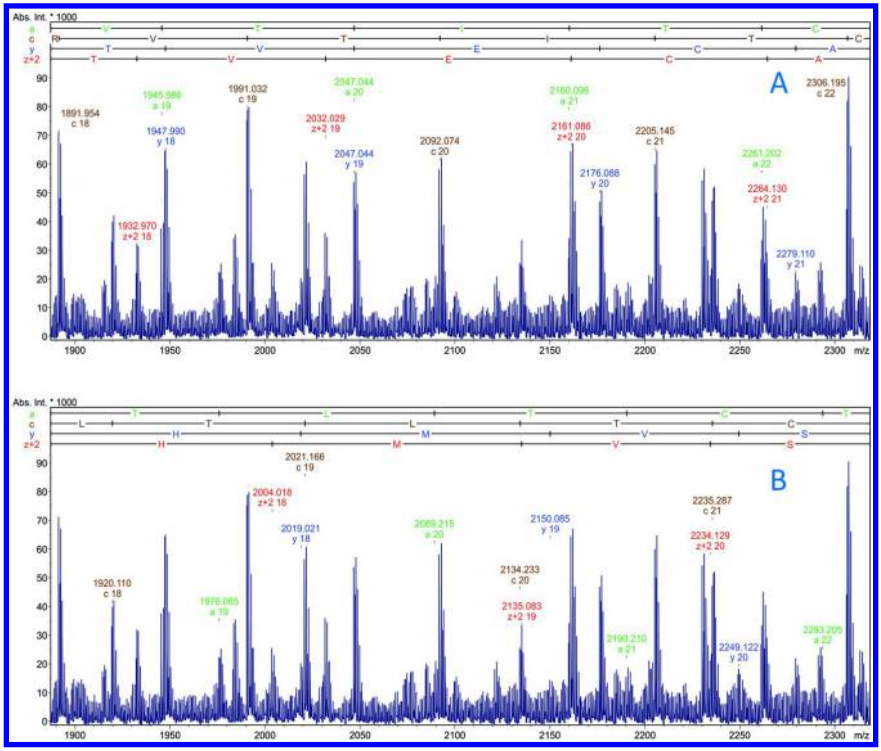

Figure 5. Comparison of sequence assignments for the light chain (A) and the heavy chain (B) in 1900-2300 Da mass region of MALDI-ISD spectrum of reduced mAb shown in Figures 3-4. (see color insert)

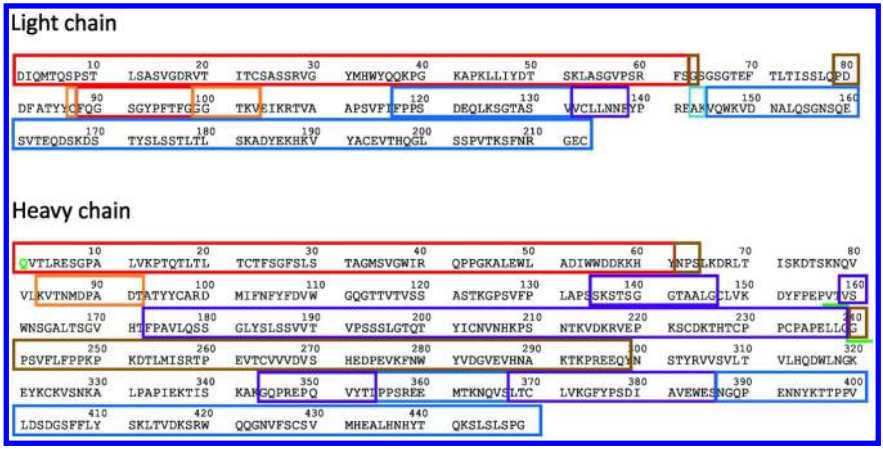

Figure 6. The sequence coverage of antibody heavy and light chains using the subunit analysis approach (reduction only) and the middle-down heavy chain domain analysis approach (digestion with IdeS and reduction). IdeS protease cleavage site between Gly239 and Gly240 is shown in green. N-terminal sequence coverage by both approaches is shown in red, by the middle-down heavy chain domain analysis approach only in brown and by the subunit analysis approach only in orange. C-terminal sequence coverage by both approaches is shown in blue, by the middle-down heavy chain domain analysis approach only in dark purple and by the subunit analysis approach only in cyan. In the heavy chain sequence Gln1 (highlighted in green) is converted to pGlu and Lys450 undergone deletion. (see color insert)

71 N-terminal residues and 76 C-terminal residues from the actual sequence of the heavy chain were assigned. Overall, the combined assignments for light chain and heavy chain covered 150 N-terminal residues and 162 C-terminal residues. The total sequence coverage of reduced antibody from a single MALDI-ISD spectrum with no LC separation prior to analysis was 47.1% using the subunit analysis approach.

Middle-Down Heavy Chain Domain Sequencing

In the first step of the middle down heavy chain domain analysis, proteolytic cleavage with IdeS protease generated only one fragment amenable to MALDI-ISD fragmentation, the Fc single chain fragment, as the second fragment $F(ab')_2$ with multiple interchain disulfide bonds does not fragment well (Figure 1B). It was previously reported that interchain disulfide bonds limit ISD fragmentation (*32*). Multiple interchain disulfide bonds in $F(ab')_2$ fragment linking two Fd fragments and two light chains likely created steric hindrances and prevented any substantial ISD fragmentation as evidenced by the lack of sequence coverage when the sequences of the light chain and the Fd fragment were matched to the ISD spectrum of antibody cleaved with IdeS.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Interpretation of the MALDI-ISD spectrum of antibody cleaved with IdeS provided an excellent fit to the sequence of Fc single chain fragment (heavy chain [240-449]). N-terminal sequence coverage extended to residue 60 and terminated there as residue 61 is asparagine (Asn300 of the intact heavy chain), which is the glycosylation site (Figures 6, 7). Multiple glycoforms and relatively low fragment ion abundance of c ions in the 6000-9000 Da range made further assignments challenging.

Three glycoforms were detected in the intact subunit analysis. More detailed analyses of intact glycoforms can be accomplished using a high resolution ESI instrument. Alternatively, a bottom-up approach focusing on glycopeptides and using either LC-MS/MS on a high resolution QTOF analyzer or LC-MALDI on a TOF/TOF analyzer may be utilized (*28*). Glycoform analysis is beyond the scope of this chapter, however.

C-terminal fragmentation covered 106 residues extending to 12000 Da with relatively few short gaps due to proline residues and in the immediate vicinity of triply charged signals of intact Fc′, Fc″ and Fc‴ glycoforms (Figure 7). This was the highest number of residues from a single terminus achieved in this work since in this approach the Fc single chain fragment is the only fragmenting component at the IdeS cleavage step. The total sequence coverage of Fc single chain fragment was 166 residues or 79.0% (Table 1).

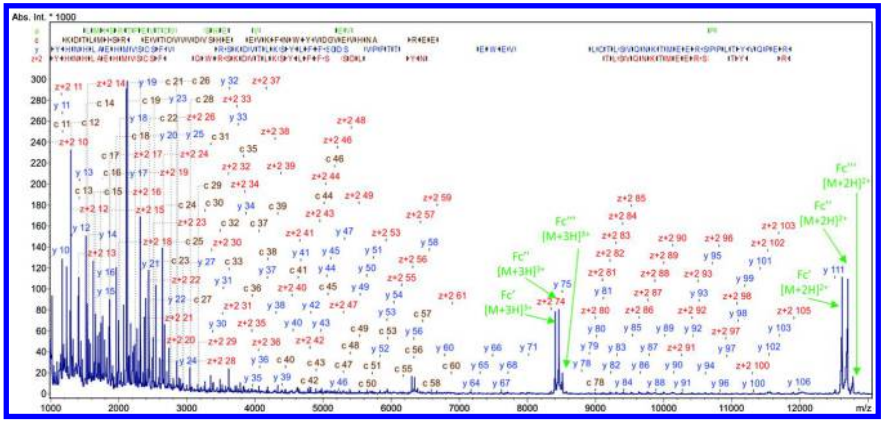

Figure 7. MALDI-ISD spectrum of mAb digested with IdeS protease before the reduction step. The sequence assignment is for Fc fragment of the heavy chain. (see color insert)

The reduction step after IdeS cleavage added the light chain subunit and the Fd fragment to the mixture containing the Fc single chain fragment. The interpretation of the MALDI-ISD spectrum began with the sequence of the Fd fragment and provided an excellent fit for the data. N-terminal coverage was 64 residues with mainly c-type fragment ions and a few a-type fragment ions, which generally have lower abundance than c ions and cover the mass range 1000-3500 Da. C-terminal coverage was 81 residues almost equally distributed between y and (z+2) ions, with a large gap between 7000 and 10000 Da where the triply charged ions of intact light chain, Fc′, Fc″, Fc‴ and Fd are situated (Figure 8).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

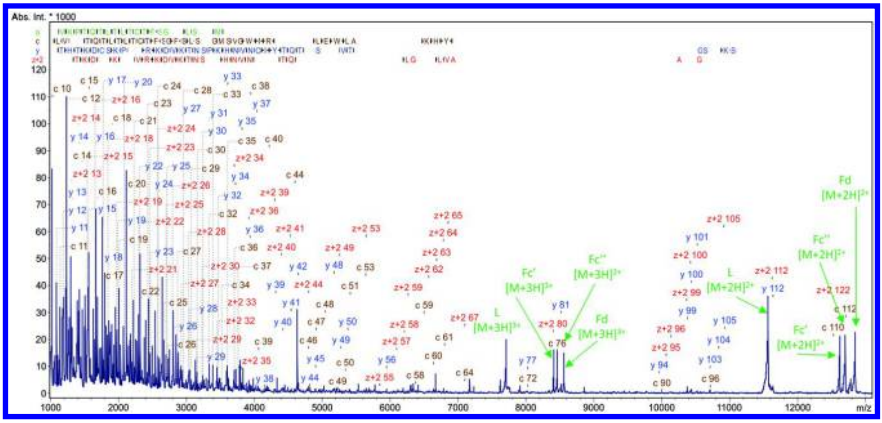

Figure 8. MALDI-ISD spectrum of mAb digested with IdeS protease and reduced with DTT. The sequence assignment is for Fd fragment (Fab heavy chain). (see color insert)

As expected, the sequence coverage of light chain in the middle-down heavy chain domain analysis approach was nearly identical to the subunit analysis approach. The total number of residues covered was 167 in the middle-down heavy chain domain analysis versus 165 in the subunit analysis (Table 1).

The sequence coverage of Fc single chain fragment was lower by 9.0% after the reduction step compared to cleavage with IdeS prior to reduction. This could be attributed to the presence of other fragmenting subunits (Fd and light chain) in the mixture, which partially suppress Fc fragment ion abundances. The C-terminal coverage dropped to 87 residues versus 106 residues while N-terminal sequence coverage was identical (Table 1).

Comparison of the middle-down heavy chain domain analysis approach (IdeS cleavage and reduction) with the subunit analysis approach (reduction only) showed a 33% increase in N-terminal coverage with 200 N-terminal residues covered versus 150. C-terminal coverage increased dramatically by 72% using the middle-down heavy chain domain approach compared to subunit analysis, with 278 C-terminal residues versus 162. The overall sequence coverage improvement was 25.1% with 72.2% sequence coverage in the middle-down heavy chain domain analysis compared to 47.1% in the subunit analysis (Table 1).

Sequence Confirmation Using T^3 Spectra

T^3 (pseudo-MS^3) spectra provide a very useful tool to further confirm the first few and the last few residues in the sequence. As previously discussed, in MALDI-ISD spectra the mass range below 1000 Da overlaps with signals from the matrix, matrix adducts and matrix clusters and thus very low molecular weight ISD fragment ions are more difficult to find. Only relatively high abundance ISD fragment ions are suitable for T^3 sequencing, and the sequence coverage is usually lower than in conventional MALDI-MS/MS measurements of peptides.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

A representative T^3 spectrum of c_{22} ISD fragment ion at *m/z* 2306.166 from light chain is shown in Figure 9. Complimentary b and y fragment ion series provided good coverage and additional confirmation for the first 22 residues of the N-terminus of light chain.

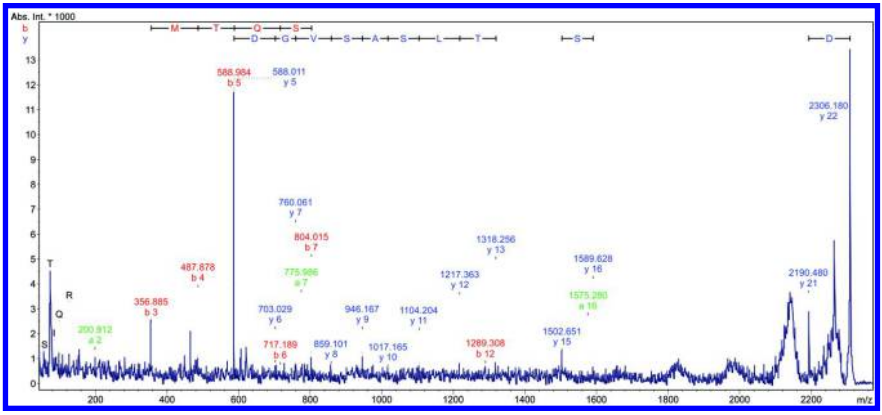

Figure 9. T^3 (pseudo-MS^3) spectrum of c_{22} ISD fragment ion at m/z 2306.166 from the light chain of antibody (sample was prepared with DAN matrix). (see color insert)

Conclusions

The rapid growth of therapeutics based on monoclonal antibodies and the emergence of biosimilar antibodies requires rapid sequence characterization methods that can verify that an antibody of interest is identical to a reference sequence. Identical amino acid sequence for a biosimilar antibody when compared to a reference antibody product is one of the requirements imposed by regulatory agencies.

Two middle-down sequencing approaches based on MALDI-ISD described in this chapter provide an alternative to previously published bottom-up proteomics approaches and top-down sequencing of intact IgGs. The subunit analysis approach (reduction only) required only 30 minutes of sample preparation and no LC separation, resulting in 47% sequence coverage of the antibody. The middle-down heavy chain domain analysis approach (IdeS cleavage and reduction) dramatically increased sequence coverage to 72% with a modest increase in sample preparation time (to 1 hour) and no prior separation. Subunit analysis allowed the detection of two terminal modification of the heavy chain sequence (conversion of Gln1 to pGlu and deletion of Lys450). The highest sequence coverage from a single terminus was achieved for the Fc single chain fragment with 106 residues from the C-terminus in the middle down heavy chain domain analysis.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Acknowledgments

The authors thank Wolfgang Jabs and Anja Resemann for helpful discussion and Matthew Willetts for reviewing the manuscript.

References

1. Brown, R. S.; Lennon, J. L. *Anal. Chem.* **1995**, *67*, 3990–3999.
2. Suckau, D.; Resemann, A. *Anal. Chem.* **2003**, *75*, 5817–5824.
3. Hardouin, J. *Mass Spectrom. Rev.* **2007**, *26*, 672–682.
4. Suckau, D.; Cornett, D. *S. Analusis* **1998**, *26*, 18–21.
5. Franzen, J. U.S. Patent 8,581,179, 2013.
6. Lennon, J. J.; Walsh, K. A. *Protein Sci.* **1999**, *8*, 2487–2493.
7. Hanisch, F. G. *Anal. Chem.* **2011**, *83*, 4829–4837.
8. Yoo, C.; Suckau, D.; Sauerland, V.; Ronk, M.; Ma, M. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 326–333.
9. Suckau, D.; Resemann, A. *J. Biomol. Technol.* **2009**, *20*, 258–262.
10. Debois, D.; Bertrand, V.; Quinton, L.; De Pauw-Gillet, M. C.; De Pauw, E. *Anal. Chem.* **2010**, *82*, 4036–4045.
11. Calligaris, D.; Longuespée, R.; Debois, D.; Asakawa, D.; Turtoi, A.; Castronovo, V.; Noël, A.; Bertrand, V.; De Pauw-Gillet, M. C.; De Pauw, E. *Anal. Chem.* **2013**, *85*, 2117–2126.
12. Hagan, N.; Smith, C. A.; Antoine, M. D.; Lin, J. S.; Feldman, A. B.; Demirev, P. A. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 773–777.
13. Kirpekar, F.; Nordhoff, E.; Larsen, L. K.; Kristiansen, K.; Roepstorff, P.; Hillenkamp, F. *Nuc. Acids Res.* **1998**, *26*, 2554–2559.
14. Kocher, T.; Engstrom, Å.; Zubarev, R. A. *Anal. Chem.* **2005**, *77*, 172–177.
15. Demeure, K.; Quinton, L.; Gabelica, V.; De Pauw, E. *Anal. Chem.* **2007**, *79*, 8678–8685.
16. Takayama, M. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1044–1049.
17. Sakakura, M.; Takayama, M. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 979–988.
18. Asakawa, D.; Sakakura, M.; Takayama, M. *Mass Spectrom. (Tokyo)* **2012**, *1*, A0002.
19. Smargiasso, N.; Quinton, L.; De Pauw, E. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 469–474.
20. Fukuyama, Y.; Iwamoto, S.; Tanaka, K. *J. Mass Spectrom.* **2006**, *41*, 191–201.
21. Horvatić, A.; Dodig, I.; Vuletić, T.; Pavoković, D.; Hameršak, Z.; Butorac, A.; Cindrić, M. *Anal. Chem.* **2013**, *858*, 3940–3947.
22. Suckau, D.; Resemann, A. U.S. Patent 7,396,686, 2004.
23. Asakawa, D.; Smargiasso, N.; De Pauw, E. *Anal. Chem.* **2014**, *86*, 2451–2457.
24. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
25. Resemann, A.; Wunderlich, D.; Rothbauer, U.; Warscheid, B.; Leonhardt, H.; Fuchser, J.; Kuhlmann, K.; Suckau, D. *Anal. Chem.* **2010**, *82*, 3283–92.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

26. Damen, C. W.; Chen, W.; Chakraborty, A. B.; van Oosterhout, M.; Mazzeo, J. R.; Gebler, J. C.; Schellens, J. H.; Rosing, H.; Beijnen, J. H. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2021–2033.
27. von Pawel-Rammingen, U.; Johansson, B. P.; Björck, L. *EMBO J.* **2002**, *21*, 1607–15.
28. Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A. *mAbs* **2013**, *5*, 699–710.
29. Wang, B.; Gucinski, A. C.; Keire, D. A.; Buhse, L. F.; Boyne, M. T., 2nd *Analyst* **2013**, *138*, 3058–3065.
30. Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86*, 3005–3012.
31. Köster, C. U.S. Patent 6,188,064, 2001.
32. Katta, V.; Chow, D. T.; Rohde, M. F. *Anal. Chem.* **1998**, *70*, 4410–4416.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch011

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Chapter 12

Automated, Online Sample Preparation for LC-MS Analyses: Affinity Capture, Digestion, and Clean-Up

David R. Colquhoun and Brian J. Feild*

Shimadzu Scientific Instruments, 7102 Riverwood Drive, Columbia, Maryland 21046, United States
***E-mail: bjfeild@shimadzu.com**

Modern mass spectrometry techniques for the analysis of proteins requires complementary sample preparation. In the study and analysis of protein biotherapeutics, high levels of reproducibility are required for many applications such as pharmacokinetic and pharmacodynamics studies, as well as process development and impurity analysis. Automation of sample preparation, from affinity capture or depletion, to digestion, desalting and reversed-phase LC-MS, allows for operator independent, highly reproducible protein and peptide analysis. This chapter discusses sample preparation and digestion techniques, and outlines automation approaches, focusing on an online, liquid chromatography- based system that simplifies protein and peptide analysis workflow by integrating most of the major steps in protein sample preparation.

Introduction and Background

Typical digestion of proteins using proteases such as trypsin has not advanced from the manual bench top protocols of the mid-20th century (*1*, *2*). While effective, these methods can cause some variability, but more importantly are time-consuming and low throughput (*3*). The introduction of automated protocols has been proposed to address the challenges of variability and speed in protein characterization and analysis workflows (*4*, *5*). Automation in protein

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

sample preparation workflows for LCMS analysis has advanced the progress of developing robust, reproducible assays for the detection of proteotypic or characteristic peptides. Successful implementation of such strategies will prove beneficial to researchers by improving speed and reproducibility and standardizing protocols. The workstation concept, a fully automated protein sample preparation system for LC-MS analysis, discussed in this chapter serves as an example to illustrate the benefits of automated protein sample preparation.

Immunoglobulin protein G (IgG) is a popular modality used for production of biopharmaceuticals. IgGs can identify and neutralize a vast array of foreign agents (Reference Material chapter/Volume 1, Chapter 1 and Mechanisms of Action chapters/Volume 1, Chapter 2). IgGs are important for the recognition of pathogens, and have been utilized as biological treatments for a number of disorders, ranging from immune deficiencies, autoimmune disease, and biological infections (*6*). The structure and function of IgGs readily lend themselves to a host of biopharmaceutical applications, including drug and substrate conjugation for delivery and immune activation (*7*). Regulatory agencies require the characterization of several properties of the therapeutic product including the primary structure (*8*). However, the analytical approaches for protein quantification and sequence mapping require a more seamless and controlled workflow. The integration of automation is therefore critical for continued advancement in this field.

Automation of digestion has not been widely adopted in the industry for a number of reasons. A workshop at the 2014 Annual Meeting of the American Society for Mass Spectrometry focused specifically on this issue (*9*), and cited the lack of research publications and perceived demonstration of successful application of automation to biotherapeutics and other accepted industry protein standards. However, automation of digestion was identified as the single most important factor in sample preparation automation in a survey conducted (S. Abbattiello, personal communication). This, and the emergence of commercial systems has spurred a renewed interest in automation, not only of digestion, but affinity capture/ depletion, desalting and reversed-phase separation. The goal of an automated system is to integrate all of the above steps (as required), offering the flexibility to carry out all or some of the steps without hardware configuration changes, and complete the experiments with minimal human interaction with the sample. Additionally, the ability to execute experiments online (i.e., directly interfaced with a mass spectrometer) also reduces the amount of human interaction and increases the automation of the workflow.

Improvements in the analytical workflow, coupled with the use of immobilized protease columns have opened the possibility of developing online digestion approaches (*5*, *10*, *11*). These advances have been spurred by a need for higher throughput and increased reproducibility. A number of options for automated digestion exist, including offline microtiter plate digestion and online, column-based liquid chromatography (LC) approaches. The benefits of reduced operator time and sample contact are optimized by LC-based approaches, while robotic plate based methods offer parallel processing.

In this chapter, we will discuss the historical digestion approaches, outlining the major sources of variability and limiting steps. We will also introduce

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

currently available automation methods and discuss the benefits and applications of these workflows, with particular emphasis on the analysis of commonly studied proteins, including IgGs, in an automated column-based protein sample preparation workflow.

Digestion

Biology and Chemistry of Digestion

Mass spectrometry (MS) has become ubiquitous in protein sequencing and quantitation workflows especially when discussing protein therapeutics and biosimilars (*12*). Mass spectrometry can provide insight into the protein drug's stability, identity, and modifications; and thus functions as a powerful tool for quality assessment (*13*). Top-down strategies (analyzing intact proteins by LC-MS and LC-MS/MS) can provide valuable information regarding protein identity, variants as well as N- and C-terminal sequences (*14*, *15*). However, decreased sensitivity and limited sequence information, as well as complicated data interpretation currently limit the widespread applicability of top-down analyses (*16*). For most protein analyses by LC-MS, a bottom-up approach (analyzing peptides from digested proteins) which can be both qualitative and quantitative, is more common (*17*).

Proteolysis uses naturally occurring enzymes, generally with known specificity, to digest proteins in a predictable manner into peptides (*17*, *18*). The degree of specificity depends on the enzyme selected. For example, trypsin generally cleaves at the C-terminal side of lysines and arginines whereas pepsin cleaves the N-terminal side of leucine and phenylalanine (*16*). Other rules related to the presence of certain amino acids adjacent to cleavage sites apply in determining the success of a cleavage (http://web.expasy.org/peptide_cutter/peptidecutter_enzymes.html) Although other enzymes can be used, trypsin has gained wide acceptance as the protease of choice for protein digestions for MS workflows, due in part to its availability, cost, stability and purity. The natural occurrence of lysines and arginines typically generates peptide products ranging in length of 9 – 15 amino acids, resulting in *m/z* values ideally suited for most MS applications. The basic residues on the *C*-terminus, combined with the free amine on the *N*-terminus facilitates positive mode ionization of most peptides and encourages predictable and sequence specific fragmentation patterns (*19*).

In order for efficient digestion to occur, it is necessary for trypsin to have access to the lysines and arginines throughout the protein. Secondary, tertiary and quaternary protein structure can restrict access to the substrate sites, giving incomplete and inconsistent digest products (*20*). These structures arise through both covalent (disulfide linkages at cysteine residues) and non-covalent (ionic, hydrophobic, hydrophilic, polar interactions, etc…). Typical workflows begin by denaturing, reducing and alkylating protein samples to disrupt these interactions and remove three-dimensional structures, thus allowing trypsin access to the lysines and arginines in the interior of the folded protein structure.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Denaturation can occur through use of high temperature, chaotropes, high concentration salts or organic solvents (*21*). Based on the method selected for denaturation, digestion efficiency and reproducibility can differ for a given protein (*20*). In most mass spectrometric workflows, denaturation of proteins often occurs in the presence of high concentration chaotropic agents, typically 6 M guanidine or 8 M urea, to achieve complete reduction and alkylation of the protein.

Following denaturation, the sample is typically reduced and alkylated in order to prevent structural reorganization of proteins. Reduction, which breaks the disulfide bonds occurring between two cysteine residues, occurs at an elevated temperature (37-60 °C) in the presence of a reducing agent in significant molar excess. Typical reducing agents include dithiothreitol (DTT), 2-mercaptoethanol and *tris* (2-carboxyethyl) phosphine (TCEP). It should be noted that if using urea as the denaturant, carbamylation of primary amines can occur at elevated temperatures and reduction should be performed at room temperature following heat denaturation (*20*). Once the disulfide bonds are broken, the protein is incubated with an excess of alkylating reagent, most commonly iodoacetamide (IAM) or iodoacetic acid (IAA) at room temperature, in the dark for 30 – 60 minutes. Protocols for these sample pre-treatment steps vary greatly between different laboratories and publications, (*13*, *22*).

At this point, seamlessly integrating the trypsin digestion protocol to the pretreatment steps becomes quite a challenge, and this creates a bottleneck in the sample workflow. This is due to the fact that the optimal conditions for protein denaturation, reduction and alkylation are not typically suitable for tryptic digestion, and some components of the mixture, e.g. TCEP, can have deleterious effects on tryptic activity. Protein solubilization and digestion conditions have dramatic effects on efficiency of enzymatic activity (*23*). Trypsin is inactive at the high concentrations of chaotrope typically used during the initial denaturant step. Therefore, it is necessary to either remove (i.e. through buffer exchange) or dilute the high concentrations of buffer components for efficient digestion to occur. Although diluting urea or guanidine HCl can provide a reproducible and high yield of tryptic peptides, detection limits of the analytical assay must be considered before performing any dilution steps. In the case of robotic systems and liquid handlers designed to carry out the benchtop digestion protocol, researchers often dilute the 6M guanidine HCl to a concentration where trypsin is active (below 1 M) (*24*). Although efficient digestion can occur, it is important to start with a sufficient concentration of biotherapeutic proteins to compensate for the need to dilute the sample while retaining the sensitivity requirements for the desired workflow.

Despite being widely accepted as a sample preparation strategy for protein analysis by mass spectrometry, there are a few drawbacks to trypsin digestion. First, trypsin undergoes autolysis, or self-digestion. When autolysis occurs, trypsin generates contaminating peptides from self cleavage, which broadens its specificity and can induce chymotryptic activity (*19*). Chymotryptic-like cleavages can result in a loss in sensitivity and assay specificity (*25*). Researchers can minimize autolytic activity in two ways: Firstly, using commercially available methylated or otherwise modified enzyme leads to a decrease in access to the autolytic substrate; and, secondly, the ratio of protein to protease is typically

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

limited to between 20:1 to 100:1 (protein: protease) (*22*, *26*). By maintaining a high protein to protease ratio, autolytic activity is minimized while still maintaining favorable kinetics for digestion. Even with these two safeguards in place, peptides associated with trypsin are generally still detected in mass spectra, indicating that autolytic activity is not completely overcome (*18*).

Keeping the protein to protease ratio fairly high (e.g., 10 – 50:1 protein:protease) often results in optimal digestion times in excess of 8 hours (*21*). Trypsin prefers to behave as an endopeptidase and initially generates peptides with missed cleavages (peptides containing multiple lysines and arginines) which are subsequently digested further into "limit peptides" (peptides containing a single lysine or arginine) later in the digestion process through exopeptidase-like activity (*22*). Because changes in the digestion products continue to be detected over a 24 hour period, adhering to strict time-dependent protocols can help control the variability observed in manual digestions. Alternate manual digestion strategies, including microwave-assisted digestion (*27*), the addition of acid labile surfactants (*28*), and filter-aided sample preparation (*29*) can aid the process by decreasing digestion time and improving the digestion efficiency through increasing protein solubility. Despite these benefits, manual sample transfers, pipetting, evaporation, resuspension and solid phase extraction represent user-dependent steps that may introduce variability.

Automation, which will be discussed in more detail in later sections, removes some of the elements of user error by tightly controlling digestion time, sample transfers, and pipetting volumes. Predictable, measurable error remains in the automation components, but is often dramatically lowered in comparison to manual protocols. For manual, overnight digestions, the time variable can have a dramatic effect if the user does not adhere to strict time protocols, or if there are unforeseen circumstances such as delays in getting to the laboratory, equipment availability or simple user error. Similarly, the effect of temperature can dramatically change digestion efficiency, and contribute to increased coefficient of variation (CVs). Likewise, while the rate of trypsin autolysis follows a second order rate (*30*), changes in the substrate-to-enzyme ratio and digest conditions can increase the rate of autolysis, thus further altering the substrate to trypsin ratio, and resulting in altered enzymatic efficiency. For tryptic digestions, the presence of trypsin inhibitors in some biological tissues and fluids can retard the kinetics of tryptic activity (*31*). Furthermore, substrate proteins can have very different kinetics for trypsin, resulting in preferential digestion of one substrate over another (*32*). This in turn leads to the production of peptides, some of which have high affinity for the binding pocket of the enzyme, thus creating an inhibitory effect that is dependent not only on the substrate of interest, but also upon the presence of off-target substrate proteins in the sample. This can result in severe limitations during in-solution digestion of samples with high protein-to-protease ratios. This effect can be overcome by utilizing immobilized enzymes which decrease the substrate-to-enzyme ratio, and with flow through methods that allow all substrates to interact with a massive excess of enzyme during the digestion step, even with small injection volumes (*10*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Digestion Automation

Automation of digestion has several important requirements including **speed**, **sensitivity**, **reproducibility**, and **accuracy**. To realize benefits over the traditional low-tech in-solution digestion protocols, automation must integrate these four components following rigorous manual validation of the processes. The amount of up-front sample preparation should be kept to a minimum, including protein extraction, reduction, alkylation and affinity enrichment. Ideally, an extracted sample should be processed entirely by the automated system in order to remove the effects of operator error. The **speed** of the assay is dependent upon the turnaround time for results, but sample preparation and analysis times in the minutes to hours range are appropriate, especially when a higher throughput is desired. However, digestion and analysis times should not result in decreased **sensitivity** and **reproducibility**. Digestion times and conditions (*e.g.*, buffer selection, temperature, flow rates, enzyme concentration, *etc.*) must be optimized to allow for adequate or complete proteolysis, which ensures the optimum amount of material for an analytical measurement.

There are a number of strategies for automated digestion. The two major approaches are offline (robotic handling) and online (column-based) digestion. The offline strategy relies on robotic liquid handling systems that transfer samples and reagents between tubes or multi-well plate platforms. With a robotic system, most steps for digestion protocols are easily handled: denaturation, reduction, alkylation, buffer exchange, digestion and quenching (*13*). These systems typically require manual transfer of samples from the automation station to the LC-MS system; however, they have the advantage of multiplexed preparation of samples independent of analysis instrumentation.

Online digestion minimizes the amount of pipetting required and the number of sample transfers. High Performance Liquid Chromatography (HPLC) systems offer high reproducibility due to minimal sample manipulation following injection by an autosampler. The Perfinity Workstation (Figure 1) (Shimadzu Scientific Instruments, Columbia, MD and Perfinity Biosciences, W. Lafayette, IN) serves as an example of a commercially available automated sample preparation instrument developed using traditional HPLC components including an autosampler, column ovens, pumps and a UV detector. The multi-position switching valves and alternating flow paths allow the seamless and reproducible transfer of sample and substrates from one column to another without the need for pipetting or manual intervention.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

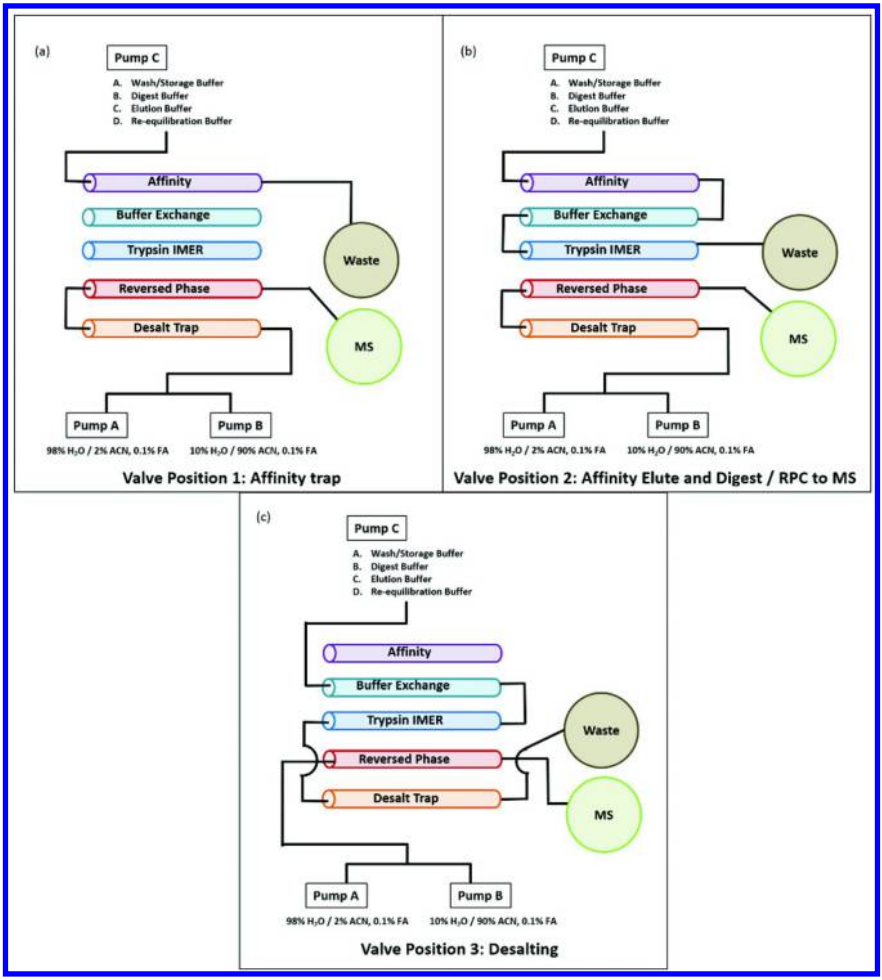

Figure 1. Configuration and valve diagram for the Perfinity Workstation – The Perfinity Workstation uses a series of columns and valves configured to traditional HPLC components to enable flexible method selection during operation without the need for replumbing. Through the five Perfinity columns (affinity capture, buffer exchange, trypsin IMER, desalting and reversed phase) and three switch valves, a range of different applications can be accommodated including affinity enrichment and proteolysis. The affinity column is kept at constant temperature (typically 25°C) using a column oven. The buffer exchange, trypsin and reversed phase columns are all placed in a second oven with a temperature optimized for digestion conditions (40° - 70°C). A quaternary pump delivers buffers to complete the biochemistry steps (affinity capture, elution from affinity, digestion and re-equilibration) and two Prominance XR pumps (Shimadzu, Kyoto, Japan) provide solvents for desalting and gradient elution of the peptides. The key steps for Perfinity Workstation can be accomplished through three configurations of the three valves: (a) affinity capture, (b) affinity elution, digestion and reversed phase to mass spectrometer, and (c) desalting. In

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

a typical experiment involving all steps (affinity capture, trypsin digestion and reversed phase chromatography) the method would change valve configurations in the following order: 1, 2, 3, 2. Wash steps are performed simultaneously as samples are prepared to increase efficiency. Changing from one workflow to another typically involves the addition or removal of columns without the need to replumb. For example, when performing digest only the buffer exchange and affinity columns would be replaced with unions. Sample pretreatment can be performed using pretreatment functionality of the autosampler, but typically this is performed offline before loading samples into the Perfinity Workstation.

Commercially available, optimized affinity columns, trypsin Immobilized Enzyme Reactor (IMER) and buffers (Perfinity, W. Lafayette, IN) facilitate method development and ensure a standardized process across multiple laboratories. The most common workflow using Perfinity technology is trypsin digestion, desalting and reversed phase separation (C. Jelinek, personal communication). Samples may be reduced and alkylated offline or with pretreatment functions of the autosampler using a generic protocol (e.g. 5 mM DTT, 60 ˚C, 1 hour; 10 mM iodoacetamide, 30 min, room temperature in the dark). The samples are then placed into a sample vial or 96-well plate and loaded into the autosampler. Following injection, the sample is pumped using a low-pressure gradient pump through an immobilized enzyme column (typically immobilized trypsin) in an optimized, proprietary digestion buffer (Perfinity Biosciences, W. Lafayette, IN) at experimentally determined optimal temperatures (40 – 60˚C for most applications) (Figure 1b). The trypsin IMER utilizes large particle (10 µm), highly porous resin whose high surface area allows a 100 to 1000-fold excess of trypsin to be immobilized over typical amounts of protein that passes through the column (*33*). The construction of the column not only provides efficient digestion, but the ability to rapidly wash the column with high flow rates without over pressuring. After passing through the trypsin IMER, the digestion products are pumped to a short C_{18} trapping column, where they are retained. Valve switching events move the desalting trap out of line with the digestion buffer and in line with a reversed phase gradient pump. The trap is then washed in mobile phase A (typically 0.1 % v/v formic acid in water) to remove any residual salts arising from the digestion buffer. After washing, a binary gradient consisting of mobile phase A and mobile phase B (typically 90 % acetonitrile in 0.1 % formic acid) is run over a predetermined time. Gradients can be as short as 5 minutes or as long as desired, based upon the detector, assay, and column configurations. Typical reversed phase gradients are carried out in less than 10 minutes. Using an automated system, such as the Perfinity Workstation, methods can be as short as 11 minutes for digestion, desalting and analytical separation, resulting in drastic time savings over traditional, in solution digestion protocols. Furthermore, samples can be queued using a batch function of the integrated software, allowing the system to be run autonomously over several days, given an adequate supply of buffers and mobile phases. A number of possible other workflows can be interfaced directly with the digestion step including affinity enrichment, protein depletion, or size exclusion, allowing orthogonal separations or enrichment steps to be combined in tandem prior to mass spectrometric detection of the tryptic peptides (*5*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

While LC automated digestion is a great leap forward over the traditional manual technology, there are also additional benefits to using an LC-based system. The biggest is that additional chemistries and columns can be added into the system to allow more complex protocols to be run, while retaining the benefits of automation, such that the system carries out automated sample preparation rather than just digestion. A common example of this is online affinity capture of proteins from complex mixtures using antibodies, and subsequent digestion and separation (*34*), which will be covered in detail below. Also, alternative methods including size exclusion chromatography, lectin and phosphopeptide affinity approaches can be automated in this style. When used in conjunction with appropriate sample collection techniques such as a plasma collection method (*26*), the human interaction with the samples are minimized to transferring them to an autosampler vial.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Improvements Achieved through Online Automation

The major benefits realized through automation are improved reproducibility, ease of use, and significant time savings. For quantitative workflows, the most critical component is reproducibility. Traditional benchtop digestions are prone to error, both in the user and process. Automating the process removes the user biases in manual digestions, as well as increasing the throughput of assays. Since online, LC-based automation works on a strictly timed schedule, the results achieved following method optimization are often highly reproducible with CVs under 10 %. An example of such improvements can be seen in Figure 2. In this example, the same NIST reference monoclonal antibody was subjected to digestion using the standard benchtop conditions and the automated Perfinity Workstation.

1. Manual In-Solution Workflow: Reduction and alkylation of the sample (10 mg/mL) was performed using DTT (5 mM for 1 hour at 60˚C) followed by iodoacetamide (10 mM, 1 hour in the dark at room temperature). Quenching was carried out using excess Tris-HCl. The final concentration after this step was 1 mg/mL. The sample (diluted as necessary) then was subjected to digestion using standard benchtop conditions (37˚C, 18 hours at a trypsin-to-mAb ratio of 1:20. Following digestion, samples were acidified using 0.5 % formic acid. Samples were dried using a SpeedVac vacuum concentration and stored at -80 °C prior to analysis. Samples were injected using a SIL-30AC (Shimadzu, Kyoto, Japan) autosampler connected to a Nexera X2 UHPLC (Shimadzu, Kyoto, Japan) system. Separation was carried out on an ACE Excel C18 column (2.1 x 100 mm, 3 µm particle size, 300 Å pore size) (Advanced Chromatography Technologies LTD, Aberdeen, Scotland). Peptides were detected using multiple reaction monitoring transitions (MRMs) that were previously optimized using the open source software package, Skyline (MacCoss Lab Software, Seattle, WA). During these experiments, every precaution was taken to ensure reproducible results including creating fresh buffers and reagents at the time of the experiment. Unavoidable variations occurred with digestion

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

times due to minor unintentional variations in processing times, although every attempt was made to keep the process as reproducible and uniform as possible.

2. Offline Reduction/Alkylation followed by Automated Digestion: Reduction and alkylation was performed in the same manner as the manual workflow. Samples were then directly injected onto the Perfinity Workstation (with no affinity extraction step) to perform automated digestion. A digestion time of 6 minutes at 50˚C using a Perfinity IMER column (2.1 x 50 mm immobilized trypsin column) was conducted.

Both sets of samples were analyzed using a LCMS-8050 triple quadrupole mass spectrometer (Shimadzu Kyoto, Japan). Peptides were separated on a 6-minute gradient from 2 – 40 % acetonitrile/ 0.1 % formic acid. Multiple reaction monitoring (the isolation of a precursor ion, subsequent fragmentation via collisionally induced dissociation in the collision cell and isolation of a specific product ion in the third quadrupole prior to detection using a pulse counting detector; MRM) transitions were generated using Skyline (*35*), and optimized for precursor, product, and collision energy. Data were imported into Skyline or processed for quantitative analysis using LabSolutions v.5.65 (Shimadzu, Kyoto, Japan). Over the course of 20 digestions across 5 days, there was a great deal of variability in the intensity of product ions produced in the benchtop method (Figure 2, top panel). The sources of this variability are unknown but are likely related to the digestion time or other conditions. While digestions were tightly controlled, operator error such as pipetting or vacuum drying times may have had an effect. The Workstation generated peptides at different relative abundances than the benchtop process, but with more reproducible CVs averaging approximately 10 % (Figure 2, lower panel). The peptide yield (as determined by area counts of measured peaks in the MRM method) was comparable between the two methods. The sample processing time was reduced from an overnight process (~18 hours) to less than an hour using the online automation.

A concurrent advantage is ease of use; since the system does not require the addition of enzymes or other digestion components, the sample can simply be added to the autosampler and the experiment executed. The Perfinity interface can be controlled to allow only a limited number of methods, and eliminate the ability to modify or edit method files, so that the same protocol is executed every time. This also translates into extensive time savings, both in man hours and time to experimental completion. Sample handling and processing is ported to the instrumentation, freeing the operator to carry out other experiments. Furthermore, with accelerated digestion times, experiments that commonly take weeks to be optimized can be readily optimized in a much shorter time frame. The Perfinity Software exists as a wrapper software around LCMS LabSolutions (Shimadzu Kyoto, Japan). The user interface completely avoids the need to write complicated time programs to execute the valve switching and timings required to complete the different steps: digestion, column washing, desalting and separation. From a user perspective, only a few parameters need to be entered into the user interface in order to develop methods: digestion time, digestion temperature and gradient conditions. For method optimization, the user can

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

quickly create methods that vary in digestion time and digestion temperature and monitor digestion efficiency by observing peak areas of limit peptides and peptides containing missed cleavages. Although the Perfinity Workstation utilizes LCMS LabSolutions software to operate, it is a vendor neutral platform that can interface with any mass spectrometer using contact closure to trigger mass spectral acquisition.

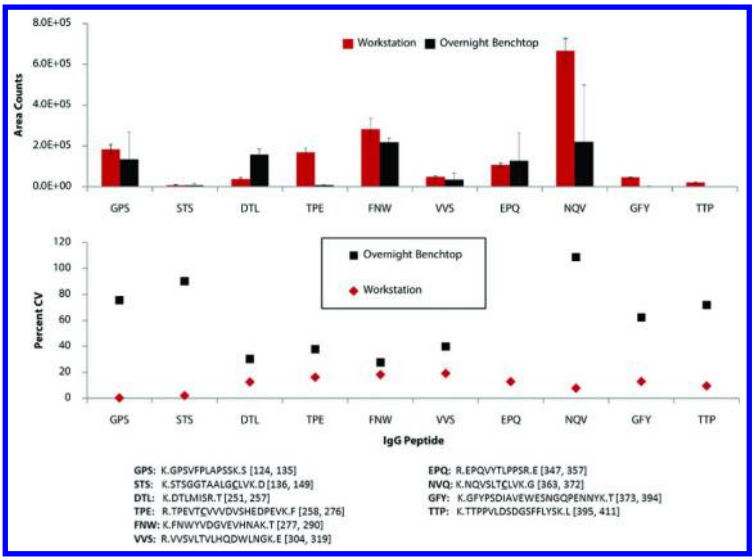

Figure 2. Comparison of overnight manual digestion and Workstation digestion. Quantitative results were exported into Skyline daily v2.5.1.5963 and processed using a standard digestion and reversed phase separation workflow. Peak area and Percent CVs were calculated for each peptide and data were exported to excel for visualization. Workstation CVs averaged 10.5 %, while the manual digestions were more varied with some CVs in excess of 100 %. Peptides are represented on the abscissa by the first three amino acids in the tryptic peptide sequence (see legend for complete amino acid sequence). Please see the text discussion for detailed method information.

Quantification and Linearity

Detection of peptides via high mass accuracy LC-MS/MS instruments can deliver a great deal of information, but the major limitation is that quantification is not readily achieved in many trapping instruments. The gold standard for MS quantification remains the single- or multiple-reaction monitoring (SRM, MRM) using triple quadrupole MS, where a transition from a precursor *m/z* to a product *m/z* is monitored. This reaction is highly specific, especially in the case of multiply charged peptides, and can be used to specifically and accurately monitor

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the relative or absolute amounts of a peptide in a sample. The latest generation of mass spectrometers are highly sensitive, accurate, and robust, and so the majority of error in the measurement is due not to the instrument but to the sample preparation. Small pipetting or dilution errors can be magnified in the sample processing and result in incorrect determination of the concentration of an analyte in a sample. This is unacceptable in bioprocess and biopharmaceutical workflows, especially when monitoring potentially subtle changes during pharmacokinetic (PK) or pharmacodynamics studies (PD). Therefore, reproducibility is critically important, more so than digest efficiency, which only contributes to sensitivity of an assay. Using a suitably fast LC-MS/MS instrument, it is possible to measure multiple peptides from the same protein in a short analytical run. For example, the NIST monoclonal antibody was digested using the Offline Reduction/Alkylation followed by Automated Digestion protocol described above. During LC-MS analysis, 27 of the possible tryptic peptides from the light and heavy chain were monitored in a single run utilizing digestion and reversed phase separation, and assessed for reproducibility and linearity through a dilution series from 0 – 500 µg/mL, using transitions predicted from the Skyline (*35*). MRMs for these 27 peptides were utilized to select four peptides and resulting transitions with optimal quantitative performance based on signal intensity, reproducibility and chromatographic resolution. The resulting peptides selected from the variable and constant regions were monitored and an MRM method was built (Figure 3). The data demonstrate that a linear response is achieved for each of the monitored peptides, and a quantitative relationship achieved based on MRM analysis.

Quantitative measurements presented in Figure 3 could be further improved by utilizing heavy (e.g., multiple ^{15}N and ^{13}C) stable isotope labeled internal standards. These include labeled peptides, and increasingly, labeled proteins (*36*). These pass through the trypsin column and are trapped on a C_{18} trap column with the target analytes, allowing further improvements in quantification accuracy and validation. Incorporating internal standards into an assay results in absolute quantification and increased confidence in sample measurement. Automated digestion which incorporates heavy standards can be achieved by either spiking intact heavy proteins in the samples prior to affinity capture, or by adding heavy peptides during the sample analysis. In the Perfinity Workstation workflow Figure 1, heavy peptides are injected directly into the trypsin column along with the enriched sample. This approach is attractive as heavy peptide standards are more readily available than synthesized heavy proteins.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

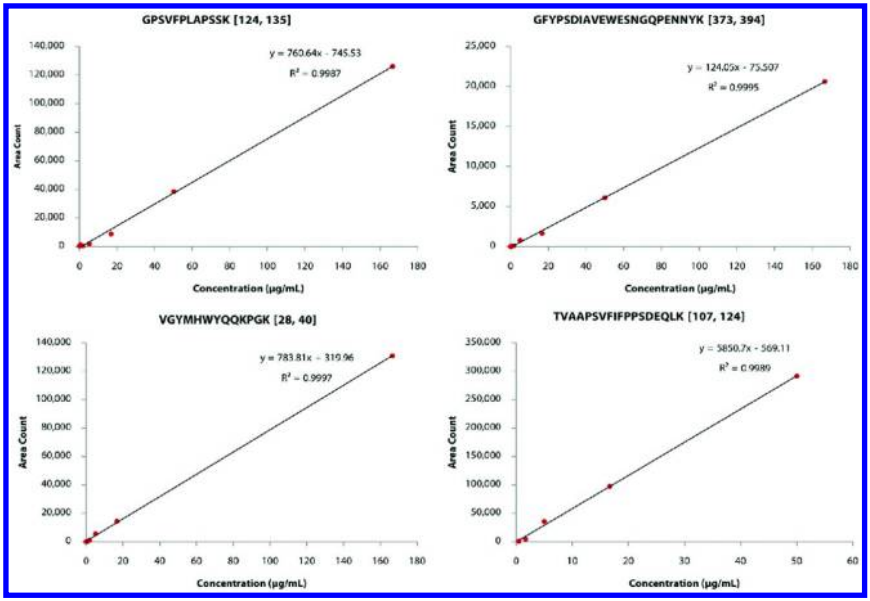

Figure 3. Quantitative analysis of peptides originating from the NISTmAb. Serial dilutions of the intact protein were digested and analyzed using a Perfinity Workstation coupled to an LCMS-8050. Peptides from both the light and heavy chain were measured, including peptides from the constant Fab, constant Fc, and variable Fab regions.

Peptide Mapping

One of the major benefits of peptide-based LC-MS/MS protein analyses is that one is able to confidently assign the amino acid sequence of detected proteins using a data dependent, probability-based database searching approach. In this method, the ions entering the MS are subjected to collision-induced dissociation, where the amide bonds are fragmented in a predictable manner by excitation energy and a collision gas (argon). In conjunction with a database populated with known amino acid sequences, the experimentally determined masses can be compared to *in silico* precursors and fragments enabling assignment of precursors to the predicted amino acid sequences.

A pertinent application for peptide mapping in primary structure confirmation and detection of posttranslational modifications of a mAb. Monoclonal antibodies consist of multiple chains that are connected by disulfide bonds. These chains are comprised of several regions; a variable region that is responsible for the specificity and strength of the antibody-antigen interaction, a hinge region, and a constant region that is highly conserved within classes of IgGs in a specific species. Accurately mapping the variable region is critical to understanding the exact identity of the mAb; since many therapeutic mAbs share a significant portion of sequence homology. As an example, the NIST monoclonal antibody was digested using the Offline Reduction/Alkylation followed by Automated Digestion protocol described above. Figure 4 illustrates an example of data-dependent

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

acquisition of digestion products from the NIST reference mAb following reduction, alkylation and on-column digestion at 50˚C for 6 minutes, followed by reversed phase separation using water and acetonitrile with 0.1 % (v/v) formic acid. The reversed phase separation consisted of a hold for 2 minutes at 5 % acetonitrile with 0.1 % (v/v) formic acid, a linear gradient from 5 – 50 % over 10 minutes, a linear gradient from 50 % – 90 % over two minutes, followed by a one minute high organic wash at 90 % and subsequent re-equilibration. By extracting ion chromatograms for several representative NISTmAb tryptic peptides (from light chain L(1-17), L(107-124), L(125-140), L(148-167), L(168-181), (189-205) and from heavy chain H(124-135), H(136-149), H(258-276), H(277-290), H(347-357), H(363-372), H(373-394), H(395-411) from the TIC (Figure 4, middle and top panels, respectively), it is evident that the predominant TIC peaks originate from the mAb, and that the peptides detected can be monitored using accurate masses of precursors. Further interrogation of the MS/MS spectra (an example of which is shown in Figure 4 LLIYDTSKLASGVPSR), and searching with Mascot (Matrix Science London, U.K.) against a database appended with the NISTmAb amino acid sequence, demonstrates that the digestion resulted in high coverage (>95 %) at a 5 % false discovery rate. Additionally, the data dependent approach identified modified peptides (shown as green shaded amino acids in Figure 4, bottom panel), including oxidation of methionine, a common PTM occurring on proteins, and carbamidomethylation of cysteine, produced during the alkylation step of the sample preparation procedure. Analysis of PTMs by LC-MS/MS represents an application of the peptide mapping strategy to measure protein damage or degradation (e.g., oxidation, deamidation) without *a priori* knowledge of the site of modification, assuming the specific modifications are considered while using database search tools (*37*, *38*). This strategy may be carried out in conjunction with intact protein analysis (by LC- or MALDI-MS), *C*- and *N*-terminal sequencing (by MALDI-ISD and Edman sequencing) for orthogonal confirmation. This type of measurements, supported by the rapid and automated digestion protocol (6 minutes), enable higher throughput analysis by peptides mapping, which is highly beneficial for biotherapeutics field applications from early screening and discovery to product development and release.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

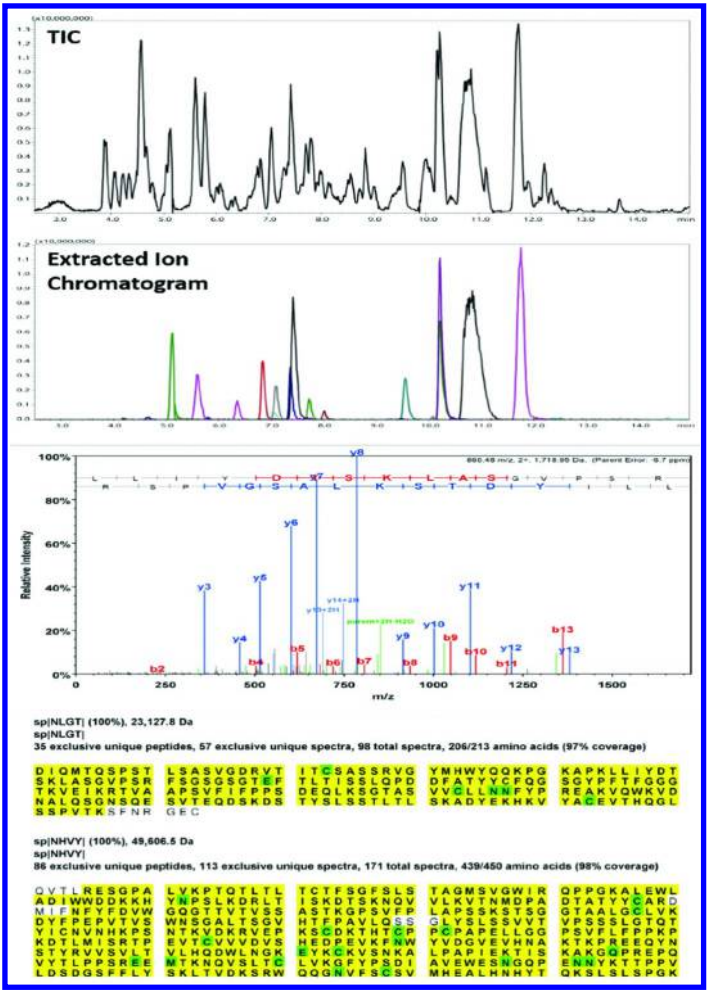

Figure 4. Peptide mapping of the NISTmAb by high mass accuracy LC-MS/MS (refer to additional details in the text). Automated digests were analyzed using an LC-MS-IT-TOF with a top 3 data-dependent strategy. Data were searched (using Mascot) against the Common repository of adventitious proteins (a database of common laboratory introduced contaminants), with the NIST and other monoclonal IgG sequences appended. The Mascot results were exported to Scaffold v.4 and researched using the X!Tandem search tool. The results were filtered for 80 % protein ID, minimum 3 peptides with a maximum 5 % peptide false discovery rate. Light (yellow) color indicates identified sequence, and dark (green) represents modified amino acids.

Affinity Capture

A major challenge in protein and peptide quantification during PK/PD experiments is the abundance of the target molecule in complex biofluids. Often, the substrate is at low concentrations relative to the matrix, and thus digestion and MS approaches can be challenging given the typical amounts of starting material (*39*, *40*). A number of potential solutions are available to target specific compounds from complex samples, including depletion of abundant proteins, specific enrichment of a subset of proteins, or affinity capture of specific target proteins. Each of these have specific benefits and disadvantages, particularly with regard to automation.

Abundant protein removal is a simple approach where high concentration proteins are specifically removed from the sample via affinity interactions; the high abundance proteins (e.g., albumin) are bound to the column, and the flowthrough is collected and analyzed (*41*). The major limitations of this approach are that depletion may not achieve 100 % capture, and therefore significant matrix or contaminants still exist. More importantly, abundant molecules such as albumin are carrier proteins, and often interact or bind the susbtrate of interest, thus reducing the reproducibility of the removal (*42*).

Affinity capture, where a specific capture antibody (or immunoglobulin-binding proteins such as Protein G or Protein A) is used to selectively enrich samples out of solution, is an accepted approach for targeting single proteins (i.e. a mAb therapeutic) from a complex mixture. The chemistry is the same as that used in enzyme-linked immunosorbent assays (ELISAs), where an capture antibody with an affinity to the substrate is utilized to capture the target protein. Following the affinity capture of the target, subsequent washes remove non-bound and weakly binding contaminants, and the target protein is then eluted from the affinity column by acidic conditions, which dissociate the interaction between capture antibody and target protein. Often during manual enrichment methods, immune complex formation is carried out in solution since the tertiary conformation of the capture antibody-target interaction may be required for optimal binding. Following in solution anti-target antibody binding to the target protein, the immune-complex is isolated and enriched using a solid support with a cross-linked moiety with affinity for the anti-target antibody such as Protein A, Protein G, or in cases where a biotinylated anti-target antibody is used monavidin or tetravidin. Alternatively, column and flowthrough type applications can be implemented using immobilized capture antibodies on the column, provided the target protein has sufficient affinity and access to the binding site of the capture antibody. This can be controlled by optimizing the flow through the column, thus lengthening the exposure of the epitope to the binding site.

Automation of the affinity capture process can be achieved in a number of ways. Ideally, the operator interaction with the sample is limited, so a flowthrough approach is desirable, where an capture antibody is immobilized onto a column via streptavidin-biotin interactions, or some other linking approach that retains capture antibody structure and activity. However, many antigen-antibody interactions require in solution incubation and so this must be carried out at the benchtop with agitation for a fixed amount of time. The immune complex can

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

then be injected onto an automated system, where the enrichment occurs via immunoglobulin-binding protein (Protein A or Protein G) interactions with the immune complex, or via streptavidin-biotin interaction in the case of biotinylated capture antibodies. There are several considerations for deciding whether to utilize either approach, and some tests must be carried out to empirically determine the optimal conditions for an experiment. The major factors are whether the antigen binds sufficiently on an affinity column (e.g., Protein A/G or avidin) given the relatively low residence time there, and whether the capture antibody is capable of being regenerated in the selected buffer conditions for elution and re-equilibration. The benefits of using the antibody column-based approach is that you can re-use a column multiple times, saving cost and also offering improved reproducibility as conditions for the immune interaction are more tightly controlled by flow rate, temperature and residence time in column, by minimizing operator interactions.

For studies involving mAbs, a relatively simple approach where the free mAb is bound to a Protein G column can be utilized. The advantage of this approach is that Protein G has a very high affinity for IgG (*43*), so a column-based method can be used. There is no need for a primary capture antibody in this configuration, so the reproducibility and sensitivity is dependent on only the binding kinetics of the mAb to the Protein G column. This type of experiment, taking the samples from affinity to digestion and MS can be completed in less than 1 hour by using the Perfinity Workstation. This can also be done using samples in a complex matrix, such as whole blood or serum., Figure 5 illustrates the affinity capture approach. The NIST reference mAb was spiked into mouse whole blood, which was diluted 1:10 and injected onto a Perfinity Protein G column. Affinity loading was carried out at 10 µL/min for 10 min, washed with 5 column volumes, and subsequently eluted through a size exclusion buffer exchange column in optimized digestion buffer onto a Perfinity IMER trypsin column. Digestion and detection parameters were as described in Figure 2 above, with the exception that no reduction or alkylation was carried out. Peptides representing the variable and conserved regions were monitored. Figure 5 shows the ability to selectively monitor enriched IgG from a complex matrix such as mouse blood at a variety of concentrations. The data correlates well to a linear curve with R^2 values of 0.9885 (GPSVFPLAPSSK) and 0.9737 (LASGVPSR) indicating that the method is robust over a range of concentrations.

Using serum and complex sample matrices introduces some challenges, including viscosity, high levels of interfering species, and the presence of contaminating IgGs. For example, immunoglobulins originating from the mouse serum in the example above would compete for binding sites on the Protein G affinity column and may be captured using the Protein G enrichment approach. Accurate quantification is still possible in this case because the total concentration of injected IgG (mouse IgG + NISTmAb) was below that of the affinity column capacity, and the monitored peptide sequences are unique to the humanized IgG (NISTmAb) versus background mouse IgG. The naturally occurring mouse IgG that are co-extracted may also carry with them antigens that are not of interest in the study. To overcome this challenge, a biotinylated anti-therapeutic mAb capture antibody coupled with an avidin column, or a therapeutic mAb selective

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

affinity column (such as an immobilized antigen), may be utilized in order to allow larger sample loading volume and specific therapeutic antibody capture. The use of mono- and tetraavidin columns for online affinity capture in conjunction with biotinylated affinity reagents can improve specificity for capture of the target antibody compared to the global antibody binding characteristics of a Protein A or Protein G column.

Using an affinity capture, digest and MS method, it is possible to detect low fmol levels of protein extracted from a complex mixture (*44*). This can be important when monitoring the dynamics of a mAb in PK studies, since monitoring may extend over long periods of time due to the long and variable half-life of IgGs circulating in the body (reviewed in (*45*)). Similarly, in PD work, the targets of immunotherapy may be low abundance and relatively variable in both half-life and concentration, depending on the kinetics of the particular therapeutic and target. Accurate determination of concentration is critical for understanding the effectiveness and biological fate of therapeutics during PK/PD studies, and using an automated affinity and digestion workflow assists in long-term reproducibility and method standardization.

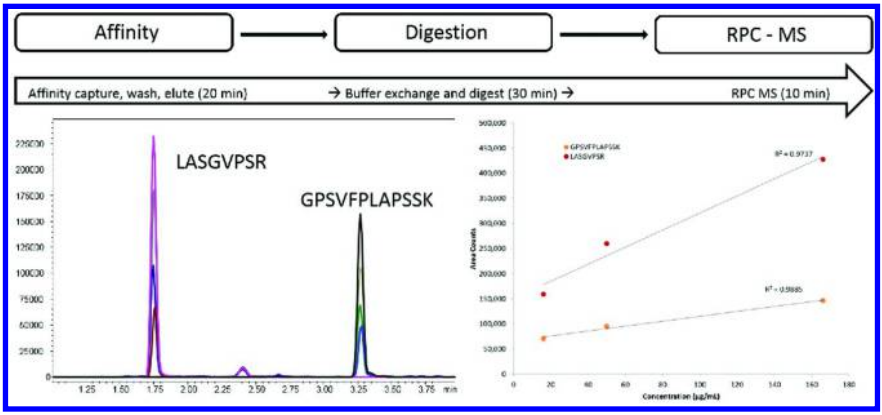

Figure 5. Affinity capture of the NIST monoclonal antibody from heat inactivated mouse blood. Antibody preparations were spiked into mouse blood and injected onto the Perfinity Workstation Affinity capture was carried out using a Protein G column, and were eluted and buffer exchanged onto an IMER trypsin column. The resulting peptides were separated by reverse phase chromatography prior to MRM analysis on an LCMS 8050 triple quadrupole. The variable colors in the chromatograms represent multiple injections at different spike concentrations. Please refer to the text for additional methods and discussions.

Other Workflows and Applications

Another area of interest in process monitoring during development of monoclonal antibody biotherapeutics is the detection and quantitation of product-related impurities (e.g. isoforms and post-translational modifications such as oxidation, glycosylation). Production processes are often dynamic such that real time monitoring is desirable. In turn, bioreactor parameters such as

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

oxygen levels, pH and nutrient concentrations can be adjusted to ensure the highest degree of product fidelity. Multiple day protocols for trypsin digestion and sample cleanup are not acceptable for many timeframes. Being able to automate analysis and generate results within an hour makes LC-MS peptide mapping for the monitoring process more attractive for industrial applications.

It should also be noted that using MS, it is possible to measure proteins in a data dependent manner, meaning that you can screen for "unknowns" without targeting a specific mass, sequence or protein. In the production of biotherapeutics, there are a number of critical steps that need to be measured, from the initial cloning and production of a biotherapeutic as well as downstream purification and production of the final product. At each step, many different process-related impurities need to be measured such as host cell proteins). Online automation of LC-MS analysis is capable of simultaneously monitoring product quality and identification and/or quantification of HCP clearance in a time regime amenable to process monitoring. Using a targeted approach involving the use of heavy-isotope labeled analogs as standards, as described in this chapter, this assay can provide truly quantitative results.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Future Directions

The detection and measurement of biotherapeutic proteins and protein/peptide targets remains a priority for many scientists. As proteins containing multiple isoforms become candidates for biotherapeutic targets, the importance of mass spectrometry for detection and differentiation of these isoforms will increase. The same can be said for biopharmaceutical process monitoring and/or PK/PD studies. Both the quantity and identity of the product, product related impurities, and process related impurities requires quantification and monitoring. The immuno-MS and immuno-MRM workflows achieved using the Perfinity Workstation show great promise as a future standard for sample preparation in that they achieve speed, reproducibility, selectivity, and resolution in a fully automated manner.

In the future, it may be possible to achieve automation through paper chromatography techniques which do not require power supplies. New technologies such as the Noviplex™ Cards can generate exactly 2.5 µl plasma from a single drop of blood without the use of centrifugation, evaporation or extraction steps (26) and interface directly with LCMS and direct ionization methods such as MALDI-TOF MS or paper spray. Using these cards in their current form, plasma is generated by passing blood through a series of porous membrane layers which filter out the red blood cells. The plasma is then collected onto a cellulose disc that can hold a specific volume based on its geometry. It is conceivable to incorporate sample preparation workflows (e.g. affinity extraction) into the Noviplex™ Cards or similar technologies for enrichment to begin at the sample collection site and to continue during sample transport. The benefits of reducing the interactions between the sample and manual manipulation are increased reproducibility coupled with higher throughput without the need for constant intervention. As this technology matures, additional workflows such

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

as post-translational modification analysis, isoform analysis and many other, potentially new applications and enrichments will result in improvements in the depth of knowledge and reproducibility of proteomics-based qualitative and quantitative workflows.

Summary and Conclusion

Automation is beneficial for the development of accurate, reproducible and rapid analysis of proteins and peptides by MS techniques. The major sample preparation processes of affinity enrichment, digestion and separation are suitable for automation, both in online and offline configurations. The increasing needs for reproducibility and throughput are driving a great deal of innovation in this area, resulting in automated workflows that offer CVs that are compatible with the requirements of modern quantitative protein analysis.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Acknowledgments

The authors would like to thank Kevin Meyer and John O'Grady of Perfinity Biosciences for their technical input and assistance with the manuscript review.

References

1. Naughton, M. A.; Sanger, F.; Hartley, B. S.; Shaw, D. C. *Biochem. J.* **1960**, *77*, 149–163.
2. Naughton, M. A.; Sanger, F. *Biochem. J.* **1961**, *78*, 156–163.
3. Walmsley, S. J.; Rudnick, P. A.; Liang, Y.; Dong, Q.; Stein, S. E.; Nesvizhskii, A. I. *J. Proteome Res.* **2013**, *12*, 5666–5680.
4. Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A. J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. *J. Proteome Res.* **2010**, *9*, 761–776.
5. Wang, S.; Regnier, F. E. *J. Chromatogr. A* **2001**, *913*, 429–436.
6. Grabowski, S. R.; Tortora, G. J. *Principles of Anatomy and Physiology*, 10th ed.; Wiley: Hoboken, NJ, 2003.
7. Li, G. *Drug Discovery Ther.* **2013**, *7*, 178–184.
8. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product Guidance fo Industry*; Silver Spring, MD: Office of Communications, Division of Drug Information, 2015.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

9. Sadagopan, N.; Abbatiello, S.; Dufield, D. In *62nd ASMS Conference on Mass Spectrometry and Allied Topics*; 2014.
10. Spross, J.; Sinz, A. *Anal. Chem.* **2010**, *82*, 1434–1443.
11. Wang, F.; Wei, X.; Zhou, H.; Liu, J.; Figeys, D.; Zou, H. *Proteomics* **2012**, *12*, 3129–3137.
12. Zhang, H.; Cui, W.; Gross, M. L. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett.* **2014**, *588*, 308–317.
13. Chelius, D.; Xiao, G.; Nichols, A. C.; Vizel, A.; He, B.; Dillon, T. M.; Rehder, D. S.; Pipes, G. D.; Kraft, E.; Oroska, A.; Treuheit, M. J.; Bondarenko, P. V. *J. Pharm. Biomed. Anal.* **2008**, *47*, 285–294.
14. Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. *Mol. Cell. Proteomics* **2012**, *11*, 1758–1767.
15. Gucinski, A. C.; Boyne, M. T. *Anal. Chem.* **2012**, *84*, 8045–8051.
16. López-Ferrer, D.; Petritis, K.; Robinson, E. W.; Hixson, K. K.; Tian, Z.; Lee, J. H.; Lee, S.-W.; Tolić, N.; Weitz, K. K.; Belov, M. E.; Smith, R. D.; Pasa-Tolić, L. *Mol. Cell. Proteomics* **2011**, *10*, M110.001479.
17. Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
18. Olsen, J. V; Ong, S.-E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3*, 608–614.
19. Vandermarliere, E.; Mueller, M.; Martens, L. *Mass Spectrom. Rev.* **2013**, *32*, 453–465.
20. Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. *J. Proteome Res.* **2010**, *9*, 5422–5437.
21. Li, F.; Fast, D.; Michael, S. *Bioanalysis* **2011**, *3*, 2459–2480.
22. Rivera-Burgos, D.; Regnier, F. E. *J. Sep. Sci.* **2013**, *36*, 454–460.
23. Ye, X.; Johann, D. J.; Hakami, R. M.; Xiao, Z.; Meng, Z.; Ulrich, R. G.; Issaq, H. J.; Veenstra, T. D.; Blonder, J. *J. Proteomics* **2009**, *73*, 112–122.
24. Richardson, J.; Shah, B.; Xiao, G.; Bondarenko, P. V.; Zhang, Z. *Anal. Biochem.* **2011**, *411*, 284–291.
25. Hildonen, S.; Halvorsen, T. G.; Reubsaet, L. *Proteomics* **2014**, *14*, 2031–2041.
26. Kim, J. H.; Inerowicz, D.; Hedrick, V.; Regnier, F. *Anal. Chem.* **2013**, *85*, 8039–8045.
27. Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6*, 586–591.
28. Ross, A. R. S.; Lee, P. J.; Smith, D. L.; Langridge, J. I.; Whetton, A. D.; Gaskell, S. J. In *Proteomics*; 2002; Vol. 2, pp. 928–936.
29. Wiśniewski, J. R.; Wiśniewski, J. R.; Zougman, A.; Zougman, A.; Nagaraj, N.; Nagaraj, N.; Mann, M.; Mann, M. *Nat. Methods* **2009**, *6*, 359–362.
30. Finehout, E. J.; Cantor, J. R.; Lee, K. H. In *Proteomics*; 2005; Vol. 5, pp. 2319–2321.
31. Clifton, J.; Huang, F.; Rucevic, M.; Cao, L.; Hixson, D.; Josic, D. *J. Proteomics* **2011**, *74*, 935, 941.
32. Fernandez-De-Cossio, J. *Anal. Chem.* **2011**, *83*, 2890–2896.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

33. Regnier, F. *Next Generation Blood Sampling*; Podium presentation presented at the 2014 Mass Spectrometry: Applications for the Clinical Laboratory Conference, San Diego, CA, March 2014.
34. Madian, A. G.; Rochelle, N. S.; Regnier, F. E. *Anal. Chem.* **2012**, *85*, 737–748.
35. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. *Bioinformatics* **2010**, *26*, 966–968.
36. Kuhn, E.; Whiteaker, J. R.; Mani, D. R.; Jackson, A. M.; Zhao, L.; Pope, M. E.; Smith, D.; Rivera, K. D.; Anderson, N. L.; Skates, S. J.; Pearson, T. W.; Paulovich, A. G.; Carr, S. A. *Mol. Cell. Proteomics* **2012**, *11*, M111.013854.
37. Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466–1467.
38. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.
39. Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
40. Boja, E. S.; Rodriguez, H. *Korean J. Lab. Med.* **2011**, *31*, 61–71.
41. Holewinski, R. J.; Jin, Z.; Powell, M. J.; Maust, M. D.; Van Eyk, J. E. *Proteomics* **2013**, *13*, 743–750.
42. Gundry, R. L.; Fu, Q.; Jelinek, C. A.; Van Eyk, J. E.; Cotter, R. J. *Proteomics: Clin. Appl.* **2007**, *1*, 73–88.
43. Saha, K.; Bender, F.; Gizeli, E. *Anal. Chem.* **2003**, *75*, 835–842.
44. *Analysis of Human IgG Using Automated Digestion Coupled Directly to LCMS-8050*; Shimadzu: Columbia, MD, 2014; pp 1–4.
45. Wang, W.; Wang, E. Q.; Balthasar, J. P. *Clin. Pharmacol. Ther.* **2008**, *84*, 548–558.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch012

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Chapter 13

Improved Identification and Quantification of Host Cell Proteins (HCPs) in Biotherapeutics Using Liquid Chromatography-Mass Spectrometry

Weibin Chen,[*,1] Catalin E. Doneanu,[1] Matthew Lauber,[2] Stephan Koza,[2] Kesh Prakash,[3] Martha Stapels,[1,4] and Kenneth J. Fountain[2]

[1]Biopharmaceutical Business Operation, Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757, United States
[2]Consumables Business Unit, Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757, United States
[3]Process and Analytical Development, Emergent BioSolutions, 300 Professional Drive, Gaithersburg, Maryland 20879, United States
[4]Present Address: Genzyme Corporation, 1 The Mountain Rd., Framingham, Massachusetts 01701, United States
[*]Phone: 508-482-2090. Fax: 508-482-3085.
E-mail: Weibin_Chen@Waters.Com

Host cell proteins (HCPs) are classified as one of the major process-related impurities in the production of recombinant therapeutic proteins. The presence of host cell proteins in the final drug product can elicit adverse patient immunogenic reaction and result in a loss of drug efficacy. Regulatory guidelines mandate the establishment of suitable analytical methods to ensure accurate determination of residual HCPs and demonstrate that contaminant levels are minimized during the bioprocess development of the therapeutic modality. Detailed analytical information on HCP composition can help to develop an efficient and robust bioprocess so the final products meet the HCP regulatory guidelines, therefore protecting patient safety. The aim of this chapter is to compare different analytical methods that are commonly used by the biopharmaceutical industry (also globally accepted

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

by regulatory bodies) for the analysis of HCPs, review the strengths and limitations of the analytical methodologies, and then propose a rational, cost-effective strategy based on liquid chromatography-mass spectrometry (LC-MS) for HCP identification and quantitation. A range of immunospecific (e.g. enzyme-linked immunosorbent immunities assays (ELISA) and Western blot) and non-specific methods (e.g. two-dimensional difference gel electrophoresis (2D-DIGE) and LC-MS) are discussed with a primary focus on the application of LC-MS approaches for HCP analysis for bioprocessing strategies and purification development. The chapter also presents several case studies on how 2D-LC-MS methods are used to acquire the identity and measure the concentration of individual HCP in biopharmaceutical samples from bioprocess development, demonstrating the power of the technique for defining a pharmaceutical product.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Introduction

Production of therapeutic proteins using recombinant cultured cells is a cost effective method for supplying commercial quantities of a drug substance. However, the manufacturing and purification processes used to obtain these recombinant products leave the potential for contamination from host cell components, such as host cell proteins (HCPs) and host cell DNA/RNA (*1–4*). Any individual HCP may have a foreign or "non-self" character and the potential to elicit an immune response in humans (*5*), resulting in adverse clinical effects (*6*) or reducing drug stability (*7*). Removal of host cell contaminants to levels deemed safe has accordingly been a major subject of bioprocess development, and is mandated by regulatory guidelines (*8–10*). In a broad sense, minimization of all forms of contamination in a therapeutic protein is a matter not only of safety, but of process robustness.

The host cells used for the production of therapeutic proteins are complex biological systems ranging from bacteria (*11*, *12*) to cell lines derived from mammalian (*13*) or insect species (*14*). The expression cells contain hundreds to thousands of HCPs and other biomolecules that could contaminate the final product. The composition and abundance of HCPs that are present in samples from different stages of purification processes and from the final drug substance are influenced by many factors. These factors obviously include the identity of the host expression system and the nature of the purification process itself (*13*). Moreover, the composition of HCPs is closely related to how the biologic of interest is expressed and the physiochemical properties of the biologic being expressed (*13*, *14*). These factors contribute either collectively or individually to the final composition of a biotherapeutic product and its HCP profile. The dynamic nature of biologics production and preparation makes it quite difficult to predict HCP profiles *a priori,* and the final description on HCPs is often learned only by testing during process development.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

The variation in manufacturing processes and expression systems implies the expression profiles of HCPs can be drastically different, creating a nearly infinite variety of HCP profiles in biopharmaceutical samples. Recognizing the complexity of HCP assays and the unique nature for each product, regulatory bodies cannot explicitly define the standard HCP detection methodology or specifications on HCP contamination (*2*, *3*). However, regulatory guidelines unequivocally require that HCPs and other impurities in biopharmaceuticals be minimized (*8–10*). And understandably, each submission is subject to examination on a case-by-case basis (*8–10*, *15*) to determine appropriate patient risk (and hence maximum allowable levels of HCPs) based on the dose, route, schedule of administration, and patient population (*16*).

The heterogeneity and variation of HCPs in recombinant drug products highlights one of the major dilemmas that the biopharmaceutical industry faces today. While the industry strives to develop a robust bioprocess to control the level of HCPs present in their drug substances, it is obvious HCPs in the final products can only be characterized to a level defined by the sensitivity and specificity of the employed assays. Improvements in testing sensitivity will reveal trace HCPs that would have previously been below limits of detection. Such information will facilitate the continuous optimization and validation of the efficiency and consistency of biotherapeutic manufacturing processes, resulting in safer and higher quality biopharmaceuticals.

Given the importance of HCP analysis for the development, optimization and validation of proprietary biopharmaceutical purification processes, there has been significant interest in the analytical development of HCP assays (*17–28*). As a result, the methodologies for HCP analysis are constantly evolving. This chapter is the second chapter in the book series focusing on HCP-related subjects (Process Impurities chapter/Volume 2, Chapter 9). A major focus of this chapter is on the recent development of HCP methods, particularly LC-MS methods for HCP analysis. We will review the major methodologies first, and then present the details of a 2D-LC-MS method for HCP analysis.

Overview on the Common Methodologies for HCP Analysis and the Limitations

A downstream process in the manufacturing of a biopharmaceutical normally includes multiple steps for the removal of HCPs (*13*, *29*). Although the concentration and composition of HCPs can vary significantly during the process, the general expectation is that the overall concentration of HCPs should decrease during the journey towards the final finishing product. However, some HCPs, depending on their physicochemical properties, may be co-purified and concentrated along with the drug substance itself (*13*, *30*, *31*). Therefore, during process development, appropriate analytical methods (assays) must be employed in order to determine and monitor the levels of HCP contaminants.

The development of appropriate analytical methods for HCP analysis has always been, and still is, focused on the identification and quantitation aspects of the assay. An optimal method must provide a quantitative measure,

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

have the capability to recognize as many HCP species as possible, if not all of them, and have a short analysis time. A range of methods for the detection and characterization of HCPs have been reported and explored in the biopharmaceutical industry (*2*, *4*, *20*). In general, the methods used for HCP analysis can be classified as either immunospecific methods, for example, ELISA and western blotting, or non-specific methods, for example, electrophoresis and MS. Among the reported methods, enzyme-linked immunosorbent assay (ELISA), gel electrophoresis (GE), western blotting and mass spectrometry (MS) are among those in widespread use (*32*). There are also other less common techniques developed for either on-line or offline detection of HCPs, and elaboration on these techniques is out of scope of the chapter. Several review papers covering these HCP methodologies were recently published. (*3*, *4*, *20*, *33*, *34*)

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Enzyme-Linked Immunosorbent Assays (ELISA)

Among all the analytical methods for HCP analysis currently available, ELISA is the most widely used and is considered the gold standard (*1*). Such assays are widely applied during process development, for example, for demonstrating manufacturing consistency (*35*), for routine quality control (*36*), and for final product release testing (*37*). Traditional ELISA uses antibodies and color change to detect and quantify a particular protein antigen present in a liquid sample. In contrast to conventional ELISA, anti-HCP ELISA measures not a single but a population of HCP species in biological samples through the application of polyclonal antibodies. Consequently, the signal acquired from anti-HCP ELISA measurement is from a number of different proteins present at low levels and reflects the collective sum of immunoreactive proteins.

The principle of ELISA for HCPs is based on specific immunoreactivity between the polyclonal antibodies and HCPs. In essence, ELISA methods attempt to measure the entire epitope population in a sample and report HCP content, expressed in units of ng/mL or parts per million (ng of HCPs per mg product). In this sense, the performance of HCP ELISA assays depends on the quality of the antibody reagents, and the assay only provides results on overall HCP population. Nevertheless, the value obtained in an ELISA assay can be a reasonable measure of HCP mass, provided that the antibody reagent and immunoassay procedures are consistent and shown to accomplish the intended purpose. The primary usage of ELISA often includes the demonstration of consistency of manufacturing processes (*36*), characterization of product purity (*35*), or monitoring host cell protein levels at several stages in biopharmaceutical development and commercialization (*37*).

The development of HCP immunoassays largely depends on the successful generation of anti-HCP antibodies, which is critically influenced by the quality with which HCPs are prepared for anti-HCP antibody generation. The most relevant samples to serve as immunogens for raising anti-HCP antibodies are protein samples that are present in the biologic manufacturing process. However, it is practically very challenging to prepare samples that contain only HCPs from a production cell line with minimal quantity of therapeutic protein (below a few ng in mg of HCPs). As a result, most anti-HCP antibodies are generated from

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

immunogens prepared from null cell lines, expression systems that have not yet been engineered to express the biologic product of interest. In yet another approach, downstream steps in the manufacturing process are used to enrich HCPs specific for a given process ("process specific" HCPs) (*3*).

ELISA methods commonly employed to quantify HCPs in biopharmaceuticals are broadly grouped into two categories: "generic" ELISA assays and process-specific ELISA assays (*3*, *38*). Whether a generic or a process-specific ELISA assay is adapted to determine HCP concentration depends on the development stage for the biopharmaceutical produced or the phase of clinical studies being performed with the materials. Generic HCP assays are supposed to measure all host cell proteins that could be present in a given cell line (e.g., CHO) cells — independent of the downstream processes used (so called platform-based approach). Such assays are based on polyclonal antibodies developed against a representative HCP preparation (normally whole cell extracts), so they have broad specificity and react with a wide variety of HCPs. Nevertheless, insufficient sensitivity with generic HCP assays and a lack of immunoreactivity with certain HCPs have been reported (*3*, *38*).

The range of protein expression profiles from varied manufacturing processes justifies development of a process-specific method for HCP detection. Process-specific assays use polyclonal antibodies produced against HCPs sampled from a given mock fermentation along with a conditioning or purification step. Such assays are usually highly specific for their corresponding HCP composite, so they are generally anticipated not to react as broadly as generic assays for an accurate determination of HCPs from different production processes. The evolution of this type of analytical development has been predominantly industry driven, because it has been essential for the development, optimization and validation of proprietary biopharmaceutical purification processes. Process-specific HCP assays are in general targeted to be in place prior to the initiation of phase III clinical trials. In contrast, generic HCP assays are often used in early phases of clinical product development, when the risk of drug failure is still high and the cost of developing a process specific assay is not yet justified. Under these circumstances, it will be very costly to develop a process-specific HCP program for each biologic molecule in the pipeline because the high labor requirement and long timelines involved for each separate program.

Although ELISA based methods for HCP determination have been widely adopted, it is important to recognize that the use of an immunoassay to measure HCPs has both advantages and disadvantages. A clear and important strength of ELISA based methods is that they can be very sensitive, capable of detecting host cell proteins in the subnanogram range, which can translate to levels of parts per billion (ppb) in final products (*26*). Because an ELISA method measures the response from a number of different proteins present at low levels, it is generally believed to be more sensitive than methods for monitoring individual HCP. This is especially useful when a recombinant product is of high purity, and there is insufficient quantity of any particular HCP species to be otherwise detected by orthogonal analytical techniques. Additionally, ELISA based methods are easily transferrable to other test sites, can be performed under GMP (good manufacturing practices) conditions, and run in a high-throughput manner (e.g. 96-well plates).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

These attributes are among the key consideration for the implementation of biopharmaceutical analytics in the industry.

The key disadvantage of ELISA for HCP analysis is that it is extremely difficult to demonstrate the HCP coverage of an ELISA measurement against all potential HCPs. As indicated above, ELISA methods almost always rely on the ability of an anti-HCP antiserum to detect HCP impurities, so the performance of the assay is tightly linked to the quality of this reagent. For example, non-immunoreactive or weakly immunoreactive HCPs cannot be detected by ELISA. To detect such proteins, an orthogonal method must be used, for instance, SDS-PAGE with a sensitive protein staining method (e.g., silver or Sypro Ruby) (*39*) or LC-MS (*40*). Secondly, ELISA provides no HCP identification and gives no information on the distribution of HCPs that are actually present in a particular protein preparation. The absence of such information is particularly disadvantageous with process-specific ELISA assay, since the assay is developed to monitor HCPs for pre-defined bioprocess conditions. Since the levels of individual HCP may be independently affected by adjustment in cell culture and purification conditions, subtle process deviations may result in the presence of an HCP species in the drug substance that an ELISA-based HCP assay is incapable of detecting. Clearly, ELISA based methods, by their nature and how they are developed, may "miss" potential protein impurities. For this reason, there is a need for orthogonal methods of HCP analysis.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Sodium-Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

For decades, polyacrylamide gel electrophoresis (PAGE) has been widely used as a technique in biopharmaceutical development laboratories to characterize protein mixtures, estimate protein molecular weight and isoelectric point, monitor protein purification and determine the heterogeneity of proteins present in a sample (*41*). The popularity of gel electrophoresis can be largely attributed to its high resolution, its ease of and near-universal use for protein separation. There are many different forms of gel electrophoresis including native PAGE, SDS-PAGE, and isoelectric focusing (IEF). Two dimensional PAGE (2D-PAGE) describes a form of gel electrophoresis that orthogonally couples IEF and SDS-PAGE techniques together in a two-step process for simultaneous resolving complex protein mixtures based on charge and size (*42*, *43*). An extension of 2D-gel electrophoresis is two dimensional difference gel electrophoresis (2D-DIGE), which can resolve multiple samples on the same gel using spectrally resolvable, size- and charge-matched fluorescent dyes (*44*). This technique is a major workhorse to find the differences in the protein expressions since it provides a convenient methodology to compare the gel images from two or more samples (*45*).

SDS-PAGE is an important analytical method for both process development and final product quality control (QC) of biopharmaceuticals. For HCP analysis, because of sample complexity, one dimensional (1D) gels generally do not have sufficient resolving power to separate individual HCP in a single gel and provide a profile for the HCP composition. One common use of 1D gels is to combine with Western blotting for the evaluation of the sensitivity of the selected anti-

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

HCP antibodies (*46*, *47*). Because 2D electrophoresis separates proteins based on two orthogonal physicochemical properties, isoelectric point (first dimension) and molecular size (second dimension) (*42*, *43*), it possesses an increased capability to resolve different HCPs in a single gel and is therefore more frequently used during up- or downstream process development and characterization (*48–53*). A 2D-PAGE gel can resolve 1000's of proteins in a single run, separating the proteins into spots in the gel. The detection of HCPs by gel electrophoresis is limited to proteins with molecular weights between 5 and 1000 kDa (*54*).

Gel electrophoresis is normally used as a complementary method to ELISA methods for the detection of non-immunogenic HCPs (*43*). One of the major applications of 2D-PAGE for HCP assay is the combination of 2D-PAGE and Western blotting to evaluate anti-HCP antibody reagents. This combination enables facile visualization of total HCP populations and facilitates assessing the percent coverage of HCPs that are immuno-reactive by an anti-HCP polyclonal antibody reagent (*55*). A typical workflow for such evaluation consists of the following steps - (1). HCP samples are separated via 2D-PAGE in replicate gels; (2). One of the 2D-PAGE gels is stained using a silver staining method for total HCP profile; (3). HCPs in the other 2D-PAGE gel are transferred to a solid support membrane such as nitrocellulose or polyvinylidene fluoride (PVDF) for Western blotting, with the polyclonal anti-HCP antibodies being evaluated; (4). Images from 2D-PAGE silver staining and Western blot membrane are overlaid and compared to yield a match ratio for the determination of HCP percentage that anti-HCP antibodies can detect. Although the procedure is laborious, it is widely used by the industry and provides a sensible measure to the quality of anti-HCP antibody reagents. Nevertheless, the overall reliability of final results is affected by experimental variation inherent to comparisons between replicate 2D-PAGE analyses (where it is challenging to ensure direct spot-to-spot comparison between two separate gels).

Resolved HCPs are often visualized in a gel by a staining method using either metals, dyes, or fluorescence (*56*). The most commonly used stains are silver and Coomassie Brilliant Blue. The sensitivity of Coomassie Brilliant Blue is in the range of 50-100 μg per band or 2D spot. Silver staining is 100-fold more sensitive than Coomassie Brilliant Blue. With a more sensitive silver-staining method, a HCP with less than 100 pg per band or spot can be detected in the SDS-PAGE gel. Coomassie Brilliant Blue staining only has 10- to 100-fold dynamic range (*53*), whereas the linearity range for sliver staining is about 40-fold. Fluorescence labeling is often found in 2D-DIGE technology, where the introduction of fluorescent reagents for protein labeling has brought substantial improvement in the reproducibility of the analysis, thus facilitating the comparison of protein patterns in gel images from multiple samples. 2D-DIGE allows multiplexing of up to five samples in one gel (*48*, *57*), provides higher sensitivity (with a detection limit of approximate 5 ng of protein), and a wider linear range ((*3*, *4*) orders of magnitude) for quantitation compared to Coomassie Brilliant Blue and silver staining methods. Similar levels of sensitivity are achieved by silver or Sypro Ruby detection (*58*). Although Sypro Ruby staining offers a much wider dynamic range than silver staining, that range is generally too narrow to provide the levels of residual HCPs present relative to product concentration as individual

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

HCP is typically present at 5-6 orders of magnitude lower concentration than the product. Additionally, staining methods do not provide the identities of HCPs, and frequently specific identification and still better sensitivities are required for gel electrophoresis as an analytical technique for HCP assays. For that reason, the immunological analysis of HCPs has proven to be indispensable as a complementary technique to gel electrophoresis. The gel spots can also be quantified and analyzed by mass spectrometry (*59*). Typically, the gel spots are excised, an in-gel digestion is performed to convert HCPs into peptides, and peptides are subsequently extracted for LC-MS analysis. MS response of the extracted peptides can provide a measure of protein abundance (e.g. using a spiked, isotope labeled peptide with the same sequence as an internal standard or a Hi3 method for quantitation, see below), and tandem mass spectrometry of the peptides can be used to confirm the identity of the protein spots/bands. In general, MS coupled with a nanoflow LC separation is capable of identifying proteins excised from a gel spot/band from Comassie Brilliant Blue or silver staining methods.

In comparison with ELISA, gel electrophoresis does not require time-consuming production of antibodies nor does it need the development of a process-specific ELISA method. As a protein separation technique, 2D-PAGE not only resolves a large number of proteins in a single run, but staining these proteins enables the relative abundances of the proteins to be quantified. However, 2D-PAGE also has certain limitations and one of the drawbacks is sample throughput (*60*). 2D-PAGE is a relatively slow, labor-intensive technique, making it difficult to automate and make amenable to high throughput analyses. Single proteins, when analyzed with 2D-PAGE, often show multiple spots due to post-translational modifications (e.g. glycosylation, deamidation, sequence extensions/truncations). This leads to added complexity in the analysis. Other potential disadvantages include large amounts of sample handling, limited reproducibility, and poor resolution of low abundance proteins, hydrophobic proteins and proteins with extreme pI values and/or molecular sizes. With the continuous technical improvement on 2D-PAGE, this limitation is gradually diminishing (*60*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Western Blotting Method

Western blotting is an antibody-dependent immune detection method (*58*) that is widely used to detect specific proteins in a given sample. This blotting technique is used to establish protein identity and purity with respect to host cell proteins. HCP samples containing the protein of interest, in either native or denatured states, are first separated with either one dimensional or two-dimensional gels, and are then transferred or blotted onto the surface of a second matrix, generally a nitrocellulose or polyvinylidene difluoride (PVDF) membrane. Typically, the membrane is then blocked with BSA or other non-signaling proteins to prevent any nonspecific binding of antibodies to the surface of the membrane. Next, the HCPs are then identified using antibodies specific to the target protein which can be detected at the site of the protein-antibody complex on the membrane. In general, there are two ways to detect the HCP-antibody complex: (1) using the

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

primary antibodies raised against HCPs, which can be directly labeled with an enzyme like Horseradish Peroxidase (HRP) or a fluorescence molecule (*61*) (2) detected indirectly with a labeled secondary antibody that specifically recognize the primary antibody (*58*).

Western blots are frequently used for confirmation of the selectivity of the antibodies used for HCP analysis (*37*, *52*, *62*, *63*), usually in combination with non-immunospecific detection methods, such as SYPRO Ruby or silver stain. It is also used for process characterization (*63*), where it is not only applied to detect or monitor HCPs, but also to provide characterization of the HCPs with regards to size (10 - 1000 kDa). In comparison with a plate-based ELISA method, and even though Western blotting can use the same anti-HCP antibodies and detect many of the same HCP, the two methods are complementary in that they may be sensitive to different sub-populations of HCPs. In the Western blotting process, a strong detergent, reducing agents, or heat is usually required to solubilize or denature the proteins prior to the electrophoresis step. These harsh conditions can damage some antigenic determinants. On the other hand, some other antibody epitopes that might be sterically hindered from binding will be exposed by denaturation. Samples analyzed by ELISA assays usually do not need to undergo such harsh treatment conditions, so HCPs are measured in relatively much more native configurations.

The advantage of the Western blotting approach is that it can separate and help identify individual HCP. ELISA methods, on the other hand, cannot differentiate or quantitate one HCP from another. The Western blotting approach can detect a protein band down to 1 ng using a typical colorimetric method. The detection limit can be improved by employing photon detection techniques such as radioactivity (In this detection method, the gamma-emitting radioisotope ^{125}I is used to label lysines in antibody. Upon binding, radiolabeled blots can be detected using X-ray film, a method known as autoradiography) or chemiluminescence with the limit of detection at around 100 pg per band, but the limit of detection of Western Blotting is essentially decided by the sample size that can be analyzed. In comparison, the detection limits for ELISA method is typically about 100 fold more sensitive than Western blotting. This is most likely because the overall response is distributed over a range of different protein bands, which could lead to signals that are too weak for reliable detection. Despite of its popularity, however, Western blotting has several disadvantages. As previously mentioned Western blotting is a time-consuming (compared to ELISA) analysis. This method also requires systematic optimization in the experimental conditions (i.e. reagent antibody, protein isolation, buffers, type of separation, gel concentration, etc.) including optimization of electrophoretically transferring proteins out of a gel onto a blotting medium, which depends on the individual proteins and blotting technique. Proteins with high molecular weights do not transfer efficiently and may remain undetected by a Western blot. Depending on its electrical charge, a protein may not be released fully from the gel or bound efficiently to the membrane during the transfer step, resulting in low recovery. In addition, Western blotting membranes are microporous substrates and are available in various pore sizes (e.g. 0.2- or 0.45- µm pore membranes are commonly selected for most analytical blotting experiments). The dimensions of the pores dictate the size

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

of the protein that can successfully bind to the membrane. Very low molecular weight proteins can pass through membrane pores without being sufficiently adsorbed. For instance, 0.45-μm pore membranes are typically recommended for proteins with a molecular weight greater than 20 kDa. Therefore, Western blotting experiments are highly demanding in terms of user experience. Additionally, a Western blot can only be performed if primary antibodies against the HCPs of interest are available.

LC-MS Analysis of HCPs: Methods, Performance Metrics, and Analytical Considerations

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

A truly meaningful assessment of the potential risks associated with residual HCPs in a drug product requires both identification and quantification of the individual HCP present, since each individual protein is likely to affect the safety profile of a drug differently. Therefore, the total HCP levels may be less relevant than the amount of specific, high-risk protein(s). Analytical approaches such as LC-MS are capable of identifying individual HCP through genome/proteome databases, thus providing more definitive information for risk assessment (*4*), process development, and defining control strategies for HCPs.

As a universal detection technique, mass spectrometry (MS) yields both qualitative and quantitative information. Most notably, it is one of the primary tools for determining protein and peptide sequences and for monitoring and identifying residual impurities in the biopharmaceutical industry (*64*). While MS is a particularly powerful tool for biotherapeutic analysis, the complexity and wide dynamic range of HCP samples preclude the direct use of MS for the identification and concentration measurement of low abundance HCPs. Even the most advanced mass spectrometers, with high scanning speeds and high resolving power, are limited in terms of in-spectrum dynamic range. Additionally, when too many analytes with differing proton affinities are concurrently introduced to a mass spectrometer, ionization suppression hinders the identification of low-abundance species and limits the analytical depth to which a sample can be probed.

The use of liquid phase separation(s) prior to MS analysis helps mitigate this issue, because the separation prior to MS detection reduces the number of analyte ions entering the mass spectrometer at any given time. Furthermore, analytes can be focused into narrow bands (or peaks) during the liquid phase separation step(s) and thereby be concentrated (relative to the original sample) prior to MS detection, improving the sensitivity of an analysis.

Liquid chromatography- mass spectrometry (LC-MS) has been recognized as a powerful tool for protein and peptide sequencing and is routinely used in proteomics and biomarker discovery studies, where low-abundance proteins are identified across a dynamic range of 3-4 orders of magnitude in a variety of samples (*65*, *66*). In such analyses, a biological protein sample is first reduced and alkylated, and then digested to peptides using a protease (typically trypsin). The protein digests are thereafter separated by LC, introduced into a mass spectrometer, and fragmented. The parent proteins are identified via searching

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

the peptide fragments against a database of protein sequences. The procedure is frequently referred to 'shotgun' proteomics. This proteomic approach possesses the capability to analyze a large number of protein targets in a single experiment, independent of the availability of anti-HCP antibodies. The distinct capability of an LC-MS/MS (liquid chromatography-tandem MS) approach to identify proteins in complex mixtures has recently led to its consideration as a means to characterize HCPs in biologics (*17–19*, *27*, *28*, *30*, *67*). This is largely because an LC-MS/MS approach can provide more definite answers to questions about the nature of HCPs, such as their identities and concentrations in a given sample (rather than a summed concentration of many HCPs). An LC-MS/MS approach also has the potential to describe a complete HCP proteome in a biopharmaceutical product.

No different than any of the aforementioned HCP assays, the primary objective for HCP analyses by LC-MS/MS is to obtain sufficient depth of HCP proteome coverage (or broadest protein identification) and to quantify the identified HCPs. Such an analytical endeavor is affected by several factors, among which are sample size, the resolution of LC separations, the sensitivity of mass spectrometers, modes for data collection, the criteria used to identify proteins from collected data, and the concentration ranges of proteins in samples. Sample size is typically of minimal concern for the biopharmaceutical industry since a bioprocess results in sample quantities that exceed the need for a typical LC-MS/MS experiment. The ideal MS platform as well as the acceptable criteria used to identify proteins in a proteomics setting has always been under constant debate in proteomics research, and detailed discussion on the best practice for these options is beyond the scope of this chapter. Instead, the focus of this section is to discuss how the final outcome of HCP proteome coverage is impacted by the resolution of LC separation methods, the ranges of protein abundances in samples, and the modes of data collection.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Performance of LC Separations and Analysis Coverage of HCP Proteomes

The implementation of LC techniques for the study of HCP proteomes generally falls into two groups — those using one-dimensional LC (1D LC) and those using two-dimensional LC (2D LC). A one-dimension LC separation is one of the basic formats in LC-MS setups for proteomics analysis as well as the characterization of biotherapeutic proteins. Reversed phase (RP) is a prevalent LC mode to couple with MS for separation of peptides, because it offers relatively high resolving power and RP-LC mobile phases are most compatible with electrospray ionization (ESI) for subsequent MS detection. To handle a complex protein digest mixture with thousands of peptide components at different concentration levels, the utmost requirement for RP-LC separation is to provide resolution among peptide species in order to facilitate the identification and quantification of the component peptides by mass spectrometry. Gradient elution is typically used for RP-LC separations of peptides and proteins. The most commonly used measure for the performance of a gradient separation is peak capacity (the maximum number of components that can be resolved in a given gradient time (*68*). Generally speaking, the higher the peak capacity, the higher the probability of separating all components in a sample. For a packed RP-LC

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

column, the peak capacity (Cp) that can be achieved for gradient separation of peptides depends on column length (L) and the particle diameter of the stationary phase (dp). The relationship between these variables and peak capacity can be simplified as follows (*69*):

$$Cp \propto L^{1/2} \text{ (for a defined particle size)}$$

$$\text{and } Cp \propto (1/dp)^{1/2} \text{ (for a constant column length)}$$

The first relationship highlights how an analyst can improve peak capacity by increasing the length of LC columns. The second relationship defines how significant gains in peak capacity can be obtained through the use of columns packed with small diameter particles. When measuring the peak capacity for two different columns of the same length but packed with 3.5 µm versus 1.8 µm particles, Gilar et al (*70*) reported that a 40% gain in peak capacity can be achieved with the 1.8 µm particle column. The use of LC columns packed with sub-2-µm diameter particles, typically termed an ultra high performance liquid chromatography (UHPLC) separation, has been a commercial success for more than a decade and a reliable means of increasing peak capacity in innumerable separations (*71*, *72*).

For a well-packed 100 mm, 3.5 µm particle column, a peak capacity of 250 can be obtained from a 1D separation and a 45-minute gradient. Peak capacities up to 1000 with extended column lengths (e.g. 800 mm) are also reported in the literature (*73*). However, the use of columns with such long lengths normally requires a corresponding adjustment in separation gradient time. For example, when a 0.15 × 800 mm column packed with 3-µm C18-bonded porous particles was used to separate peptides from a cell lysate digestion, the separation required more than 3 hours (*74*). Although a higher peak capacity can be achieved by extending the column lengths and gradient time, the number of separated components per unit time, or the productivity of separation, can be rather low. In addition, the number of peptides identified in LC-MS/MS experiments does not proportionally increase with augmented peak capacity, especially for samples with a wide dynamic concentration range (*73*) such as biotherapeutic samples containing low abundance HCPs. This observation is largely due to the fact that extended gradients dilute peak volumes, decreases analyte concentrations and thus MS responses of eluting peptides in LC-MS/MS analysis. Since peptides from HCPs are typically in low abundance, such a decrease in concentration would negatively impact the quality of fragmentation spectra, on which identifications are typically made.

Alternatively, it is often useful to improve chromatographic peak capacity through the use of 2D chromatography, where a separation is achieved based on the coupling of two separation mechanisms, two that are ideally orthogonal. Since RP-LC separation is generally preferred when coupling to MS, almost all 2D LC proteomics analyses to date have been based on a RP second dimension in combination with another LC mechanism for fractionating samples in a first dimension separation. In this sense, choice of another LC dimension is mainly determined by its orthogonality to the 2nd dimension RP separation.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Orthogonality is largely determined by the analyte retention mechanisms imposed by different separation modes. Although a variety of physicochemical properties that contribute to peptide retention on LC columns can be exploited to design different 2D LC systems, true orthogonality is difficult to achieve because no LC separation is achieved completely based on a single property alone. Correlations among molecular weight/size, hydrophobicity/polarity, charge, and isoelectric point (pI) of peptides/proteins reduce the effective orthogonality of any two separation mechanisms. Nevertheless, Gilar et al reported several 2D configurations, including a system consisting of high pH RP coupled to low pH RP (so called 2D-RP/RP), can still achieve a high-resolution separation despite that the coupling lacks complete orthogonality (*75*).

The peak capacity of an ideal 2D system (that is based on true orthogonal 2D separation) can be calculated as the product of individual peak capacities from each dimension ($Cp_{2D} = Cp_{1st} \times Cp_{2nd}$). This formula is applicable when the 2D system is operated such that each dimension can reach the optimal peak capacities without much concern of analytical throughput. In reality, in the interest of maintaining analytical throughput for advanced MS analysis of complex mixtures, the first dimension undergoes only a limited number of fractionation steps (e.g. < 20) rather than a high-resolution gradient separation. Consequently, the total peak capacity obtained from 2D LC analyses is usually much lower than the theoretical one. Therefore, the overall peak capacity is more appropriately calculated as the product of the number of fraction steps in the first dimension and the peak capacity of the second dimension. Based on this assumption, a 2D-LC system that undergoes 10-step fraction in the first dimension can generate a total peak capacity of 2500, which is a peak capacity 10 times that obtained with a typical one dimensional RP separation (take 250 as a typical peak capacity in 1D separation). This value can readily surpass the peak capacity acquired from a long column operated at an extended separation time. In addition, the use of a short but efficient column for the second dimension separation in a 2D configuration can improve the productivity of a separation and peptide identification, since relatively fast gradients can be applied without comprising the peak capacity of the system.

For peptide separations, the most widely practiced 2D-chromatography separation schemes are based on the coupling of strong cation exchange (SCX) and low-pH RP chromatography (*76*). Peptide fractions are eluted from the first-dimension SCX column by applying a series of step gradients of salt solutions with increasing ionic strengths. At each step, peptides are subsequently separated on the second dimension RP column using a linear acetonitrile gradient. In principle, retention in SCX is driven by an analyte's charge. Because the majority of tryptic peptides in a protein digest are doubly and triply charged at the pH used for the SCX separation, the distribution of peptide fractions is bimodal, with 2+ and 3+ charged peptides eluting in clusters. As a result, significant splitting of high-abundance peptides across multiple fractions has been reported (*75*). In addition, SCX separations tend to suffer from poor reproducibility and significant peptide losses due to secondary hydrophobic interactions originating from SCX sorbents (*75*).

As mentioned before, a different 2D-LC methodology has been developed, based on a RP separation under basic (pH 10) conditions in the first dimension

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

followed by a low-pH RP separation in the second dimension (*77*, *78*). It has previously been demonstrated that pH can significantly alter the selectivity of peptide separations in RP chromatography (*75*, *77*, *79*, *80*). This observation prompted significant interest in coupling two RP columns (*75*, *77–81*), operated at two pH extremes (pH 10 and pH 2.6), where the ability to robustly perform first dimension pH 10 peptide separations was made possible by development of organosilica hybrid C18 stationary phases, most notably ethylene bridge hybrid materials (*81*). Although the coupling of high-pH RP/low-pH RP separations was shown to be less orthogonal than a traditional SCX/RP multidimensional system for the separation of complex peptide mixtures in proteomic experiments (*75*), the separation resolution offered by the high-pH RP in the first chromatographic dimension is far superior to the SCX separation. A RP separation elutes peptides almost equally over the entire retention window (trapezoidal distribution of peptides) allowing for a greater spread of peptides across the same number of fractions (*79*). Taken together, the advantages offered by a 2D-LC system based on high-pH/low-pH RP separations translate into better chromatographic performance. In the case of an HCP assay, these chromatographic advantages can help to provide more HCP identifications and/or better sequence coverage for low abundance HCPs.

The use of 2D-LC for LC-MS analysis has the advantages of resolution, efficiency and sensitivity. However, in comparison with one-dimensional LC (1D-LC), 2D-LC increases the time needed for analysis (lower throughput) and requires more sophisticated instrumentations. 1D-LC methods offer better throughput and instrument robustness. As such, the selection of an LC-MS-based HCP analytical strategy should be made based on the fundamental understanding of the limitation and advantages of each LC operation modes. The use of 2D LC-MS methods can potentially lead to the identification of more HCPs and may be desirable for analyzing highly purified drug products; whereas 1D LC-MS offers can be used as a higher-throughput screening tool, e.g., to test therapeutic samples for process development. These samples typically contain much higher levels of HCPs than a final drug product.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Sample Loading Capability of LC Columns and Coverage of HCP Proteome

As reported in literature, the majority of HCPs remaining in biopharmaceutical samples are attributed to a weak interaction between the HCPs and therapeutic proteins rather than an HCP-resin association (*82*). The potential for HCPs to bind to or in some fashion associate with the therapeutic protein can greatly complicate biopharmaceutical process development and implementation. In this scenario, it is important to analyze the sample without depleting the therapeutic protein (or HCP enrichment) as a depletion approach could potentially bias the native HCP profile and result in an HCP distribution that is not a true representation of HCPs in the sample. Accordingly, analytical approaches that can produce HCP profiles without needing to remove therapeutic proteins from samples are always preferred and advantageous.

In this regard, achieving higher peak capacity in LC separations is important for the analysis of HCP samples because the resolving power can render a greater

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

chance of separating low-abundance HCP peptides from other components, most critically the peptides from the overwhelmingly abundant therapeutic protein in a biopharmaceutical substance/product. Resolution of these low abundance peptides minimizes ionization suppression, spectral crowding of MS and MS/MS spectra, and the frequency of where nominally detectable species are not detected due to detector dynamic range limitations or are under-sampled in ion selection for MS/MS fragmentation in 'shotgun' type of experiments.

The identification of low abundance HCPs from a complex protein digest mixture is further influenced by other factors. Among them, LC column sample capacity is a critical factor. Sample capacity is important to consider in HCP analysis as it is a metric that defines how much sample can be injected onto a column before overload (peak broadening/tailing) is prominently observed. Biotherapeutic protein samples for HCP analysis have unique properties in terms of protein concentration distribution. First of all, the concentrations of individual protein in the samples have a very wide protein concentration range, spanning at least 5-6 orders of magnitude. This is different from typical proteomics samples, where multiple components may be present at relatively higher levels and the distribution of protein concentration for the proteome is narrower (Plasma proteome is an exception). Secondly, despite the wide concentration range, the concentration of individual protein in the samples is drastically different, as therapeutic protein accounts for the majority of the protein mass (typically > 99 %). The concentration range of total HCPs in most samples constitutes only the bottom 1-2 orders of the concentration range (1 -100 ppm), and there are no or very few other protein species with concentrations greater than 100 ppm. This characteristic of an HCP sample highlights the major difference in sample compositions encountered in HCP versus generic proteomics analyses, and why there is a need for LC-MS/MS methodologies in HCP analysis that possess the capability to identify and quantify proteins at very low concentrations while probing deep into the trace level and sparse host-cell proteome present in a biotherapeutic sample.

At the very basic level, the dynamic range that a proteomics analysis can cover is determined by the detection limit of a mass spectrometer, the amount of sample injected for LC-MS analysis and the electrospray ionization efficiency. The use of more sensitive mass spectrometers can certainly help to measure and detect low concentration peptides, thus improving the proteome coverage. However, for a given mass spectrometer under an optimized electrospray ionization condition, the dynamic range is mainly determined by the amount of sample that is loaded onto and separated by a LC column. In theory, the more materials loaded onto a system, the higher the concentration gets for the lower abundance peptides in the RP-LC eluent - until the column sample capacity is reached. Therefore, a common tactic in proteomics research is to use a large sample size to increase the concentration of low abundance proteins above identification thresholds. This approach can aid identification of the low concentration proteins. But for biological samples with a wide range of protein concentration, peptides from the higher abundance proteins ultimately overload the column, and the number of proteins identified, or the proteome coverage, will not necessarily increase continuously with the increase in sample loading. From a 12-hr LC-MS analysis experiment, only 853

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

proteins can be identified in human plasma (*69*), which represents one of the most complex proteomes given its range of protein concentrations that span 10 orders of magnitude. Meanwhile, a similar experiment can be used to identify >2000 proteins in *S. oneidensis* proteome (*74*, *83*). In this case, the distribution of protein concentrations in the sample is a key factor that decides the achievable coverage. It is reported that a 50 μm × 400 mm capillary column (packed with 1.4 μm porous C18 particles) can chromatograph 10-100 μg of protein digests before the column sample capacity is reached (*69*, *84*). With this amount of sample, a proteomics LC-MS/MS analysis can cover a range of protein concentrations of about 4-5 orders of magnitude (*85*, *86*). If less material is injected onto a larger size (internal diameter) capillary column for a one-dimensional RP separation (e.g. 1ug of materials injected onto a 75 μm × 200 mm column), the dynamic range of LC-MS/MS experiment is expected to be 3-4 orders of magnitude. This implies that an analysis based on a one-dimensional RP separation is likely to reliably identify HCPs down to a level of about 100 ppm (molar). This conjecture is supported by two recent publications (*27*, *67*) on HCP analysis using one-dimensional LC-MS/MS. Thompson et al (*67*) reported the dynamic range of 1D LC-MS-based HCP analysis was only about of 3 orders of magnitude (*103*) when 0.5 μg of protein digest was loaded onto a 100 μm C18 column for HCP analysis. In a separate report by Bondarenko et al (*27*), 1D nanoflow LC coupled with an ion-trap mass spectrometer identified two HCPs from an antibody drug product with the lower HCP abundance being 88 ppm.

On the other hand, when sufficient amounts of material are available, 2D-LC offers a more flexible and capable mechanism to detect low abundance proteins. In 2D-LC, samples are separated into multiple fractions prior to LC-MS analysis. As a result, the amount of the sample that can be loaded in the analysis depends to a large extent on the column dimension in the first dimension, which can be much larger than the second dimension column and therefore used to load large quantities of sample without concern of column breakthrough (no adsorption). In addition, the pre-fractionation and corresponding reduction of sample complexity is highly beneficial to most analyses. In general, each fraction contains a "simplified" mixture of peptides, enabling identification and possibly quantitation of more peptides and their counterpart proteins, including those at low abundance. At the same time, fractionation can provide information about analytes without any additional analytical effort. This information can be used together with the tandem mass spectrometry data in the validation of peptide-spectrum matches (*87*).

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Data Collection Mode and HCP Proteome Coverage

The aforementioned characteristics of biotherapeutic protein samples suggest that, in order to detect low abundance HCPs by LC-MS/MS, it is normally unavoidable to overload the chromatographic system with peptides from the therapeutic protein itself. Based on this consideration, a 2D-LC system paired with a modern MS instrument and a 75 μm ID, sub-2μm particle column in the second dimension, typically requires about 10 μg of therapeutic protein digest to be loaded onto an appropriately sized first dimension column (i.e. a 300 μm

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

ID column) in order to identify HCPs that are present at low single digit ppm concentrations in a monoclonal antibody sample (20 mg/mL) (*88*). In such an analysis, about 220,000 ions (excluding background ions, roughly 82,000 components) are generally produced and detected by MS in a single-step fraction. According to theory (*89*, *90*), the peak capacity should exceed the number of components in a sample by a factor of 100 if 98% of those components are to be resolved. However, the peak capacity of such a 2D-UHPLC system is often less than 5000, falling short of the required degree of separation. These observations clearly indicate that many components are co-eluting in a single chromatographic peak during LC-MS analysis of HCPs.

Co-elution of HCP peptides with peptides from the high abundance therapeutic protein creates a significant challenge for MS detection of the low abundance HCP peptides. Multiple high-abundance peptides entering the mass spectrometer at any given time can saturate the instrument's detector, creating interfering "ringing" responses, or bias the ion-influx measurement in ion-trap instruments to properly gate ion trapping time for the detection of HCP peptides. Although high-resolution MS instrumentation does not need all components to be individually separated, prior to analysis, to be detected. each peptide precursor ion still needs to be individually selected by the mass spectrometer for fragmentation in a process referred to as data-dependent acquisition ('DDA'). The rate at which a mass spectrometer can perform the fragmentation (i.e. switch between MS and MS/MS) determines the sampling depth and the dynamic range of the analysis. Because the concentration of the biopharmaceutical molecule is at least three orders of magnitude higher than the HCPs, the signals of the peptides from the biotherapeutic protein dominate an LC run. As a result, peptides from HCPs frequently co-elute with multiple peptides present at much higher abundances. This situation quickly overwhelms the MS/MS acquisition rate of even the fastest instruments, because peptide precursors are selected for MS/MS fragmentation based on their intensities. Since the low intensity ions are those of interest for HCP identification, the use of data-dependent techniques results in preferential fragmentation of high-abundance peptides (coming from the biopharmaceutical), with less chance for the fragmentation of HCP peptides. In a data-dependent approach, the MS instrument is greatly biased toward "uninformative" peptide fragmentation of the biopharmaceutical. This results in a reduced number of HCP peptide identifications, and greatly increases the variability of results from LC-MS analyses (*91*, *92*).

A commonly used strategy to increase peptide sampling is the use of the mass exclusion list combined with DDA. This strategy relies upon peptide mass information from previous analyses to dictate the data acquisition of the current run. A mass exclusion list contains the masses of previously identified peptides and their retention times, so that when ions in a particular elution time window are being ranked for potential CID spectrum acquisition, peptides on the exclusion list are discarded, and new, less intense precursors undergo fragmentation. This results in a secondary set of peptide identifications that is largely novel. The masses of these new peptides are then appended to the mass exclusion list, and additional LC-MS/MS analyses are iteratively performed until the desired degree of characterization has been achieved, or until no additional unique peptides are

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

identified. While this approach is effective with regard to improving proteome coverage via increasing the number of peptide identifications, it is also clear that there remain significant limitations upon its ability to exhaustively fragment all the ions generated during a LC-MS analysis. It was reported that only ~20% of all features are sampled even after 6 runs in a 90-minute gradient for HeLa S3 cell lysate (*93*). In addition, the efficiency in terms of instrumentation time for generating meaningful HCP identification is greatly reduced because the majority of identification is made for therapeutic proteins. The implementation of this strategy into acquisition methods is also problematic because LC-MS runs, database searches, and preparation of the exclusion lists become segmented in time.

Compared with data-dependent acquisition, data independent acquisition methods, such as multiplexed data acquisition (MS^E), have proven to be advantageous to provide in-depth sampling (*94–98*) and operation efficiency. Such methods do not involve any precursor selection during the fragmentation process. All peptide precursor ions, whether they have high or low intensity, have an equal opportunity to be fragmented, thus offering a higher duty cycle, improved chromatographic peak sampling, and richer, more reproducible mass spectra (*97*, *98*). As a result, this technique provides an efficient way to sample the low intensity peptide precursors when there is a wide range of peptide abundances. For the identification of HCPs in biopharmaceuticals, where low abundance proteins are the subject of interest, multiplexed MS^E methods seem to be a rational choice. Along with 2D-LC, MS^E can provide reproducible, high-sensitivity MS and tandem MS data for all peptides eluting from the chromatography column. Additionally, MS^E has also shown a potential to offer not only to be used to identify proteins but also to quantify them within a single analysis via the use of the summed intensity of "Hi3" tryptic peptides of every identified protein. Hi3 peptides reference those three tryptic peptides that produce the top 3 most intensive molecular ions for a protein in mass spectrometric analysis. It was previously demonstrated that the Hi3 peptide responses for any given protein is similar; therefore, the summed intensity of Hi3 peptides from a spiked protein standard with a known amount may be used to estimate the molar amount of any protein in a complex digest mixture (*99*, *100*). This implies that *a priori* unknown proteins might be identified and quantified, in a single experiment, based on the calibration of their Hi3 peptide intensities with those of spiked protein standards. Such broad, label-free quantification represents a powerful approach in quantitative protein profiling.

Application of 2D-LC-MS^E Methods for HCP Analysis in Support of Bioprocess Development

A number of reports on the use of 2D-LC-MS for HCP analysis have appeared in the literature since the first paper on this subject was published in early 2012 (*19*). These applications have touched upon different areas of bioprocess development, including the identification and quantitation of HCPs from samples prepared from different purification protocols with protein-A (*19*),

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

tracking the HCP profile changes across multiple-stage purification steps (*30*, *101*); and evaluating the effects of process changes on residual HCP profile to facilitate bioprocess development (*17*, *18*). More importantly, these studies have demonstrated the ability of the 2D-LC-MS^E approach to identify, and quantify, individual HCP present in biotherapeutic proteins in an objective manner. In the following section, several case studies (Case I, II and III) are presented that demonstrate how the HCP information acquired by 2D-LC-MS approach is used to provide comprehensive, and accurate, characterization of HCPs in biopharmaceutical samples.

Case I: Identifying HCPs in NISTmAb by 2D-LC-MS^E and HCP ELISA-Assay: Protocol and Summary Results

Detailed protocols are provided on the implementation of high pH/low pH 2D-LC-MS^E for analysis of HCPs in NISTmAb.

Sample − NISTmAb

The NISTmAb sample provided (Candidate RM 8670, lot 3f1b) is a IgG1k monoclonal antibody produced in a murine myeloma cell line. The sample has undergone industry standard upstream and downstream purification to remove process related impurities. The material was provided for analysis at concentration of 100 mg/mL in 12.5 mM L-Histidine, 12.5 mM L-Histidine HCl (pH 6.0).

Sample Preparation − Trypsin Digestion Protocol for NISTmAb

Twenty-five microliters (25 µL) of NISTmAb solution (100 mg/mL) was drawn and mixed with 626 µl of 50 mM ammonium bicarbonate (NH_4HCO_3) solution. The resulting protein solution was mixed with 40 µL of 1.0 % (w/v) RapiGest (Waters Corp, in 50 mM ammonium bicarbonate). The mixture was incubated at 60 °C for 15 min to denature the proteins. After denaturation, 34 µL of 500 mM DTT (in 50 mM NH_4HCO_3) was added to the vial (the final DTT concentration: 25 mM DTT) and the proteins were reduced for 30 min at 60 °C. After cooled down to room temperature, the sample was then mixed with 35 µl of 500 mM iodoacetamide (IAM in 50 mM NH_4HCO_3) to alkylate the proteins at room temperature in the dark for 30 min. The sample was finally quenched with 20 µl of 500 mM DTT (in 50 mM NH_4HCO_3) before proteolysis by Trypsin. Enzymatic digestion (Enzyme: Substrate = 1:17) was performed at 37°C overnight with porcine trypsin (Promega, Madison, WI, USA). After digestion, the RapiGest surfactant was decomposed by adding 5 µL of pure TFA and the samples were incubated for 30 min at 37 °C, and centrifuged (10 min at 10,000 rpm) to remove the insoluble component of the degraded RapiGest. After adjusting the pH of the supernatant solution to pH 10 using 25 µL of 28% (w/w) ammonium hydroxide (pH 11), the digestion solution was added to 100 µL of 0.2 mM ammonium formate (pH 10). Finally, the digestion solution was spiked with 90 µL of a protein standard digest mix (Waters Corporation, Milford, MA),

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

which consists of 11 nM Bovine serum albumin (BSA), 44 mM rabbit glycogen phosphorylase b (PHO), 222 mM yeast alcohol dehydrogenase (ADH) and 2 mM yeast enolase (ENL). The digestion protocol was designed to produce the same volume of peptide digest for each sample (1 mL), irrespective of the initial mAb concentration. Because the injection volume for each sample was kept the same for each 2D-LC-MS^E experiment (100 µL), the amount of the spiked proteins loaded on-column was constant for every injection: 800 fmoles ADH, 320 fmoles PHO, 80 fmoles BSA and 16 fmoles ENL.

The protein standard digest mix that was spiked in the HCP sample during sample preparation stage serves two different purposes: (1). It provides an internal quantitation standard so a label-free quantitation (Hi3 method) (*99*, *100*) can be simultaneously performed along with the identification of HCPs; (2). Since the four protein standards were spiked at various concentration levels (ranging from a single ppm to hundreds of ppm), the identification of some or all of the four proteins can provide a judicious measure to properly reflect the HCP concentration level at which the employed 2D-LC-MS^E approach is capable of detecting, thus validating the assay.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Experimental Conditions

2D-LC System Setup and Chromatographic Conditions

An ACQUITY™ UPLC® M-Class system (Waters Corporation, Milford, MA, USA) equipped with online 2D-LC technology was used to perform the peptide separations. A schematic diagram illustrating the operation of the 2D-LC system is presented in Figure 1. The first chromatographic dimension performs peptide fractionation under basic (pH 10) conditions on a reversed-phase Peptide XBridge BEH C18 300Å 5µm 1.0 x 50 mm column (Waters Corporation) at a flow rate of 10 µL/min. Mobile phase A was 20 mM ammonium formate in water (pH 10), and mobile phase B was pure ACN. A 0.3 × 25mm trap column packed with 5 µm Symmetry C18 (ACQUITY UPLC M-Class Symmetry C18 2D HCP Trap, 100Å, 5 µm, 300 µm × 25 mm; Waters Corporation, Milford, MA, USA) was used to trap peptides eluted from the first dimension. Peptide fractions were eluted in step gradients from the first dimension column and mixed on-line with 100 µL/min of 0.1% TFA v/v aqueous solution (1:10 dilution) before being trapped on the trapping column (Figure 1). The ten-fold online dilution reduces the organic content and the pH of the mobile phase so that peptides can be effectively retained on the trap column before the second dimension separation. The mobile phases for the second chromatographic dimension (low pH RP) were 0.1% FA v/v in water (mobile phase A) and 0.1% FA v/v in ACN (mobile phase B). The second dimension column was a 0.3 mm × 150 mm sub-2µm C18 column (ACQUITY UPLC M-Class HSS T3, 1.8 µm, 300 µm × 150 mm; Waters Corporation, Milford, MA, USA). The flow rate for the second dimension separation was set at 8 µL/min and the column was maintained at 60 °C. A 30-min gradient from 3 to 35% B was employed for the second dimension separation. The column was washed using 85% B for 2 min and re-equilibrated at 3% B for

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

10 min before returning to the next step of fractionation. Under such operating conditions, a peak capacity of greater than 2,500 can be routinely achieved.

The step elution gradients for the first dimension were optimized such that approximately the same amount of peptides was eluted off at each step. Throughout the study, 10-step fractionation was used, and the step gradients used for the first dimension separation consisted of elution with 10.7, 12.4, 14.0, 15.4, 16.7, 18.6, 20.4, 25.0, 30.0 and 50% B. The fractionation process started immediately after the completion of sample loading (15 min at 10 µL/min with 3% B), and each elution step was completed in 15 min (using a flow rate of 10 µL/min). The ACQUITY™ UPLC® M-Class 2D system was operated in a completely on-line manner, and the whole procedure did not involve any offline fraction collection and re-injection. During the entire process of the 10-step fractionation, the high pH mobile phase continued to flow in the first dimension for 30 minutes while prior "fraction" was eluted by the gradient in the second dimension.

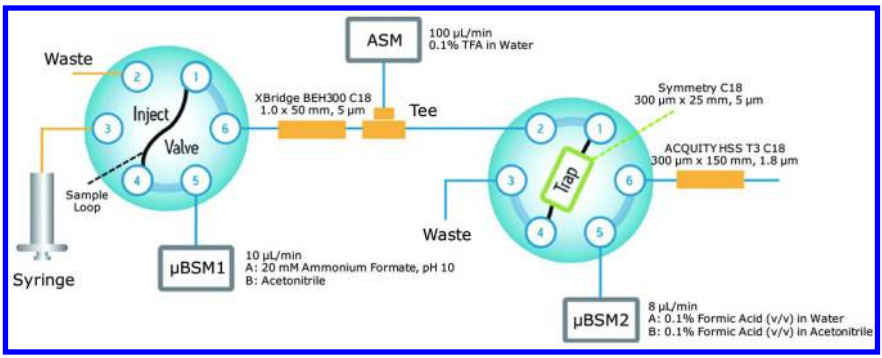

Figure 1. Fluidic configuration of the 2-dimensional chromatography with online dilution. The column dimension and the mobile phase compositions are provided on the figure. The figure illustrates the fluidic path during sample loading. For other detailed 2D-LC configurations such as peptide fractionation using the first chromatographic dimension (high pH reversed phase), peptide trapping; and peptide separation in the second dimension (low pH reversed phase), please refer to (19). µBSM1-µBinary Solvent Manager 1; µBSM2- µBinary Solvent ManagerBBBBASM-Auxiliary Solvent Manager. (Figure adapted with permission from the American Chemical Society, ref (107)). (see color insert)

MS System and MS Settings

A multiplexed data acquisition method (MS^E) was employed for the mass spectrometric analysis. The LC-MS^E data was acquired using a quadrupole time-of-flight mass spectrometer (Synapt G2-S HDMS, Waters Corporation, Milford, MA) equipped with a microscale ESI probe fitted with a small bore (50 µm ID) stainless steel capillary (Waters Corporation, p/n 186007529). For all measurements, the mass spectrometer was operated in positive ESI ion mode with a typical resolving power of 20,000 FWHM. Data were acquired in continuum

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

mode over *m/z* range of 100 — 1990, using a capillary voltage of 3.0 kV, a source temperature of 100 °C and a sample cone voltage of 30 V. The desolvation temperature was set to 250 °C and the desolvation gas flow rate was 500 L/hour.

The LC-MS^E data was collected by alternating the collision energy of the MS instrument between low energy (MS) and elevated energy (MS^E) without precursor selection. The spectral acquisition time at each energy setting was 0.5 seconds such that one spectrum of MS and MS^E data was acquired every 0.6 seconds. In the low energy MS mode, the data were collected at a constant Trap-cell collision energy of 5 eV and a constant Transfer-cell collision energy of 4 eV. While in the MS^E mode, the collision energy was ramped from 20 to 45 eV. A solution of 50 fmol/μL Glu1-fibrinopeptide B (GFP) in 50% acetonitrile with 0.1% FA was used as a lock-mass solution. The solution was delivered at a flow rate of 5 μL/min using an auxiliary pump of the ACQUITY M-class 2D-LC system. The lock-mass data were sampled every 4 min using 0.5 sec scans over a mass range of 100-1990 m/z.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Sample Amount Injected for Analysis

The ACQUITY UPLC M-Class 2D system was configured with a 100 μL volume sample loop and a 250 μL sample syringe. During the analysis of the reference standards, the injection volume for each sample was kept the same for each 2D-LC-MS^E experiment (100 μL). Therefore, the total amount of protein digest for HCP identification was 250 μg for NISTmAb.

Data Processing and Database Searching for HCP Identification

LC-MS^E data were processed using ProteinLynx Global Server (PLGS) 3.0.2 software (Waters Corporation, Milford, MA, USA) (*97*, *102*) in order to automate the identification of HCPs. For each reference sample, all of the MS^E data from each fractionation step were digitally combined into a single file using PLGS (*97*, *102*). The low-energy and high-energy (MS^E) data were background subtracted, de-isotoped and charge-state reduced to the corresponding monoisotopic peaks. Each monoisotopic peak was then lock-mass corrected to yield the accurate mass measurement. Fragment ions and their corresponding precursor ions were automatically aligned (grouped) together based on the retention time profiles of the ions (*97*, *102*). Processed spectra were searched against a mouse protein database from Swiss Prot containing 16,644 mouse protein entries. The protein sequences of four spiked-in proteins (ADH, PHO, BSA, ENL), the sequence of porcine trypsin, Protein A (*S. aureus*) sequence, and the heavy/light chain sequences of the NISTmAb were added to the database. The final custom database also included an equal number of entries of randomized (decoy) sequences (one random sequence for each true sequence), containing a total of 33,302 entries in the database. The decoy strategy was used to control the false positive rate. The search was limited to tryptic peptides with one potential missed cleavage. The mass tolerance allowed for the low—energy precursor ions was 15 ppm, while the mass tolerance of elevated-energy fragment ions was set to 20 ppm.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

A false discovery rate (FDR) of 10% was used during peptide identification with PLGS. This low stringency setting minimized false negatives particularly when the number of HCPs returns from the search is small. Protein match criteria were employed to compensate for this low fidelity matching. Specifically, a protein was reported as being "identified" only when the protein was identified in at least of 2 out 3 replicate runs. For the details on how protein assignment is accomplished, please refer to the published work (*102*).

Results

Chromatographic Performance of a 2D-LC System for Separation of Overloaded Digests from Therapeutic Proteins

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Tryptic cleavage generates multiple peptides per protein such that biotherapeutic samples typically consist of hundreds sometime thousands of peptides, if a large number of HCPs are present. These peptides span a wide concentration range. The ability of the 2D-LC system to separate such a complex mixture was investigated, and the chromatograms from a 10-step separation of NISTmAb digest are depicted in Figure 2. As can be seen, a majority of the separation space in each chromatogram is occupied by a number of very broad chromatographic peaks resulting from the overloading of peptides from the mAb digest. In addition, some narrow and well-defined chromatographic peaks can be observed scattered between those broad chromatographic peaks. Many of these small peaks are from some of the HCP peptides that are subject to MS analysis for HCP identification. For comparison, pane A in Figure 2 displays a chromatogram acquired from a one dimensional separation for the same digest using a long, shallow gradient on a 250 mm length, sub-2μm particle C18 column. The comparison of the chromatograms from both the 1D and 2D separation reveal an interesting observation - many small chromatographic peaks observed in an individual fraction from the 2D separation eluted with retention times otherwise occupied by overloaded, broad peaks in other fractions. This observation suggests that these potential HCP peptides would have co-eluted with high-abundance, therapeutic-derived peptides if a separation based on a single dimension was relied upon. It is perceivable that the additional resolution of the small peaks afforded by 2D-LC increases the chance for their identification and the confidence of their quantitation.

The implementation of 2D-LC (high-pH RP/low-pH RP) was accomplished using a micro-scale column (0.3 ×150 mm) configuration (see the detailed description in experimental setup). This configuration is adopted in HCP analysis to strike a balance between the sensitivity requirement for low-abundance protein identification and the robustness of the experimental operation. Nanoflow chromatography (e.g. 75 μm ID columns) is frequently employed in proteomics analysis. Although it is highly sensitive, nanoflow chromatography generally lacks the robustness needed for routine, industry-wide application, because nano-columns and nano-ESI emitters are prone to clogging, are easily broken, and because nanoLC fittings are notoriously challenging to connect without introducing chromatography-affecting unswept volumes. Coupling to a standard

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

electrospray probe equipped with a narrow bore stainless steel capillary (50 μm), micro-scale chromatography can deliver a highly efficient separation (Figure 2) with a modest increase (16 fold) in the sample consumption to match up the sensitivity achieved in nanoflow chromatography.

Figure 3 shows the total ion chromatogram (pane A) and an extracted ion chromatogram (pane B) for a peptide (*m/z* 594.31.) from a HCP (Uniquitin conjugating enzyme E2 variant, 21 ppm) in the NISTmAb that was identified via micro-scale 2D-LC-MS^E in fraction 9 at a retention time of ~17 minutes. This peptide was reasonably well resolved by the 2D separation, though many co-eluting species could still be observed in the MS spectrum at that retention time (Figure 3(C)). Regardless, MS^E fragmentation along with PLGS data processing was capable of confidently producing the identity of the HCP peptide, as exemplified in Figure 3(D). This example clearly illustrates why there is a need to combine high-resolution separations with non-targeted multiplexing peptide fragmentation approaches to identify low abundance HCP peptides.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

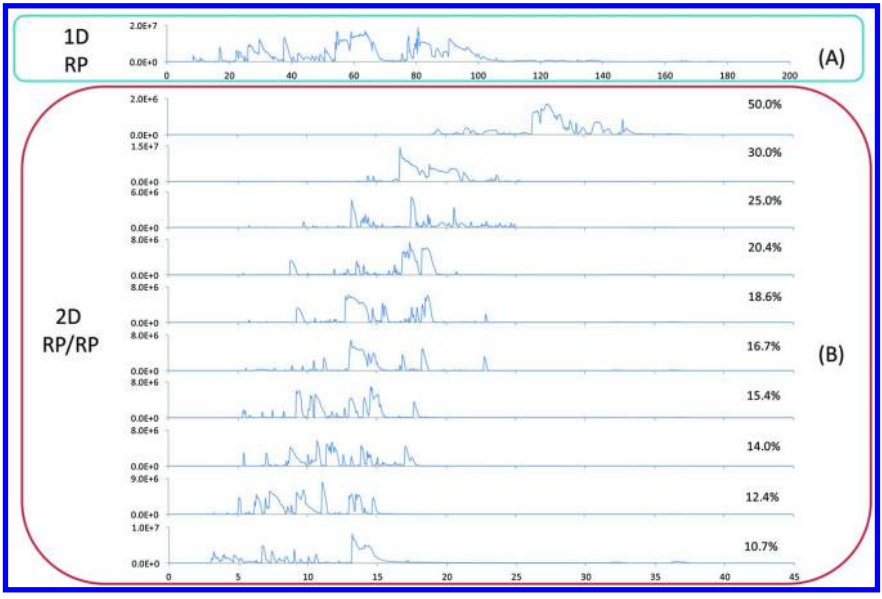

Figure 2. Total ion chromatograms of NISTmAb digest from a single-dimension LC separation (A) and a 10-step 2D-LC separation (B). The percentage values listed are the percentage of acetonitrile used in the step-gradient for the fractionation of peptide in the first dimension column. (see color insert)

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

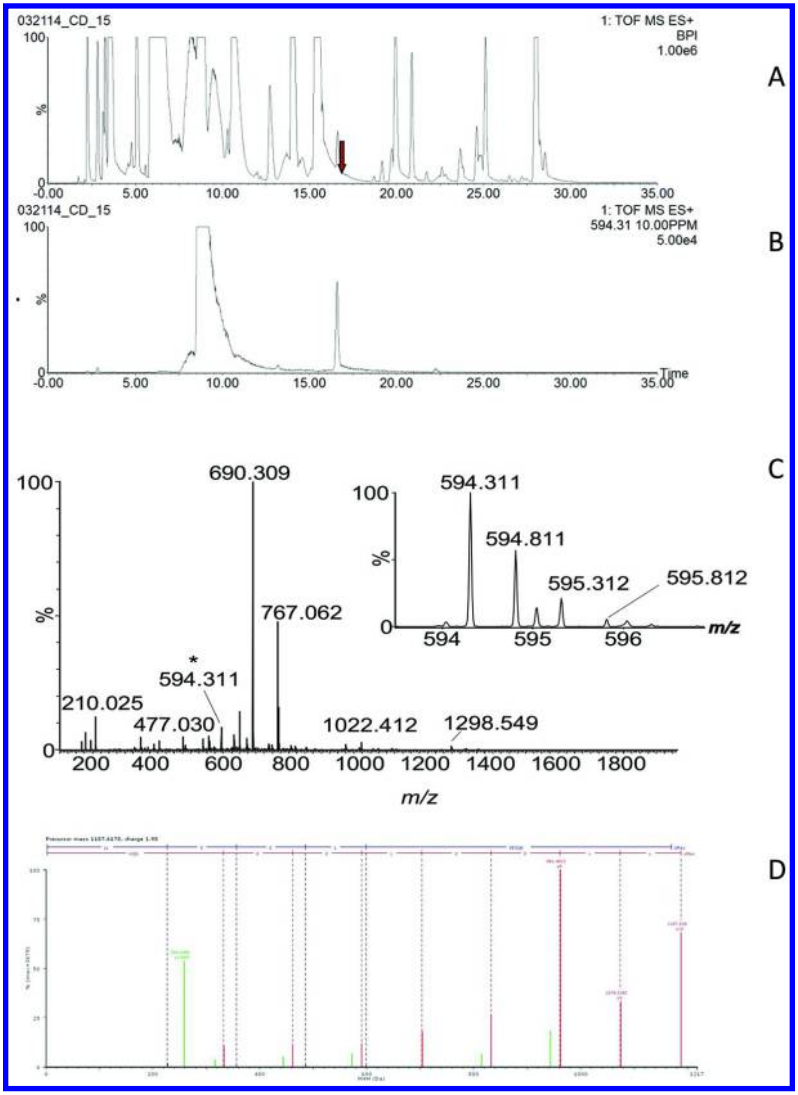

Figure 3. Identification of a low-abundance peptide (LLEELEEGQK) from a HCP (Uniquitin conjugating enzyme E2 variant, 21 ppm) in NISTmAb. (A) Total ion chromatogram (base peak chromatogram) from the 2nd dimension separation of the peptides eluted from the 2nd-step fractionation in the 1st dimension analysis; (B) Extracted mass chromatogram (± 10 ppm) for the HCP peptide (LLEELEEGQK) identified in the NISTmAb; (C) MS spectrum showing the ion signals observed in the same elution time window of the HCP peptide LLEELEEGQK (labeled by asterisk). The isotopic distribution of the peptide is shown in the insert. The spectrum is composed of 10 combined scans across the entire chromatographic peak-width; (D) High energy MS^E fragmentation spectrum of peptide LLEELEEGQK. The spectrum was taken from the data processing software PLGS. (see color insert)

Using the 10-fraction 2D-LC-MS^E experimental conditions described above, we set out to identify and quantify HCPs present in this NISTmAb. Results from this work are summarized in Table I. Each of the HCP in Table I was identified in at least 2 of 3 replicates. In addition, three of the four spiked proteins (PHO, ADH and BSA) were identified in all mAb digests. ENL, which has the lowest concentration among all the spiked proteins, was not identified in the sample.

Table I. Host Cell Proteins (HCPs) Identified in NISTmAb (Bold) and Their Corresponding Concentrations (ppm, Expressed in ng HCP/mg total protein). The protein concentrations were calculated using top three best responding peptides in ESI-MS (*97*, *98*) from an average of three replicate injections. Also included in the Table are the spiked protein digest standards and the calculated respective molar amount.

Protein No.	Accession Number	Protein Description	Sequence Coverage (%)	Average MW (kDa)	Amount on Column fmoles	Amount on Column ng	Concentration ng/mL	Concentration ppm	RSD (%)
1	P00330	Alcohol Dehydrogenase Yeast (ADH) - 2000 fmoles	59.9	36.7	1512	55	22196	222	2.3
2	**P05064**	**Fructose Biphosphate Aldolase A Isoform**	**74.4**	**39.3**	**914**	**36**	**14368**	**144**	**18.3**
3	P00489	Glycogen phosphorylase rabbit (PHO) - 400 fmoles	37.2	97.1	400	39	15536	155	0.0
4	**P05063**	**Fructose Biphosphate Aldolase C Isoform**	**55.6**	**39.3**	**584**	**23**	**9180**	**92**	**29.5**
5	**Q922R8**	**Protein Disulfide Isomerase A6**	**34.1**	**48.1**	**260**	**13**	**5002**	**50**	**26.2**
6	P02769	Bovine Serum Albumin (BSA) - 100 fmoles	28.0	66.3	173	11	4588	46	11.1
7	**P08101**	**Low Affinity Immunoglobulin Gamma Fc Region Receptor**	**22.1**	**36.7**	**304**	**11**	**4463**	**45**	**21.0**
8	**Q9D2M8**	**Uniquitin Conjugating Enzyme E2 Variant**	**28.3**	**16.3**	**316**	**5**	**2060**	**21**	**33.2**
9	P00924	Yeast Enolase 1 (ENL) - 20 fmoles	-	46.6	20	1	373	4	-
Total HCP Concentration (ppm)								351	
Total Sample Purity (%)								99.965%	

As discussed above, the concentration of an identified protein in a complex mixture can be estimated using the Hi3 method (*99*, *100*). The inclusion of the spiked proteins in NISTmAb allows us to use the Hi3 method to quantify the concentration of each HCP discovered. For this purpose, PHO was selected as an internal reference, of which a known amount of 400 fmoles was loaded on-column in every 2D-LC-MS^E experiment. From the ratio between the averaged intensity of the top-3 best responding peptides from each identified HCP and the internal reference PHO, the molar amounts loaded on column for all identified proteins (including the other spiked ones) can be calculated. Based on the average molecular weight of each protein, the on-column HCP amount (in nanograms) can then be readily calculated. This value can then be used to back-calculate the protein concentration in the original sample, after taking into account the volume changes during the sample preparation. Finally the HCP concentration in the NISTmAb is expressed in ppm (as ng of HCP/mg total protein) based on the total protein concentrations (provided by the sample supplier). The ppm values for the spiked internal standards are also included in the table for reference.

Table I shows that five HCPs (names in bold) were identified and quantified in the current 2D-LC-MS^E analysis. Among them, two isoforms of fructose biophosphate aldolase comprise 67 % of the total HCP mass detected in the NISTmAb, with the A isoform at a concentration of 144 ppm (as ng of HCP/mg total protein) and the C isoform at 92 ppm. Additionally, three other HCPs, each

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

having a relatively low concentration, were also identified: protein disulfide isomerase A6 (50 ppm), low affinity immunoglobulin gamma Fc region receptor (45 ppm), and ubiquitin conjugation enzyme E2 variant (21 ppm).

The NISTmAb is a product that has undergone industry standard upstream and downstream purification to remove process related impurities. Therefore, its HCP content is expected to be low. Indeed, the total HCP concentration measured by a generic HCP ELISA assay indicates it contains <10 ppm of HCPs (1.8 ng/mg; Process Impurities chapter/Volume 2, Chapter 9). The discrepancy between the ELISA result and 2D-LC-MS^E analysis may be ascribed to the following reasons (1). The ELISA method may underestimate the HCP concentration because of the low reactivity of the multi-component ELISA kits. The employed ELISA kit in the current study was a platform kit from a single manufacturer, normally developed for a number of commonly used biopharmaceutical production cell lines. The reactive antibodies in the ELISA kits may not adequately react with all HCPs potentially present in NISTmAb since every biopharmaceutical product follows its unique process. The discrepancy between the ELISA results and the MS-based methods was also reported previously (*18*, *19*), and indicates that MS-based methods are able to more comprehensively profile the HCP proteome . (2). the 2D-LC-MS^E-based method covers somewhat different HCP population from ELISA. Since a MS method makes a measurement based on the responses of identified peptides from HCPs, multiple forms of HCP species (including protein clips, isoforms, etc.) in the samples, regardless their reactivities towards antibodies in the ELISA kits, would be detectable by MS. This suggests that MS-based method can potentially combine different protein isoforms together and overestimate the concentration of a particular HCP isoform. (3). The Hi3 method used to calculate the HCP concentration may have resulted in a very rough estimate about the protein concentration under analysis. In a recent study, the correlation between the top 3 intensity and protein load for trace proteins in up to 100,000-fold excess drug product was thoroughly investigated. Because every HCP exhibits different MS response factors, it is estimated Hi3 method could potentially quantify unknown HCPs within an expected error of 1.5-fold (high) and 1.8-fold (low) of their actual value (*18*).

The difference observed between the ELISA results and the LC-MS^E assay in the current exercise suggests that analytical results from orthogonal analytical methods may not agree well with each other all the time. This observation brings in an interesting question on how to identify the source of error in the HCP assays and/or bridge the analytical results obtained from the two assays in a development environment. One of the orthogonal methods that can be applied to test those results is an antigen specific ELISA assay. The identification of HCP by MS methods provides the names of HCPs, from which an antigen specific ELISA assay can be developed to target specific HCPs. Schenauer et al (*17*, *18*) recently used antigen specific ELISA assay to validate the LC-MS^E quantitation results in a HCP analysis, and found the results from the two assays agree very well. Secondly, the quantitation results from LC-MS^E HCP assay may also be examined or validated by an alternative LC-MRM (multiple reaction monitoring) method with synthetic AQUA peptides (*103*) as internal standards. Methods employing AQUA peptides are commonly practiced in analytical laboratories to

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

obtain the absolute quantitation measurement for the proteins of interests from various biological matrix (e.g. plasma, urine, etc.). In a separate study, Doneanu et al (*19*) compared the absolute quantification of three selected HCPs, based on LC-MRM methods and spiking AQUA peptides, against the results from the Hi3 quantitation method, and found that the Hi3 and MRM data sets correlated reasonably well.

The experimental results for NISTmAbs show that a 10-fraction 2D-LC-MS^E analysis can confidently identify HCPs down to a level of 20 ppm for a sample containing a highly concentrated mAb. This is a noteworthy performance, particularly in light of the limited capabilities of an alternative 1D separation, such as the one previously discussed and portrayed in Figure 2A. For comparison, Pane A in Figure 2 displays a chromatogram acquired from a one-dimensional separation for the same digest using a long, shallow gradient on a relatively long sub-2μm particle C18 column (0.3 × 250 mm). Using that 1D-LC-MS^E platform, not a single HCP (or spiked-in protein) could be detected in the NISTmAb. Yet with a 10-step 2D-LC analysis, 5 different HCPs and 3 of the 4 spiked-in proteins could be detected. The limit of detection for a long gradient, 1D-LC based technique might therefore be over an order of magnitude higher than that of an appropriately designed 2D-LC based technique. It is worth pointing out that limits of detection of current methodologies is influenced by many experimental variables, such as the number of fractionation steps used in a 2D separation and the molecular masses of therapeutic drugs. Therapeutic drugs with higher MW typically generate more peptides (and their derivatives) during the sample preparation (digestion) step. Since these peptides are the major interfering source for the HCP identification, the size (masses) of the therapeutic drugs directly impact the level of HCP that can be identified for a given method. For this reason, in-depth discussion on the performance of the 2D-LC-MS^E approach for HCP coverage will be reserved to several previous publications and not included in any greater detail in this chapter (*18*, *19*, *30*).

The results shown in this section demonstrates the performance of the 2D-LC-MS^E approach for HCP analysis of a highly concentrated reference mAb standard. Detailed protocols are included here so the analysis results can be corroborated by interested readers. In the following sections, two case studies from published literature on HCP profiling in true biopharmaceutical settings are provided to further illustrate the significance of information acquired by the 2D-LC-MS^E approach.

Case II: Tracking Removal of Individual HCP during Purification

Tracking the clearance of residual HCPs is an essential part of process development and characterization, and regulatory guidelines demand that suitable analytical methods be established to demonstrate the minimization of HCPs across purification stages toward the final drug product. In comparison with the measurement of total HCP content by ELISA assay, quantitative information on HCP concentration and distribution for in-process samples should help elucidate

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

mechanisms behind how individual HCP are co-purified with mAb products, or are selectively eliminated, at certain steps of a purification process. Furthermore, quantitative tracking of individual HCP through purification is of significance to rational process development because it allows clearance factors at different purification steps to be assessed and correlated with physiochemical properties of individual HCP. This, in turn, should allow well-informed, as opposed to empirical, purification improvements that target HCP deemed problematic. This rational approach will facilitate bioprocess development, improve our understanding of purification effects, and increase the knowledge base.

Two separate reports on using 2D-LC-MS^E to acquire the identity and concentration of individual HCP in biotherapeutic products during purification have been published. Zhang et al (*30*) recently reported a study on comprehensive tracking of HCPs during monoclonal antibody purifications using a 2D-LC-MS^E method. Individual HCP present at various purification steps were identified and quantified in several therapeutic mAb samples produced from Chinese hamster ovary (CHO) cells. In this study, the authors tracked individual HCP from cell culture fluid to a Protein A eluate pool to a subsequent viral inactivation pool and, in some cases, even further downstream. Approximately 500 HCPs were confidently identified in cell culture fluid. This HCP population was confirmed to decline progressively through the purification scheme until no HCP could be confidently identified in samples that had been subjected to a polishing step (cation-exchange chromatography). In addition, it was shown that the 2D-LC-MS^E methodology could be used for reproducible identification and quantification among replicate analyses. Furthermore, Zhang and co-workers suggested that HCPs could be confidently identified and quantified to ~13 ppm, which agrees well with a value previously documented in a different analysis setting for the individual limit of quantification of HCPs (*18*).

Literature suggests that the majority of HCPs remaining in the in-process pools after Protein A purification step are attributed to a weak interaction between the HCPs and the therapeutic mAb rather than an HCP-resin association (*82*). In an effort to understand what factors may contribute to the HCP population after Protein A purification, the authors analyzed Protein A purified pools of nine mAbs, all using approximately similar platform cell culture conditions and purified by similar protocols. The results showed that the Protein A eluate pool of nine different mAbs contained widely differing concentration distributions of HCPs. The number of HCPs identified ranged from 8 to 74, with total HCP levels ranging from 337 to 7450 ppm. However, the bulk of the total HCP content in each case came from a small subset of normally intracellular HCPs that are highly abundant in cell culture conditioned media. The fact that different mAbs produced a profile that largely consisted of a group of highly abundant upstream HCPs suggests that mAbs interact weakly, through non-specific binding, with different members of these cellular proteins to varying degrees. This interaction is probably driven in a concentration dependent manner between highly abundant HCPs and mAbs, which are bound at high effective concentration to the affinity resin. These observations imply that downstream HCP content is dependent on the extent of cell lysis that occurs during cell culture/harvest.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

In a separate publication, the 2D-LC-MS^E analytical methodology was applied to the identification and quantification of HCPs present in preparations of a recombinant human protein expressed in eukaryotic cell and produced using an alternative purification process (*101*). The aim of the study was to corroborate the HCP ELISA assay results that were obtained for bioprocessing, and provide insight into the clearance of HCPs during purification. At each purification step (4 steps in total), the samples were collected, spiked with four protein standards (derived from a different organism than the host cell), and digested with a protease. The digests were analyzed by the 2D-LC-MS^E approach to identify and quantify HCPs from the samples collected. Seventy eight HCPs were identified for cell culture fluid samples with a total HCP concentration of ~ 21,000 ppm, but only 4 of these proteins were detected in the final purified, finished product (a total HCP concentration of 20 ppm). The 2D-LC-MS^E assay results were compared against ELISA measurements for the same samples, and the results correlated well for all the samples collected except the sample from the second purification step, for which a large discrepancy in the total HCP concentration between the measurement of the two techniques was observed (5,600 ppm (2D-LC-MS^E) *vs* 1,100 ppm (ELSIA)). The difference is believed to be caused by HCPs that do not signal strongly in the ELISA assay, a well-known weakness of ELISA assay. The results show that the LC-MS assay has advantages to offer not only the total concentration of HCPs in each sample but also the identity and quantitation of each individual HCP. This information helps to understand the relative contribution of purification schemes to the nature and concentrations of HCP impurities in biopharmaceutical samples, and provide insights for streamlining the purification process.

Case III: Understanding the Effects of Process Changes on the Profile of Residual HCPs in Drug Substance

To ensure the safety and efficacy of biopharmaceutical products administered to patients, manufacturers are expected to measure and control impurities in their drug substance (DS), including the host cell components and process additives that may leach into the process stream and final product. The removal of HCPs is a major consideration during bioprocess development, and the clearance is frequently reported as one of the benchmarks to demonstrate the establishment of a robust and well-controlled bioprocess. However, since HCP levels in the final drug substance depend on many intricate factors such as cell lines and purification protocols from both up- and down-stream processes, little is currently known about the identity of HCPs present in drug substance or how these HCPs survive rigorous purification schemes. The traditional ELISA assay, because of its inability to identify and quantify individual HCP, does not provide an adequate information to answer these questions and provide sound analytical support to aid the design of a rational bioprocess development. As a result, process development under those circumstances has to undertake a trial-and-error approach.

The following study illustrates how the detailed information acquired by the 2D-LC-MS^E approach can help probe the impact of process change, major or

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

minor, on HCP profiles of several peptibody (Pbs) therapeutics (*17*). The authors first analyzed the HCP content of one purified peptibody (Pb1) as a function of process changes during development. Drug substances from four distinct purification processes during the development were analyzed by the 2D-LC-MS^E method and ELISA assay to determine the levels of DS HCP. While ELISA assay results showed very little difference in total DS HCP levels between these processes (the level of HCPs ranges from 1 to 5 ppm), 2D-LC-MS^E approach identified significantly different residual HCP impurity profiles between the same materials. Two HCPs were identified in Process 1 DS; 11 HCPs in Process 2 DS, 3 HCPs in Process 3 DS, and 6 HCPs in Process 4 DS. In contrast to the similar HCP ELISA results, the total HCP concentration measured by 2D-LC-MS^E illustrates a quite different picture. The process change from Process 1 to Process 2 results in a significant increase in total detectable HCPs from 26 to 391 ppm. The concentration of DS HCPs in Process 3 and Process 4 was 195 ppm and 166 ppm, respectively.

The advantages of 2D-LC-MS^E approach for HCP analysis are further displayed by the correlations between the measured DS HCP levels and the changes among the four bioprocesses (*17*). For example, there was one common HCP found in both Process 1 and Process 2, DnaK. Single protein DnaK ELISA assay with anti-DnaK antibody was also performed to measure the concentration of DnaK in DS from Process 1 and Process 2. Average values of 8 and 73 ppm DnaK were obtained for the lots of Process 1 and 2 DS tested, respectively. These values agree well with the results from 2D-LC-MS^E measurement, which showed the DnaK level increased from ~ 20 to ~ 96 ppm between the lots of Process 1 and 2 DS. However, with HCP ELISA, the increased amount of DnaK in DS from Process 2 was not reflected at all as HCP ELISA assay indicates no changes in the total HCP levels. More interestingly, for Process 3 which was the result of modifications made to Process 2 with the goal of removing the particular HCP protein DnaK, the HCP ELISA was unable to detect a significant HCP level difference in Process 3 compared to Process 2 while this is readily tracked by either the 2D-LC-MS^E or DnaK ELISA. The failure to detect the HCP level changes in the process development suggests that ELISA is fundamentally limited to support the design of an efficient process that can explicitly exploit the physicochemical properties of the HCP.

The study demonstrates that the composition of residual HCPs as determined by MS^E in a biotherapeutic DS may sensitively reflect significant differences in both composition and quantity of individual HCP as a result of process changes. Compared to ELISA, the breadth of information obtained using MS^E can provide a more comprehensive, and accurate, characterization of DS HCPs, thereby facilitating process development as well as more rationally assessing potential safety risks posed by individual, identified HCPs.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Summary and Outlook

HCP is a critical quality attribute of biopharmaceutical products. Despite the recognized importance of controlling HCP levels, our understanding of the impact of problematic HCPs on drug efficacy and safety are rather limited. The dilemma is largely attributed to the insufficient myopic information provided by the HCP ELISA assay commonly used today. While the HCP-ELISA method is sensitive to a subset of the HCPs present in a biotherapeutic, the method is limited due to its inability to provide fundamental information on the distribution and identities of the entire HCP proteome. When an individual anti-HCP antibody is available, the Western blotting method is capable of identifying individual HCP but its sensitivity is limited. In addition, the number of commercially available anti-HCP antibodies is limited. 2D SDS-PAGE provides a general picture of the HCP proteome in a sample, but it requires additional analytical techniques to identify the separated HCPs. While all of these HCP analysis methods have been practiced widely in different stages of bioprocess development, these techniques all face significant challenges in terms of illustrating a comprehensive HCP proteome for biotherapeutic products. As the industry moves towards the Quality by Design initiative, a better understanding of the identities, concentrations and composition of HCPs seems warranted.

Mass Spectrometry-based methods can provide a comprehensive assessment of the HCPs present in biotherapeutic products. The information acquired from such an approach is not only specific (down to individual protein name), but also sensitive, allowing for the quantitation of individual proteins at low ppm levels. While an HCP ELISA method in combination with protein purity analyses by chromatographic and electrophoretic separations are most commonly employed for the routine analysis of biotherapeutic preparations, it is compelling to envision the use of a MS method as a means to orthogonally confirm product purity in the execution of an in-depth assessment of biotherapeutic purity.

Future development of the technique should aim to increase the sample throughput of the assay and further improve the detection limit of HCPs. For a single ten-step of 2D separation, up to 30 hours of instrument time may be required to process one biotherapeutic sample. In a modern bioprocess development project, analysis of multiple samples is necessary to find an optimized purification scheme. This throughput (30-hr per sample) will face significant challenge to meet development need. In addition, it would be highly desirable to improve the limit of detection of LC-MS assay so HCPs at sub-ppm level can be confidently identified. Most of the current MS-based methods can only identify HCPs at the single ppm level at their best performance (*17–19*, *27*, *30*, *92*), and the sensitivity is lower in comparison to ELISA assay. For any HCP assay, confidence in identification and quantification, reproducibility, accuracy, sensitivity and throughput are all important factors that need to be taken into account.

The ultimate goal of HCP analysis is to identify and quantify every protein present in the host-cell proteome that has copurified with the target protein therapeutic. Understanding the presence or absence of every host cell protein in a biotherapeutic as well as its relation to manufacturing changes is key to better understanding bioprocess development. The use of multiple HCP assays is

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

still the best strategy to ensure that a complete picture of the impurities present throughout the manufacturing process is known, and thus, a robust HCP removal strategy can be implemented. While the current technology limits the use of 2D-LC-MS^E methods in a routine QC setting, the information gained can be valuable to the Quality by Design product development initiative, and ultimately for addressing HCP levels as an integral part of both innovator and biosimilar recombinant therapeutic protein regulatory-driven comparability assessments (*104–106*).

References

1. Bracewell, D. G.; Smales, M. *Bioanalysis* **2013**, *5*, 123–126.
2. Wolter, T.; Richter, A. *BioProcess Int.* **2005**, *2*, 40–46.
3. Champion, K.; Madden, H.; Dougherty, J.; Shacter, E. *BioProcess Int.* **2005**, *3*, 52–58.
4. Wang, X.; Hunter, A. K.; Mozier, N.M. *Biotechnol. Bioeng.* **2009**, *103*, 446–458.
5. Wadhwa, M.; Skog, A. L.; Bird, C.; Ragnhammar, P.; Lilljefors, M.; Gaines-das, R.; Mellstedt, H.; Thorpe, R. *Clin. Cancer Res.* **1999**, *5*, 1353–1361.
6. Sharma, B. *Biotechnol. Adv.* **2007**, *25*, 325–331.
7. Gao, S. X.; Zhang, Y.; Stansberry-Perkins, K.; Buko, A.; Bai, S.; Nguyen, V.; Brader, M. L. *Biotechnol. Bioeng.* **2011**, *108*, 977–982.
8. *CPMP Position Statement on DNA and Host Cell Proteins (HCP) Impurities, Routine Testing versus Validation Studies*; European Medicines Agency: London, 1997.
9. Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use. U.S. Food and Drug Administration, 1997. http://www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/OtherRecommendationsforManufacturers/UCM153182.pdf (accessed June 2015).
10. *Guidance for Industry. Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products*; U.S. Food and Drug Administration: Washington, DC, 1999.
11. Hart, R. A.; Rinas, U.; Bailey, J. E. *J. Biol. Chem.* **1990**, *265*, 12728–12733.
12. Rinas, U.; Bailey, J. E. *Appl. Microbiol. Biotechnol.* **1992**, *37*, 609–614.
13. Shukla, A. A.; Hinckley, P. *Biotechnol. Prog.* **2008**, *24*, 1115–1121.
14. Rinas, U.; Boone, T. C.; Bailey, J. E. *J. Biotechnol.* **1993**, *28*, 313–320.
15. Robey, F.A. In *Methods in Protein Sequence Analysis*; Walsh, K. A., Ed.; Humana Press: Clifton, NJ, 1986; pp 67−78.
16. Simmerman, H.; Donnelly, R. P. *BioProcess Int.* **2005**, *3*, 32–40.
17. Schenauer, M. R.; Flynn, G. C.; Goetze, A. M. *Biotechnol. Prog.* **2013**, *29*, 951–957.
18. Schenauer, M. R.; Flynn, G. C.; Goetze, A. M. *Anal. Biochem.* **2012**, *428*, 150–157.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

19. Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Stapels, M.; Chen, W. *mAbs* **2012**, *4*, 24–44.
20. Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A. *Biotechnol. J.* **2013**, *8*, 655–670.
21. Jin, M.; Szapiel, N.; Zhang, J.; Hickey, J.; Ghose, S. *Biotechnol. Bioeng.* **2010**, *105*, 306–316.
22. Krawitz, D. C.; Forrest, W.; Moreno, G. T.; Kittleson, J.; Champion, K. M. *Proteomics* **2006**, *6*, 94–110.
23. Dagouassat, N.; Haeuw, J.; Robillard, V.; Damien, F.; Libon, C.; Corvania, N.; Lawny, F.; Nguyen, T.; Bonnefoy, J.; Beck, A. *J. Immunol. Methods* **2001**, *251*, 151–159.
24. Valente, K. N.; Schaefer, A. K.; Kempton, H. R.; Lenhoff, A. M.; Lee, K. H. *Biotechnol J.* **2014**, *9*, 87–99.
25. Capito, F.; Skudas, R.; Kolmar, H.; Stanislawski, B. *Biotechnol. Bioeng.* **2013**, *110*, 252–259.
26. Zhu, G.; Sun, L.; Linkous, T.; Kernaghan, D.; McGivney, J. B.; Dovichi, N. J. *Electrophoresis* **2014**, *35*, 1448–1452.
27. Xiao, G.; Ren, D.; Bondarenko, P. V. *Paper presented at the 62st Annual ASMS Conference on Mass Spectrometry and Allied Topics*, June 15−19, 2014, Baltimore, MD.
28. Reisinger, V.; Toll, H.; Mayer, R. E.; Visser, J.; Wolschin, F. *Anal. Biochem.* **2014**, *463*, 1–6, DOI: 10.1016/j.ab.2014.06.005.
29. Gagnon P. In *Process Scale Bioseparations for the Biopharmaceutical Industry*; Shukla, A., Gadam, S., Etzel, M., Eds.; CRC Press Taylor and Francis: Boca Raton, FL, 2007; pp 491—505.
30. Zhang, Q.; Goetze, A. M.; Cui, H.; Wylie, J.; Trimble, S.; Hewig, A.; Flynn, G. C. *mAbs* **2014**, *6*, 659–670.
31. Yigzaw, Y.; Piper, R.; Tran, M.; Shukla, A. A. *Biotechnol. Prog.* **2006**, *22*, 288–296.
32. Hoffman, K. *BioPharm. Int.* **2000**, *13*, 38–45.
33. Zhu-Shimoni, J.; Yu, C.; Nishihara, J.; Wong, R. M.; Gunawan, F.; Lin, M.; Krawitz, D.; Liu, P.; Sandoval, W.; Vanderlaan, M. *Biotechnol. Bioeng.* **2014**, DOI: 10.1002/bit.25327.
34. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianférani, S. *Anal. Chem.* **2013**, *85*, 715–736.
35. Wang, X.; Schomogy, T.; Wells, K.; Mozier, N. M. *BioProcess Int.* **2010**, *8*, 18–24.
36. Flatman, S.; Alam, I.; Gerard, J.; Mussa, N. *J. Chromatogr. B* **2007**, *848*, 79–87.
37. Savino, E.; Hu, B.; Sellers, J.; Sobjak, A.; Majewski, N.; Fenton, S.; Yang, T. *BioProcess Int.* **2011**, *9*, 68–75.
38. Schwertner, D.; Kirchner, M. *BioProcess Int.* **2010**, *8*, 56–62.
39. Hunter, A. K.; Wang, X.; Suda, E. J.; Herberg, J. T.; Shell, R. E.; Thomas, K. E.; Dufield, R. L.; Gustafson, M. E.; Mozier, N. M.; Ho, S. V. *Biotechnol. Prog.* **2009**, *25*, 446–453.
40. Bomans, K.; Lang, A.; Roedl, V.; Adolf, L.; Kyriosoglou, K.; Diepold, K.; Eberl, G.; Mølhøj, M.; Strauss, U.; Schmalz, C.; Vogel, R.; Reusch, D.;

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Wegele, H.; Wiedmann, M.; Bulau, P. *PLoS ONE* **2013**, *8*, e81639, DOI: 10.1371/journal.pone.0081639.
41. Garfin, D. In *Analytical Techniques for Biopharmaceutical Development*; Rodrigueez-Diaz, R.,Wehr, T., Tuck, S., Eds.; Marcel Dekker: New York, 2005; pp 113−160.
42. Gorg, A.; Postel, W.; Gunther, S. *Electrophoresis* **1998**, *9*, 531–546.
43. O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007–4021.
44. Unlu, M.; Morgan, M. E.; Minden, J. S. *Electrophoresis* **1997**, *18*, 2071–2077.
45. Alban, A.; David, S. O.; Bjorkesten, L.; Anderson, C.; Sloge, E.; Lewis, S.; Currie, I. *Proteomics* **2003**, *3*, 36–44.
46. Wan, M.; Wang, Y.; Rabideau, S.; Moreadith, R. *J. Pharm. Biomed. Anal.* **2002**, *28*, 953–963.
47. Wang, X; Schomogy, T.; Wells, K.; Mozier, N. M. *BioProcess Int.* **2010**, *8*, 18–24.
48. Jin, M.; Szapiel, N.; Zhang, J.; Hickey, J.; Ghose, S. *Biotechnol. Bioeng.* **2009**, *105*, 306–316.
49. Bartlow, P.; Uechi, G. T.; Cardamone, Jr., J. J.; Sultana, T. *Protein Expression Purif.* **2011**, *78*, 216–224.
50. Grzeskowiak, J. K.; Tscheliessnig, A.; Toh, P. C.; Chusainow, J. *Protein Expression Purif.* **2009**, *66*, 58–65.
51. Grzeskowiak, J. K.; Tscheliessnig, A.; Toh, P. C.; Chusainow, J.; Lee, Y. Y.; Wong, N.; Jungbauer, A. *J. Chromatogr. A* **2009**, *1216*, 4902–4912.
52. Krawitz, D. C.; Forrest, W.; Moreno, G. T.; Kittleson, J.; Champion, K. M. *Proteomics* **2006**, *6*, 94–110.
53. Grzeskowiak, J. K.; Tscheliessnig, A.; Wu, M. W.; Toh, P. C.; Chusainow, J.; Lee, Y. Y.; Wong, N.; Jungbauer, A. *Electrophoresis* **2010**, *31*, 1862–1872.
54. Berrill, A.; Ho, S. V.; Bracewell, D. G. *Biotechnol. Prog.* **2010**, *26*, 881–887.
55. Berkelman, T.; Sun, C.; Harbers, A.; Bandhakavi, S. *BioProcess Int.* **2013**, *11*, 50–61.
56. Rabilloud, T. *Methods Mol. Biol.* **2012**, *893*, 61–73.
57. Marouga, R.; David, S.; Hawkins, E. *Anal. Bioanal. Chem.* **2005**, *382*, 669–678.
58. Speicher, D. W Electrophoretic Separation of Proteins. In *Current Protocols in Protein Science*; John Wiley and Sons: Hoboken, NJ, 2008; Section II, Chapter 10, pp 10.2.1−10.5.1.
59. Krawitz, D. C.; Forrest, W.; Moreno, G. T.; Kittleson, J.; Champion, K. M. *Proteomics* **2006**, *6*, 94–110.
60. Issaq, H. J.; Veenstra, T. D. *BioTechniques* **2008**, *44*, 697–700.
61. Zhu, D.; Saul, A. J.; Miles, A. P. *J. Immunol. Methods* **2005**, *306*, 40–50.
62. Wan, M.; Wang, Y.; Rabideau, S.; Moreadith, R.; Schrimshera, J.; Conna, G. *J. Pharm. Biomed. Anal.* **2002**, *28*, 953–963.
63. Flatman, S.; Alam, I.; Gerard, J.; Mussa, N. *J. Chromatogr. B* **2007**, *848*, 79–87.
64. Ludwig, R.; Bongers, J.; Tao, L.; Huang, Y.; Fu, J.; Wu, W.; Liu, P.; Song, H.; Russell, R. In *Characterization of Protein Therapeutics using mass Spectrometry*; Chen, G. Ed.; Springer: New York, 2013; pp 207−277.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

65. Dowling, P.; Meleady, P.; Henry, M.; Clynes, P. *Bioanalysis* **2010**, *2*, 1609–1615.
66. Chen, G.; Pramanik, B. N. *Drug Discovery Today* **2009**, *14*, 465–471.
67. Thompson, H. J.; Chung, W. K.; Zhu, M.; Tie, L.; Lu, Y.; Aboulaich, N.; Strouse, R.; Mo, W. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 855–860.
68. Giddings, J.C. *Anal. Chem.* **1967**, *39*, 1027–1028.
69. Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E.A.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 3090–3100.
70. Gilar, M.; Daly, A. E.; Kele, M.; Neue, U. D.; Gebler, J. C. *J. Chromatogr. A* **2005**, *1061*, 183–192.
71. Swartz, M. E. *J. Liq. Chromatogr. Related Technol.* **2005**, *28*, 1253–1263.
72. Liu, H.; Finch, J. W.; Lavallee, M. J.; Collamati, R. A.; Benevides, C. C.; Gebler, J. C. *J. Chromatogr. A* **2007**, *1147*, 30–36.
73. Hsieh, E. J.; Bereman, M. S.; Durand, S.; Valaskovic, G. A.; MacCoss, M. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 148–153.
74. Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T.P.; Anderson, G. A.; Tang, K.; Pasa-Tolic, L.; Veenstra, T. D.; Lipton, M. S.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 1766–1775.
75. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Anal Chem.* **2005**, *77*, 6426–6434.
76. Washburn, M. P.; Wolters, D. A.; Yates, J. R., III *Nat Biotechnol.* **2001**, *19*, 242–247.
77. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *J. Sep. Sci.* **2005**, *28*, 1694–1703.
78. Yang, F.; Shen, Y.; Camp, D. G., II; Smith, R. D. *Expert Rev. Proteomics* **2012**, *9*, 129–134.
79. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G. *J. Proteome Res.* **2007**, *6*, 4363–4373.
80. Nakamura, T.; Kuromitsu, J.; Oda, Y. *J. Proteome Res.* **2008**, *7*, 1007–1011.
81. Wyndham, K. D.; O'Gara, J. E.; Walter, T. H.; Glose, K. H.; Lawrence, N. L.; Alden, B. A.; Izzo, G. S.; Hudalla, C. J.; Iraneta, P. C. *Anal. Chem.* **2003**, *75*, 6781–6788.
82. Nogal, B.; Chhiba, K.; Emery, J. C. *Biotechnol. Prog.* **2012**, *28*, 454–458.
83. Luo, Q.; Shen, Y.; Hixson, K. K.; Zhao, R.; Yang, F.; Moore, R. J.; Mottaz, H. M.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 5028–5035.
84. Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R., III *Anal. Chem.* **2006**, *78*, 5109–5118.
85. Shen, Y.; Smith, R. D.; Unger, K. K.; Kumar, D.; Lubda, D. *Anal. Chem.* **2005**, *77*, 6692–6701.
86. Shen, Y.; Jacobs, J. M.; Camp, D. G., II; Fang, R.; Moore, R. J.; Smith, R. D.; Xiao, W.; Davis, R. W.; Tompkins, R. G. *Anal. Chem.* **2004**, *76*, 1134–1144.
87. Manadas, B.; Mendes, V. M.; English, J.; Dunn, M. J. *Expert Rev. Proteomics* **2010**, *7*, 655–663.
88. Stapels, M.; Doneanu, C. E.; Fadgen, K. *Paper presented at the 61st Annual ASMS Conference on Mass Spectrometry and Allied Topics*, June 9−13, 2013, Minneapolis, MN.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

89. Davis, J. M.; Giddings, J. C. *Anal. Chem.* **1985**, *57*, 2168–2177.
90. Giddings, J. C. *J. Chromatogr. A* **1995**, *703*, 3–15.
91. Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. *J Proteome Res.* **2004**, *3*, 1082–1085.
92. Bell, A.; Rogers, R.; Kowski, T.; Robert; B. *Paper presented at the 61st Annual ASMS Conference on Mass Spectrometry and Allied Topics*, June 9−13, 2013, Minneapolis, MN.
93. Rudomin, E. L.; Carr, S. A.; Jaffe, J. D. *J Proteome Res.* **2009**, *8*, 3154–3160.
94. Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 792–803.
95. Purvine, S.; Eppel, J.; Yi, E.; Goodlett, D. *Proteomics* **2003**, *3*, 847–850.
96. Silva, J. C.; Denny, R.; Dorschel, A.; Gorenstein, M.; Kass, I.; Li, G. *J. Anal. Chem.* **2005**, *77*, 2187–2200.
97. Geromanos, S. J.; Vissers, J.; Silva, J. C.; Dorschel, C. A.; Li, G. Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I. *Proteomics* **2009**, *9*, 1683–1695.
98. Blackburn, K.; Mbeunkui, F.; Mitra, S. K.; Mentzel, T.; Goshe, M. B. *J. Proteome Res.* **2010**, *9*, 3621–3637.
99. Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Li, G.Z.; Richardson, K.; Wall, D.; Geromanos, S. *J. Mol. Cell. Proteomics* **2006**, *5*, 589–607.
100. Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. *Mol. Cell. Proteomics* **2006**, *5*, 144–156.
101. Doneanu, C. E.; Monchois, V.; Chen, W. *Paper presented at the 60th Annual Conference on Mass Spectrometry and Allied Topics*, May 20−24, 2012, Vancouver, Canada.
102. Li, G. Z.; Vissers, J.; Silva, J.; Golick, D.; Gorenstein, M.; Geromanos, S. J. *Proteomics* **2009**, *9*, 1696–719.
103. Stemmann, O.; Zou, H.; Gerber, S. A.; Gygi, S. P.; Kirschner, M. W. *Cell* **2001**, *107*, 715–726.
104. Reichert, J. M.; Beck, A. *mAbs* **2009**, *1*, 394–416.
105. Reichert, J. M. *mAbs* **2011**, *3*, 223–240.
106. Beck, A.; Reichert, J. M. *mAbs* **2013**, *5*, 621–623.
107. Doneanu, C. E.; Anderson, M.; Williams, B. J.; Lauber, M. A.; Chakraborty, A.; Chen, W. *Anal. Chem.* **2015**, DOI: 10.1021/acs.analchem.5b02103.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch013

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Chapter 14

Towards a Comprehensive Bioinformatic Analysis of the NIST Reference mAb

Yong J. Kil, Marshall Bern, Kevin Crowell, Doron Kletter, Nicholas Bern, Wilfred Tang, Eric Carlson, and Christopher Becker*

Protein Metrics, Inc., San Carlos, California 94070, United States
***E-mail: becker@proteinmetrics.com**

An emerging need in the biopharmaceutical industry is for new bioinformatics tools that have been developed for comprehensively characterizing therapeutic proteins and efficiently analyzing large sets of liquid chromatography-tandem mass spectrometry and UV data. The focus of this chapter is the demonstration of new software that integrates complementary data and returns results with a high degree of confidence aided by advanced inspection tools to manually verify any result and avoid risk of false or missing information. These software tools (Byologic® and Byomap™) have been applied to the NIST reference mAb for high and low concentration sequences, variants and modifications and associated report elements are presented. They can read any mass spectrometer vendor format, offering a single set of tools to compare information between groups and departments.

Introduction

Pharmaceutical companies bear the responsibility to apply the best analytical techniques to assure safety and efficacy of their products. Counter-pressure to the need for the most advanced and detailed analytics is the corporate desire to reach the marketplace quickly and avoid potential issues with regulatory agencies, and the humanitarian desire to rush life-extending medicines to patients.

This tension is magnified as analytical tools are providing data with ever finer detail and sensitivity, and megabyte files become gigabyte files, etc. A

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

common lament among analytical chemists is that the software is not keeping up with advances in instrumentation, with arguably mass spectrometry providing the greatest advancements and breadth of application and thus the largest need for advances in related data processing software and informatics.

The goal of new bioinformatics tools is not just to keep abreast of the newest technology but also, paradoxically, to provide answers more quickly and easily than ever, despite having to do so with greater detail and accuracy with these more complex data sets.

There is an ever changing landscape of assays too. Notably it has only been in the last few years that there were concerns about sequence variants (*1–3*), but now sequence variant analysis is commonly performed. At the same time there is interest in transitioning tried and true assays like forced degradation studies by peptide mapping analysis to take advantage of more complete mass spectral data in an automated or semi-automated fashion.

This chapter will focus on the analysis of mass spectrometric data, primarily of so-called bottom-up or shotgun data employing digestive enzymes, but it will also include some top-down or middle-down data which can be seen as an orthogonal approach. To exemplify the novel bioinformatics techniques discussed, the chapter focuses on data that has been generated on the NIST reference mAb. The bottom-up data will focus on the following:

- Primary Sequence Verification
- Novel search capabilities
- Sequence variant analysis (SVA)
- Oxidation and glycation
- Glycosylation via intact glycopeptide analysis
- Peptide map annotation and comparisons

An integrated platform of software developed by Protein Metrics is discussed herein that includes:

• Preview™: a pre-search tool for quickly sampling bottom-up data to measure mass errors, digestion specificity and modifications.

• Byonic™: a search engine for peptide/protein identification.

• Byologic®: inspection and quantification software, for analyzing and reporting on low and medium concentration variants and modifications; *the principle software under discussion*.

• Byomap™: peptide mapping software for map annotations and comparisons including "feature finding" and reporting

NISTmAb LC-MS/MS data was generated on a Thermo Fisher Scientific Orbitrap Elite mass spectrometer from digested mAb samples, analyzed at high resolution for the precursors (MS1), and both high and low resolution (separate runs) for the fragments (MS2) using low-energy collision-induced dissociation (CID) or high energy collisional dissociation (HCD) for bottom-up data. Middle-down data used this equipment but also with electron-transfer dissociation (ETD). This NISTmAb data is representative of industry-relevant characterization data and serves as the basis for illustrating these software applications in this chapter.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

The basic question in most any mass spectral analysis, especially when trace abundances or complex structures are involved is: what is present in the sample? After an identification has been made, the question then becomes: is the identification a true-positive or false-positive, and if true, how much of the molecule is present? Thus the bioinformatics method begins with identification. For most applications with therapeutic proteins, the sequence is well known and thus a database search program is an appropriate tool.

The Protein Metrics' database search program Byonic (*4*, *5*) provides a sensitive search and is flexible, allowing many modifications to be searched at once. Byonic is applied to a purified protein in the current chapter, and can also be applied to a complex mixture as done in proteomics experiments. It is the starting point for the other software tools which focus specifically on biopharmaceutical development needs. Byonic has special features such as searching for glycosylated peptides and proteins, searching for amino acid substitutions while allowing for other modifications, and a novel "wildcard" or "blind modification" search. The wildcard search allows any possible modification on a peptide or protein, totally independent of whether the modification is anticipated or even known; therefore the modification need not be found in unimod database of protein modifications for use in mass spectrometry applications (www.unimod.org). Searching all of unimod database unnecessarily will commonly generate avoidable false positives. The Byonic wildcard search can be performed while also allowing common variable modifications such as oxidation, deamidation, and over-alkylation. Thus for example, a wildcard search of the NISTmAb data revealed unanticipated iron adducts. The wildcard search can also detect sequence insertions and deletions. Wildcard searches can be part of a powerful combination strategy for SVA with searches in combination with those with specific amino acid substitutions.

One of the basic requirements for therapeutic protein characterization is that of the primary structure verification. By the time a biomolecule leaves Discovery and enters Development, the target sequence is known; development scientists need to verify that changes in cell-lines and in production scale, have not yielded an undesired primary structure changes due to random mutations or degradation. In general this is not difficult but nevertheless essential. A search by Byonic provides graphical and tabular sequence coverage information, along with probability scoring and tandem (MS2) spectral annotation and inspection. The Byologic and Byomap software programs add to this primary sequence verification by integration with additional tools such as MS-level (MS1) visualization of spectra, automatic quantification of observed peptides by extracted ion chromatograms, placement of sequence within a traditional peptide map, and even fragmentation coverage showing all of the observed cleavage positions from one or many MS2 spectra to prove not just the presence of specific amino acids but their order as well. Some of these views will be apparent as these software tools are elucidated below.

In contrast to primary structure verification, SVA is a situation where special care is needed to identify and quantify variants at trace levels, such as at levels of 1% or even 0.1% relative to the wildtype, form. Because SVA requires reliable detection at such trace levels, even low signal-to-noise and low scoring peptide spectrum matches (PSMs) have to be considered, causing false positives

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

identifications to be generated. Any complete analysis of sequence variants requires that these putative identifications be inspected and that false positives be rejected prior to reporting.

However, low signal is only one of the challenges. SVA, and modification evaluation in general, requires analysis of data sets that may comprise thousands of single-MS and tens or hundreds of thousands of tandem-MS spectra, often coming from multiple sample digests. Standard proteomics software provides at best a partial solution for the analysis of heterogeneous low-abundance components.

For the examples in this chapter, numerous Byonic searches were performed, varying the digestion specificity (fully specific or semi-specific), known modifications (oxidations, deamidations, pyroglutamate (pyro-Glu) N-terminus, cationic Na^{+}, Ca^{++}, Fe^{++} adducts, dehydration, formylation, carbamylation, etc.), and amino acid substitutions, as well as enabling wildcard search (which allows any mass delta within a user-settable range on any one residue or terminus). More than one search can be a useful method for managing the search space and optimizing sensitivity and specificity of a search in cases where many modifications are possible, even though Byonic can handle large numbers of modifications. The Protein Metrics pre-search tool, Preview™ software, was useful before performing the full Byonic search by sampling the data to find common modifications, understand the digestion specificity, and evaluate any mass calibration errors (*6*). Generally, wildcard search is performed with most other modifications turned off. Byonic results were imported into Byologic for subsequent validation and quantification.

Protein Metrics' Byologic® is new bioinformatics software, arguably the first built especially for detailed characterization of biopharmaceuticals, combining not only quantification tools but also inspection capabilities allowing an analyst to reject false negatives and confirm true positives before reporting.

Byologic software presents all the relevant information for evaluating candidate identifications from either a single sample or a "family" of related samples on a single interactive "dashboard", thus improving expert analyst productivity and accuracy. See Figure 1. A coverage view graphically shows spectrum assignments, with modifications highlighted. The Peptide view is part of the software containing an interactive table by which the user selects peptides with a sequence variant, or some modification such as oxidation, deamidation, glycation, or glycosylation, along with the corresponding wildtypes. Once a variant or modified peptide is selected, automatically there are side-by-side graphical comparisons to the wildtype results, including elution profile and extracted ion chromatogram (XIC), annotated MS2 spectra, and MS1 isotope plots. The quantification of variant/modification relative to the wildtype is performed by the ratios of the XICs by label-free comparison. The following discussion focuses on the capabilities and novel features of the software from Protein Metrics for analysis of biotherapeutic proteins, using a single representative dataset on the NISTmAb as the example. It should be noted that relative quantities for post-translational modifications are dependent on sample preparation conditions, instrument settings, and analysis conditions as discussed in more detail in the PTMs chapter/Volume 2, Chapter 3.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

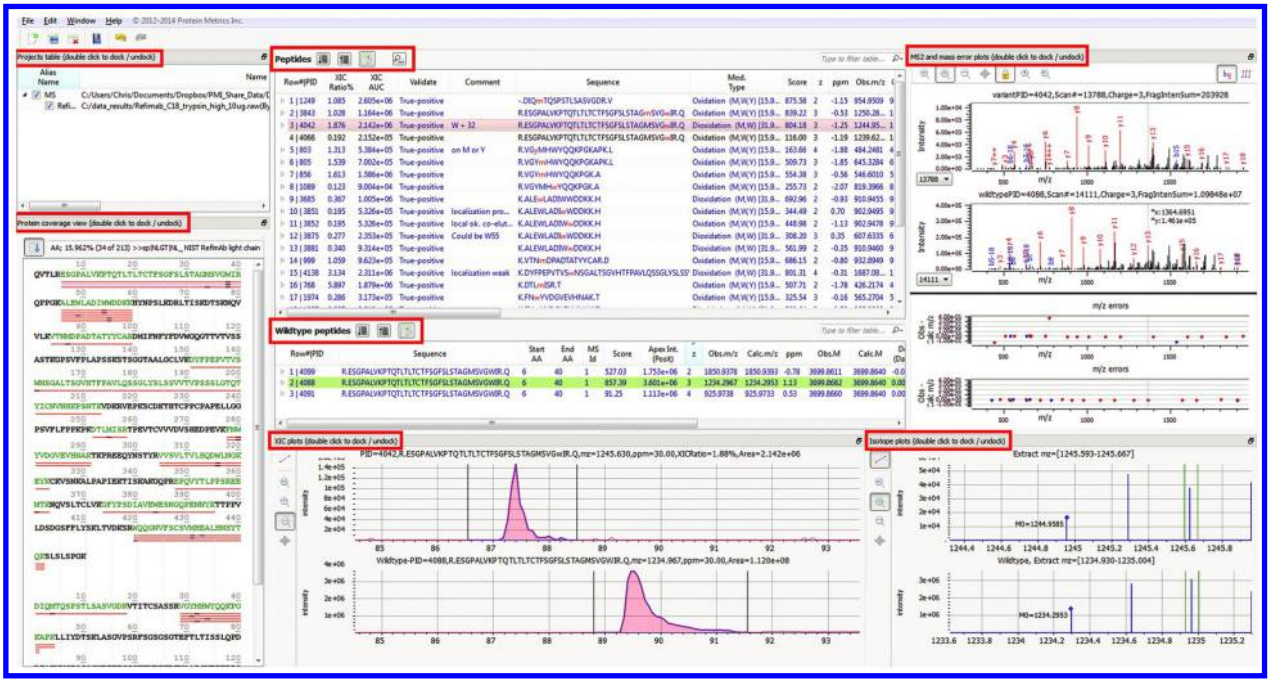

Figure 1. View of the interactive dashboard of Byologic. The red boxes highlight the different "views" of the dashboard. There is a (1) project view for the data files in use, a (2) protein (sequence) coverage view displaying the entire protein sequence along with the observed modified peptides, (3) peptides table and (4) wildtype peptides table with information such as predicted and observed masses, mass errors, XIC areas, retention times and retention time prediction, and modified-wildtype side-by-side plots in the (5) XIC plots view, (6) MS2 view showing annotations and associated fragment errors, and (7) isotope plots view for MS1 data.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Sequence Variant Analysis

In particular for SVA, many or even most putative sequence variants can be rejected by a human expert as false positives due to alternative, more likely, explanations. False positives can arise from such cases as:

- Incorrect precursor mass, charge, or monoisotope setting. For example, a precursor mass off-by-4 Da occurs quite often when the first 3 isotope peaks are locked out by dynamic exclusion; the MS2 spectrum then might match a peptide with a P→T substitution. These false positives are easily distinguished by the MS1 isotope plots.
- Adducts: +22 Da can match Na^+ adducts or D→H. These can be distinguished by accurate precursor mass or elution time.
- Sample preparation artifacts. +28 Da can match in vitro formylation, or S→D, T→E, K→R, Q→R, or A→V. The first 3 possibilities give exactly the same mass delta.
- Oxidations can occur on essentially all amino acids, and for example oxidation of Phe is ambiguous with an F→Y variant, or A→S.
- Methylation can be confused with G→A or S→T.
- Deamidation of N and Q are ambiguous with N→D, Q→E, plus off-by-one monoisotopic assignments can give such mis-assignments in either direction.
- Over-alkylation can confuse with a modification such as G→N at +57 or G→D at +58.
- A substantial table of such exact and near-by possible confusions is available from the corresponding author.

Byologic can save a great deal of analyst time by efficient evaluation of all the relevant data in one interactive dashboard. A serious comprehensive study to detect trace sequence variants has been known to consume weeks of skilled analyst time, involving multiple digests with complementary enzymes (depending on the sample). The Byologic program has been found to drastically reduce this time by providing a consistent, integrated framework to review all relevant data, thus freeing the analyst for other duties, and most importantly, speeding development timelines.

When reviewing data in Byologic, there is no set order to examining the different data fields, but generally the analyst will begin with the Peptides table, which contains the list of all of the putative modified peptides and associated data fields. Data fields (sortable columns) of the table include but are not limited to (1) the amino acid sequence, (2) modification name and position, (3) XIC area and time integration start/stop times, (4) which protein (if there are more than one) and which data analysis file is providing the information (if more than one), (5) retention time maximum and scan times, (6) sequence start and stop amino acid positions within the protein, observed and theoretical *m/z* and total mass and associated mass errors, (7) identification score, (8) predicted retention time and comparison of this prediction relative to the associated wildtype peptide, (9) number of scans associated with the modified and wildtype peptides. The Peptide

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

table can be filtered, for example, to show just sequence variants, deamidation or oxidation or combinations of modifications. Also included within the Peptides table are fields for the analyst to validate or reject the putative identification, a comment field and a field for a reviewer to add their own comments or decision.

The analyst generally has to look at more than one of the other views to come to a conclusion about whether the identification is accurate. But even within the Peptides table itself there may be important clues; perhaps there is one of the potential confusable situations listed in the bullet points above, perhaps the ppm mass error is inconsistent with the errors of know correct identifications, perhaps the predicted retention time is significantly different than that observed. Putative variants that appear only along with another modification yet not by themselves are also generally suspect. These are all examples of "red flags." Next, the analyst might examine the isotope plot. An "off-by-x" error is a clear sign of some misidentification, and this is seen when the obvious monoisotopic peak is not the one indicated in the plot by the blue diamond which reflects the input *m/z* for the search engine. Next could be an examination of the graphical MS2 annotation alongside that of the wildtype peptide. This examination can reveal many different indicators that the identification of the peptide, and the location and nature of the modification, is correct or incorrect. Finally, the XIC plots must be examined not only to check, and if necessary to adjust, the integration window, but also to look for situations such as in-source modification of the wildtype peptide that would rule out a valid identification. In-source modification of peptides typically manifests itself when elution of variant forms coincides with elution of a wildtype peptide (i.e. are artifacts of ionization).

Figure 2 shows an example of side-by-side MS2 annotation comparison of variant and wildtype. There are various graphic tools in the MS2 annotation plots such as a cursor which reveals exact masses of the fragments and displays alignment between fragment ions, plus zooming and panning etc. Figure 3 shows a table of Byologic inspection and quantification results from a SVA search by Byonic from the NIST reference material. These results show all the peptide spectrum matches (PSMs) for two different LC-MS/MS runs from the Orbitrap Elite instrument with CID, one run with high resolution and the other with low resolution MS2 spectra. Note that despite the very low concentrations (as low as 0.02%), there is close agreement in the relative abundance as measured by XIC ratios for the two separate LC-MS analyses. This exemplary variant search was based on 11 commonly observed sequence variants associated with ribosomal translation errors. A more complete SVA would be performed using a larger list of substitutions, potentially all ~ 380 possibilities, plus including results from a separate wildcard analysis by the Byonic search engine.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

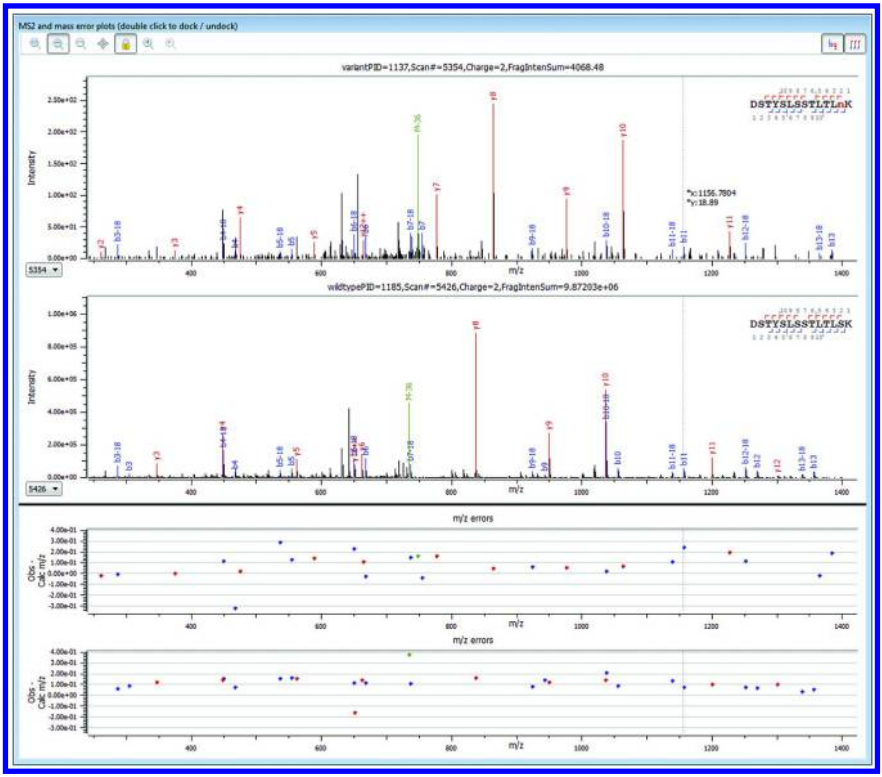

Figure 2. Side-by-side comparison of MS2 annotation of variant (top spectrum) and wildtype (bottom spectrum) for the NIST light chain peptide DSTYSLSSTLTL(S→N)K. The cursor is sitting on b11 which shows correspondence between the variant and the wildtype. All b-ions corresponds until b13, as expected for this site-specific variant. As expected for this variant, all of the y-series shows the 27.0109 Da shift.

XIC Ratio%	XIC AUC	Sequence	Mod. Type	MS res.
0.049	1.32E+05	K.STSGGTAALdCLVK.D	Gly->Asp (G) [58.0055]	high
0.016	4.53E+04	R.TPEVTCVVVDVnHEDPEVK.F	Ser->Asn (S) [27.0109]	high
0.024	6.70E+04	R.TPEVTCVVVDVnHEDPEVK.F	Ser->Asn (S) [27.0109]	low
0.172	6.35E+05	K.TTPPVLDSDGSFFLYnK.L	Ser->Asn (S) [27.0109]	low
0.111	1.29E+05	K.DSTYSLSSTLTLnK.A	Ser->Asn (S) [27.0109]	high
0.179	2.00E+05	K.DSTYSLSSTLTLnK.A	Ser->Asn (S) [27.0109]	low
0.183	2.52E+05	K.HKVYACEVTHQdLSSPVTK.S	Gly->Asp (G) [58.0055]	high
0.134	1.79E+05	K.HKVYACEVTHQdLSSPVTK.S	Gly->Asp (G) [58.0055]	low

Figure 3. Table of SVA peptide spectrum matches (PSMs) for two LC-MS/MS analyses of the NIST reference mAb, collected with high and low MS2 spectral resolution setting and showing their relative abundance (XIC ratio %) relative to the wildtype. MS res. refers to high or low resolution for the MS2 spectra, all in CID mode. There are additional data fields not shown such as observed versus predicted masses, mass errors, retention times, and identification scores.

Oxidation

An analysis was also performed for oxidation of the NIST reference mAb. Oxidation and deamidation discussed below can both occur from sample handling and are a common degradation path for most biologics; any forced degradation or subsequent stability study evaluates these modifications. In the present example, the starting point was a Byonic search for mono-oxidation on methionine, and mono- and di-oxidation on tryptophan.

A number of different oxidations were observed on methionine and tryptophan, as shown in Figure 4 for the heavy chain. It is common for oxidation at specific residues to result in chromatographic elution having multiple maxima; the software can treat these different elution times separately or integrated together. Relative abundances ranged from about 0.1% to 6%.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

XIC Ratio%	Sequence	Mod.Type
1.09	-.DIQmTQSPSTLSASVGDR.V	Oxidation +16
1.61	R.VGYmHWYQQKPGK.A	Oxidation +16
1.54	R.VGYmHWYQQKPGKAPK.L	Oxidation +16
0.12	R.VGYMHwYQQKPGK.A	Oxidation +16
0.19	R.ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGwIR.Q	Oxidation +16
1.90	R.ESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGwIR.Q	Dioxidation +32
1.03	R.ESGPALVKPTQTLTLTCTFSGFSLSTAGmSVGwIR.Q	Oxi +16, Dioxi +32
0.37	K.ALEwLADIWWDDKK.H	Dioxidation +32
0.19	K.ALEWLADIWwDDKK.H	Oxidation +16
0.34	K.ALEWLADIWwDDKK.H	Dioxidation +32
1.06	K.VTNmDPADTATYYCAR.D	Oxidation +16
3.13	K.DYFPEPVTVSwNSGALTSGVHTFPAVLQSSGLYS... ... LSSVVTVPSSSLGTQTYICNVNHKPSNTK.V	Dioxidation +32
5.90	K.DTLmISR.T	Oxidation +16
0.29	K.FNwYVDGVEVHNAK.T	Oxidation +16
0.29	K.FNwYVDGVEVHNAK.T	Dioxidation +32
1.36	R.VVSVLTVLHQDwLNGK.E	Dioxidation +32
0.86	R.VVSVLTVLHQDwLNGKEYK.C	Dioxidation +32
2.14	R.EPQVYTLPPSREEmTK.N	Oxidation +16
0.22	K.GFYPSDIAVEwESNGQPENNYK.T	Dioxidation +32
1.00	R.wQQGNVFSCSVMHEALHNHYTQK.S	Oxidation +16
1.88	R.WQQGNVFSCSVmHEALHNHYTQK.S	Oxidation +16

Figure 4. Table of oxidized peptides for LC-MS analysis of the heavy chain of NIST reference mAb. There are additional data fields not shown such as observed versus predicted masses, mass errors, retention times, and identification scores.

Deamidation and Ammonia Loss

Another important class of analysis is that of deamidation and related ammonia loss. Again, these are modifications which can be strongly influenced by sample handling including deliberate stress tests. The case of deamidation is a special case because of partial isotopic overlap with the generally much more

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

abundant wildtype peptide. Care generally is taken to achieve chromatographic separation between the wildtype and deamidated forms, and there often are more than one elution maximum for the deamidated forms. Figure 5 presents a table of deamidated and ammonia loss peptides and associated percent abundance based on the XIC ratio to the wildtype peptides.

XIC Ratio%	XIC AUC	Comment	Sequence	Mod. Names	Scan Time
0.35	6.68E+04	later eluter, double hump	K.nQVVLK.V	Deamidated(N,Q) [0.9840]	29.00
0.08	2.64E+05	1st peak	K.FnWYVDGVEVHNAK.T	Deamidated(N,Q) [0.9840]	56.04
0.95	3.18E+06	3rd peak, pretty clean	K.FNWYVDGVEVHnAK.T	Deamidated(N,Q) [0.9840]	58.89
0.77	2.58E+06	2nd peak;mix	K.FNWYVDGVEVHnAK.T	Deamidated(N,Q) [0.9840]	58.45
3.72	7.48E+06	late eluter	R.VVSVLTVLHQDWLnGK.E	Deamidated(N,Q) [0.9840]	81.29
13.68	1.35E+07	2nd peak, large deamidation	R.VVSVLTVLHQDWLnGKEYK.C	Deamidated(N,Q) [0.9840]	77.24
1.97	1.70E+06	3rd peak	R.VVSVLTVLHQDWLnGKEYK.C	Deamidated(N,Q) [0.9840]	78.51
8.30	1.10E+07	first peak	K.GFYPSDIAVEWESnGQPENNYK.T	Deamidated(N,Q) [0.9840]	66.23
10.10	1.36E+07	2nd-3rd pks blur, and mix	K.GFYPSDIAVEWESnGQPENNYK.T	Deamidated(N,Q) [0.9840]	67.28
6.84	8.89E+06	2nd peak, mixed, partly resolved	K.GFYPSDIAVEWESNGQPEnNYK.T	Deamidated(N,Q) [0.9840]	67.35
2.04	9.30E+06	very early peak	K.SGTASVVCLLnNFYPR.E	Deamidated(N,Q) [0.9840]	69.81
1.18	5.33E+06	late peak	K.SGTASVVCLLnNFYPR.E	Deamidated(N,Q) [0.9840]	83.12
0.50	6.11E+05	early eluter	K.VDnALQSGNSQESVTEQDSK.D	Deamidated(N,Q) [0.9840]	31.89
1.07	1.32E+06	late eluter but not clean	K.VDnALQSGNSQESVTEQDSK.D	Deamidated(N,Q) [0.9840]	33.37
0.2	1.44E+05		K.VTnMDPADTATYYCAR.D	Ammonia-loss(N) [-17.0265]	51.74
0.2	6.31E+05		K.FnWYVDGVEVHNAK.T	Ammonia-loss(N) [-17.0265]	67.24
0.4	1.36E+06		K.FNWYVDGVEVHnAK.T	Ammonia-loss(N) [-17.0265]	60.65
9.8	2.59E+07		R.VVSVLTVLHQDWLnGK.E	Ammonia-loss(N) [-17.0265]	83.31
9.7	3.26E+06		R.VVSVLTVLHQDWLnGKEYK.C	Ammonia-loss(N) [-17.0265]	79.03
6.4	8.26E+06	two pks	K.GFYPSDIAVEWESnGQPENNYK.T	Ammonia-loss(N) [-17.0265]	68.36
6.4	8.26E+06	two pks	K.GFYPSDIAVEWESNGQPEnNYK.T	Ammonia-loss(N) [-17.0265]	67.50
1.0	7.83E+05		R.WQQGNVFSCSVMHEALHnHYTQK.S	Ammonia-loss(N) [-17.0265]	59.00
0.7	3.16E+06		K.SGTASVVCLLnNFYPR.E	Ammonia-loss(N) [-17.0265]	79.59
0.3	3.99E+05		K.VDnALQSGNSQESVTEQDSK.D	Ammonia-loss(N) [-17.0265]	34.39
0.3	3.99E+05		K.VDNALQSGnSQESVTEQDSK.D	Ammonia-loss(N) [-17.0265]	34.26

Figure 5. Table of deamidated and ammonia loss peptides for LC-MS/MS analysis of the NIST reference mAb sorted on Mod.Name. There are additional data fields not shown such as observed versus predicted masses, mass errors, retention times, and identification scores. The comments field can be used to capture user-defined informal comments related to selection of maxima in cases of multiple elution maxima. Deamidated species generally elute as multiple variant forms: iso-Asp and Asp. Species with iso-Asp generally elute in front, whereas species containing Asp variant elute after of the non-deamidated species. Further, deamidation can produce both L and D enantiomer forms, which can chromatographically separate.

Glycation

The next investigation for Byologic, with Byonic search results and raw data as input, focused on glycation. The preliminary data for trypsin digest of the reference mAb showed several glycation sites, but comparison to the wildtype is inappropriate, or at least problematic, because of the poor trypsin efficiency to cut at lysines modified by glycation. An examination was performed with glu-C digestion where two of these glycations were able to be identified and quantified relative to the wildtype. The sequences and relative XICs are as follows, and a graphical XIC comparison from Byologic for the light chain peptide is displayed in Figure 6.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

E.AKVQW**k**VDNALQSGNSQE.S	light chain	0.34% XIC ratio
L.GCLV**k**DYFPEPVTVSW.N	heavy chain	0.67% XIC ratio

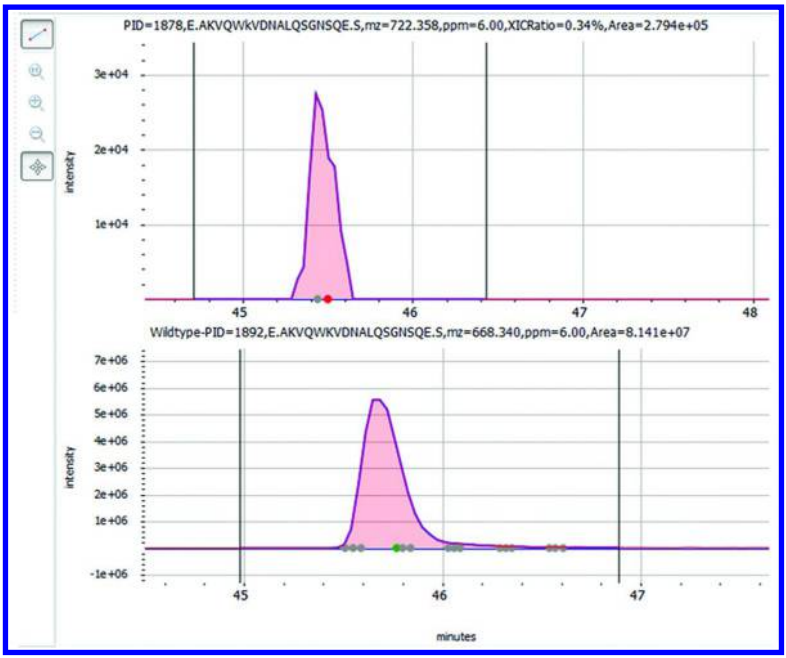

Figure 6. From Byologic, the comparison of XICs is shown for the modified (top) and wildtype (bottom) of glycated E.AKVQW**k**VDNALQSGNSQE.S*, and their relative XIC areas.*

Intact Glycosylated Peptides

The Byonic search engine is unique in its ability to identify intact glycosylated peptides and proteins. Byonic can identify glycopeptides using a variety of information, tuned to the fragmentation method. For example, for ETD fragmentation, it is common to keep the glycan attached and see backbone fragments; this is the optimal situation for identification of the intact glycopeptide. Higher energy collision induced dissociation (HCD) often provides useful information on immonium ions, intact peptide with partial fragmentation of glycans (peptide + glycan fragments), and sometimes on partial backbone fragmentation. Low-energy CID fragmentation does sometimes produce backbone fragment ions, but generally results in water loss which is of no significant value. However there is a special option for Byonic which for HCD/CID allows identification only through the intact mass of the glycopeptide. The Byonic score reflects this level of information and in general both MS2 and MS1 information is used to make the identification. The MS1 mass accuracy is important too, and for this NIST mass spectral data, for HCD spectra, the accuracy for these intact glycopeptides was generally within 1 ppm helping give strong confidence to the assignments.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Although, generally the glycosylation site is known for mAbs, for many other therapeutic proteins it is can be critical to have site-specific glycosylation analysis.

Using Byologic with the NISTmAb HCD data, the relative abundances were determined for the N-glycans for the combination of the overlapping sequences K.TKPREEQYnSTYR.V and R.EEQYnSTYR.V. For these glycopeptides, the XIC curves showed high signal-to-noise, similar to Figure 6. The results are below.

Glycan composition	Proposed as:	XIC areas %
HexNAc(2)Hex(5)	M5, Man5	1.2%
HexNAc(3)Hex(3)Fuc(1)	FA1, M3G0F	6.3%
HexNAc(4)Hex(3)Fuc(1)	FA2, G0F	43.1%
HexNAc(4)Hex(4)Fuc(1)	FA2G1, G1F	39.2%
HexNAc(4)Hex(5)Fuc(1)	FA2G2, G2F	8.4%
HexNAc(4)Hex(6)Fuc(1)	FA2G2Ga1	1.8%

As mentioned, the MS2 spectrum of a glycosylated peptide, and its associated annotation, depend strongly on whether the dissociation method is by ETD, HCD or CID. An example of ETD fragmentation, for a HexNAc(4)Hex(5)Fuc(1) structure (proposed structure FA2G2, G2F) is shown in Figure 7; this data was from blood plasma taken with an Orbitrap Elite instrument (data courtesy of Dr. R. Viner, Thermo Fisher) and analyzed with Byonic software. The NIST reference mAb data for bottom-up analysis was available just for HCD and CID, although the top-down/middle-down data in the next section, including for an intact glycopeptide, was acquired with ETD.

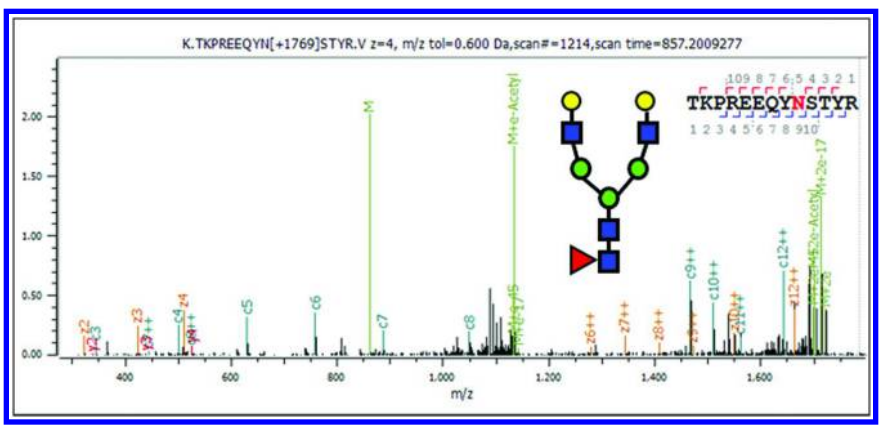

Figure 7. This ETD MS2 spectrum shows a glycopeptide with HexNAc(4)Hex(5)Fuc(1) (proposed FA2G2, G2F) from an IgG.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Top-Down/Middle-Down Mass Spectrometry

The results so far have used digested, bottom-up, protein analysis. A growing orthogonal capability of interest, is to use non-digested protein analysis to perform "top-down" sequencing analysis. An alternative is to use IdeS digestion to produce major ~ 25 kDa components of IgG's, called middle-down analysis. This approach has some advantages by observing the various intact forms of the protein, including with glycans attached. Instrumentation is also rapidly improving for these types of measurements. This approach is expected also to minimize relative ionization efficiencies due to different glycans, or oxidative states, etc., because of the larger capacity of charge formation.

Figure 8 shows an expanded view of a Byonic annotated Orbitrap Elite ETD high resolution MS2 fragment analysis of the NIST reference mAb after IdeS digestion. Multiple isotopes per fragment ion are evaluated before making the assignment, fitting to a theoretical (avergine) isotope distribution before confirmation; and a fragmentation mass table is also available. The requirement for assignment is adjustable but the default setting used here requires observation of the three most intense predicted isotopes for the fragment ion, within a settable mass tolerance, which in this case was 10 ppm. This provides a high degree of stringency. Note that because of the density of the MS2 spectra, high resolution/high accuracy is a practical necessity to generate these types of results. The "fragmentation coverage" is shown by the diagram in the Byonic viewer, and two examples for ETD fragmentation in Figure 9. The Byologic software has a unique capability to generate a fragmentation map by integrating over multiple input sample data files. For these experiments, ETD gave the most extensive fragmentation coverage but HCD and CID gave nearly as complete coverage. The use of high mass accuracy and the examination of multiple isotopic peaks per fragment ion avoids accidental (false assignment). A false discovery rate is tested for this dense spectrum and found to be < 5% by artificially varying the mass of the residues and observing the number of false assignments.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

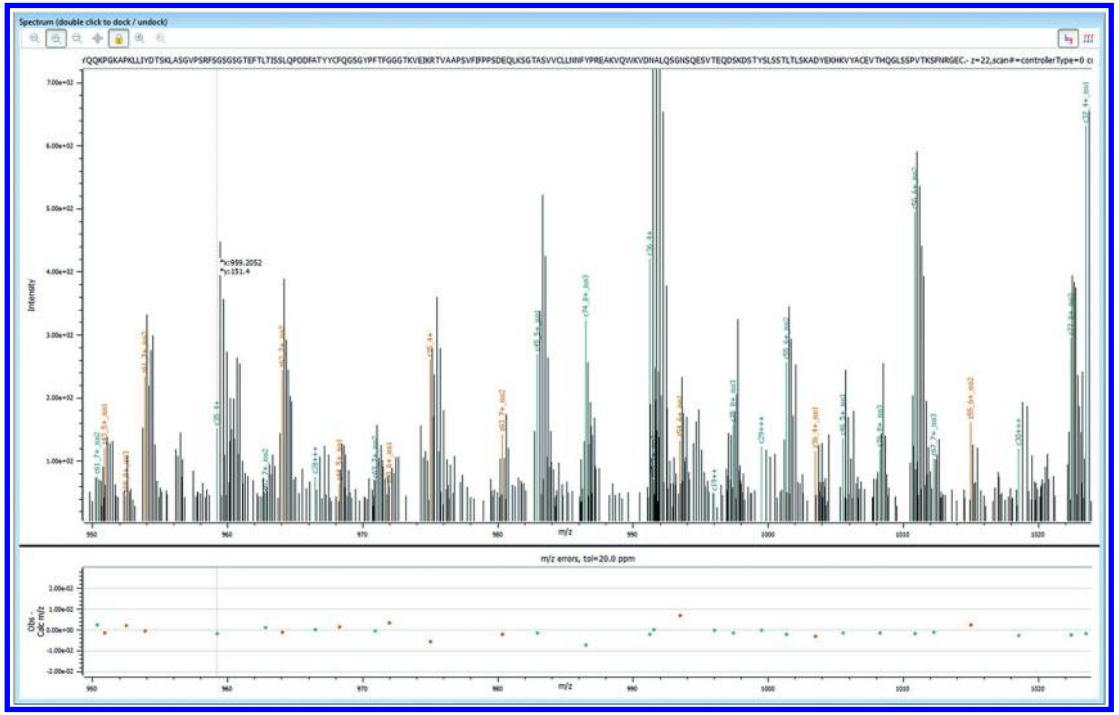

Figure 8. A zoomed-in region of an annotated high-resolution/high accuracy ETD MS2 spectrum in the range of ~ 950 – 1020 m/z for a middle-down spectrum of the NISTmAb light chain generated on a Thermo Fisher Orbitrap Elite instrument. A tabular fragmentation map is an associated export function. The cursor is highlighting a specific m/z peak at 959.2052 with associated MS2 fragment error near 0 ppm.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

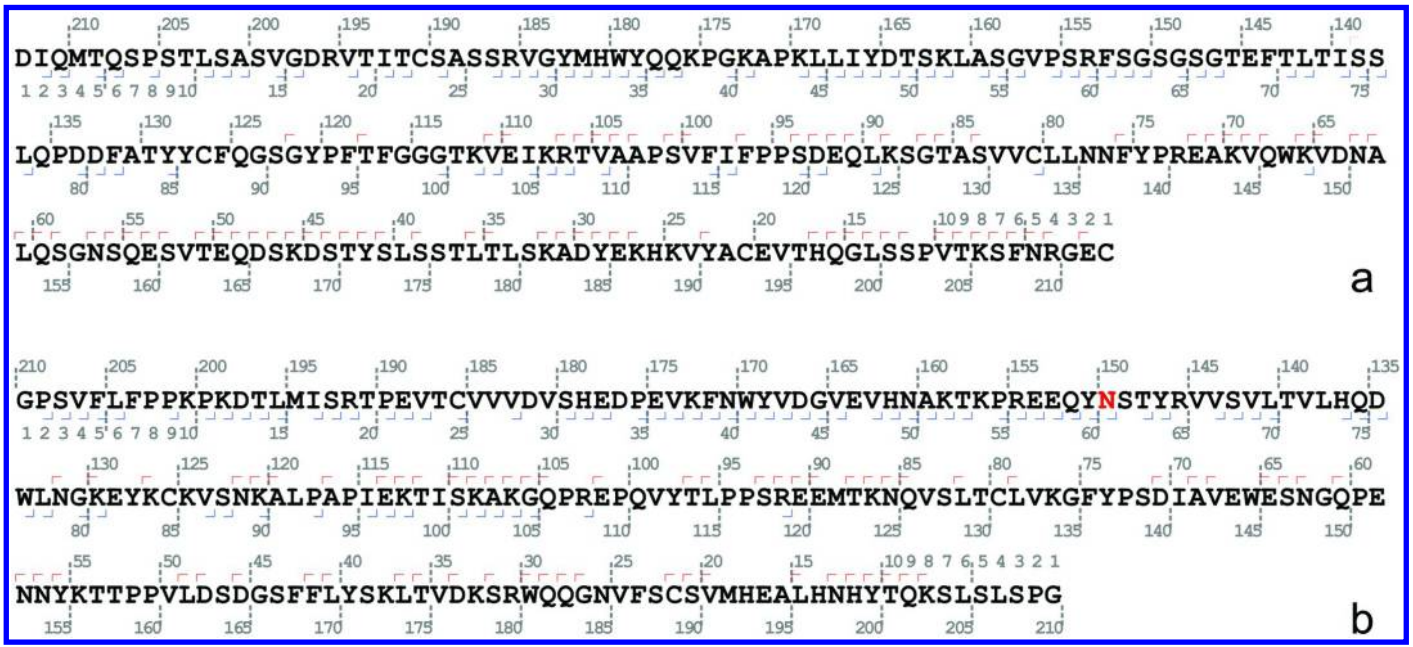

Figure 9. Fragmentation maps by ETD for the NIST reference mAb (a) light chain, charge 21+, observed monoisotopic m/z 1101.641 (b) Fc fragment with intact HexNAc(4)Hex(4)Fuc(1) glycosylation, charge 25+, monoisotopic m/z 1016.307. Note similar spectra are observed from neighboring charge states.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

General Peptide Mapping

So far the analysis presented in this chapter has been focused on new bioinformatics methods to reveal and quantify details about sequence variants, degradants and other post-translational modifications. Peptide mapping is decades old and there are various software programs available for this function but available informatics tools for analyzing biotechnology products have presented a bottleneck in workflow and lack of sophisticated analytics. Thus there are opportunities for new bioinformatics approaches to peptide mapping software to increases the speed and efficacy of (a) annotating a given peptide map, and (b) quantitatively comparing multiple peptide maps. Consequently, a new peptide mapping software called Byomap™ was designed and built. Significantly, Byomap can annotate either TIC or UV chromatograms, and the program can be used to graphically associate UV traces with mass spec data even in cases where the UV data is collected on a separate instrument than the mass spectrometric data.

A peptide map of the NIST reference mAb is seen in Figure 10. A key part of the Byomap™ program is to automatically calculate a baseline, assign peaks, and perform peak area calculations; the analyst may make adjustments if desired. To make annotation easy and reliable, accurate mass MS1 matching to candidates is used, further facilitated by qualifying the time correlation of elution of each of the individual ions in the MS1 spectra with the temporal behavior of the peptide map peak under evaluation. Candidates come from database search results and/or an *in silico* list. The analyst has a choice of annotating manually or in an automated fashion. Manual annotation involves clicking each peak guided by the MS1 view, which displays candidates that are within the selected *m/z* tolerance; see an example in Figure 11. Alternatively, the annotation can be performed automatically, again using MS1 *m/z* tolerance, with the analyst having the option to review and edit any entries. Thus with the automatic approach, a peptide map can be quantified and annotated literally in seconds. Various export functions are available for graphics and tables.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

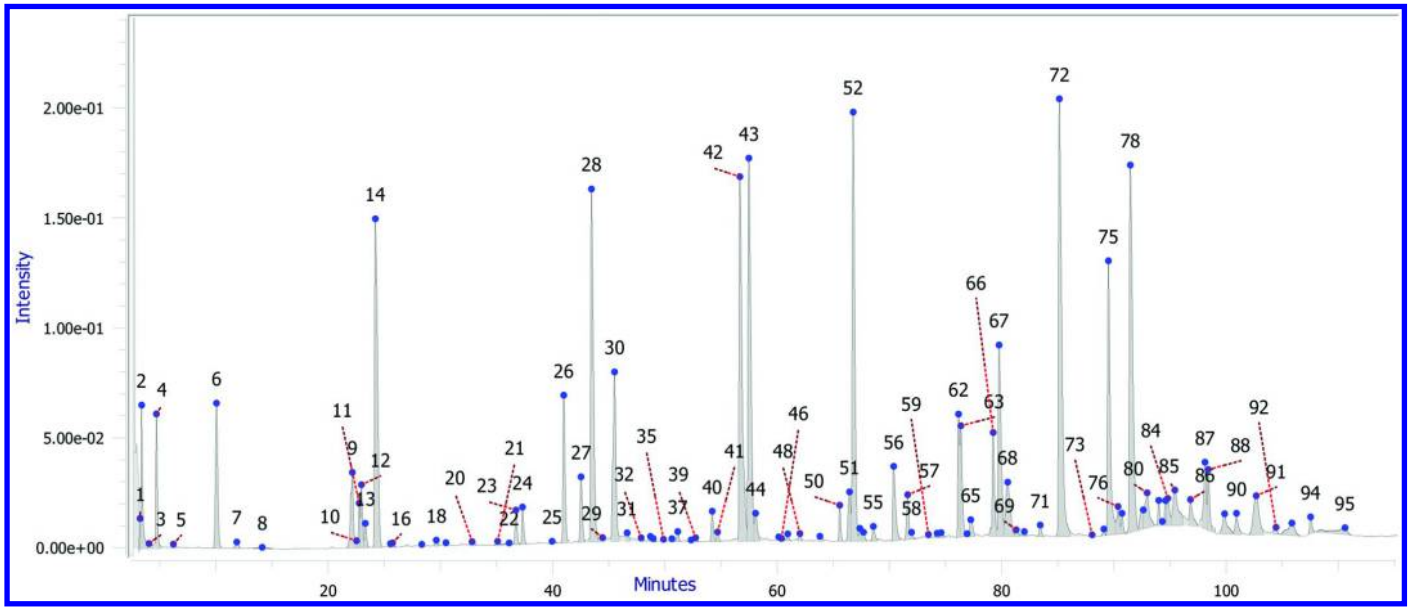

Figure 10. UV peptide map of the NIST reference mAb with automatic baseline, peak labeling and peak integration.

In State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; Schiel, et al.; ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

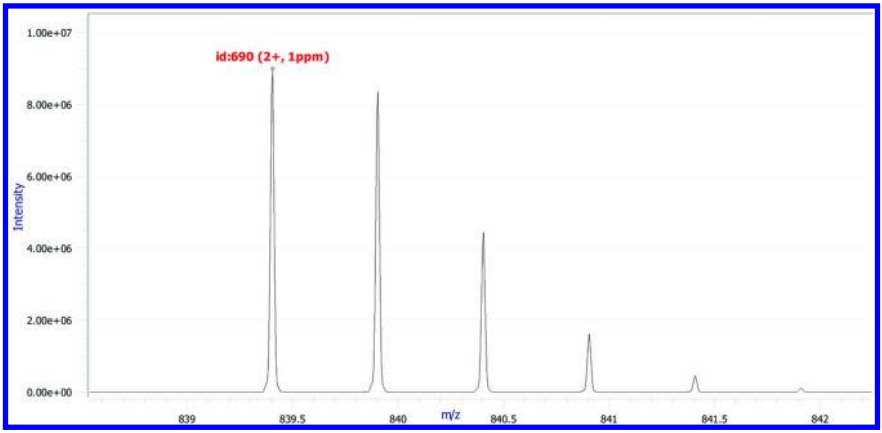

Figure 11. Candidate assignment for the peptide FNWYVDGVEVHNAK with the aid of accurate mass and charge assignment, as well as time correlation, drawn from candidates from database search and/or in silico generated list.

After a reference peptide map has been annotated, it can be used to quantitatively compare up to ten peptide maps to the reference map, again automatically determining baselines and peak areas, quantifying all the peaks and flagging any that are out of a pre-set quantitative range. To help ensure that the proper peaks are being compared to the reference, in the commonly occurring instance of chromatographic shift from LC (or CE) runs at different recordings, a nonlinear alignment function is available to help the analyst confirm proper alignment of all the map peaks. See an example in the zoomed in time window in Figure 12.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

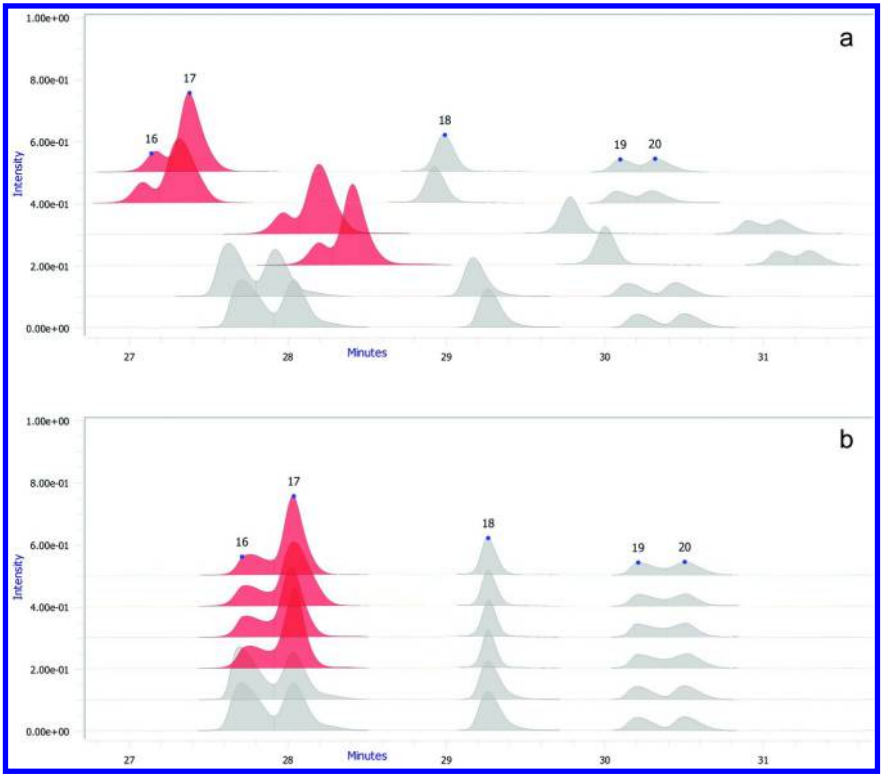

Figure 12. A zoomed in region of a peptide map with 5 different UV maps compared to a reference map (bottom trace). Part (a) shows the raw data with chromatographic shift associated with data acquired days, weeks or months apart. Part (b) shows the same data after nonlinear time alignment to check that the peaks are properly aligned. Note the red-colored highlighted peaks are out of range for peak areas relative to the reference, and hence are being flagged graphically; they are also flagged in the table associated with all the data for the peaks and their assigned peptides (not shown). The peak areas used for determining whether a peptide map peak is out of range, and hence flagged, are always determined before any nonlinear time adjustment to avoid any distortion of the areas.

Summary

To fully capitalize on the rapid advances in analytical instruments such as mass spectrometers, biopharmaceutical researchers require bioinformatics tools able to process and manage the large influx of associated data from these techniques. As described in this chapter, researchers are now using software tools such as Protein Metrics' Byonic, Byologic and Byomap to bring together multiple sets of data to automatically process large sets of data in order to fully characterize biologic materials and advance drug candidates through development milestones.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

With the ability to confidently characterize the product down to trace components identifying degradants, impurities, post-translational modifications, sequence variants, and to quantitatively compare data sets, bioinformatics tools like Byologic and Byomap are also enabling researchers to incorporate these techniques that historically took expert analysts weeks of effort into a routine part of the development process. This allows development groups to share data across their organization and to quickly report on comparability. Using these advanced bioinformatics tools to turn data into knowledge supports regulatory filings and advances Quality by Design (QbD) initiatives to track quality attributes as products move through development.

Acknowledgments

The authors thank Drs. John Schiel, Trina Formolo and Lisa Kilpatrick of NIST for providing the raw data and helpful discussions. The authors also thank Dr. Boyan Zhang, Dr. Richard Seipert and Mr. Michael Kim of Genentech for their help with the development of the Byologic and Byomap software.

References

1. Guo, D.; Gao, A.; Michels, D. A.; Feeney, L.; Eng, M.; Chan, B.; Laird, M. W.; Zhang, B.; Yu, X. C.; Joly, J.; Snedecor, B.; Shen, A. *Biotechnol. Bioeng.* **2010**, *107*, 163–171.
2. Yang, Y.; Strahan, A.; Li, C.; Shen, A.; Liu, H.; Ouyang, J.; Katta, V.; Francissen, K.; Zhang, B. *mAbs* **2010**, *2*, 285–298.
3. Zhang, Z.; Shah, B.; Bondarenko, P. V. *Biochemistry* **2013**, *52*, 8165–8176.
4. Bern, M.; Cai, Y.; Goldberg, D. *Anal. Chem.* **2007**, *79*, 1393–1400.
5. Bern, M.; Kil, Y. J.; Becker, C. *Curr. Protoc. Bioinf.* **2012**Chapter 13, Unit13 20. DOI:10.1002/0471250953.bi1320s40.
6. Kil, Y. J.; Becker, C.; Sandoval, W.; Goldberg, D.; Bern, M. *Anal. Chem.* **2011**, *83*, 5259–5267.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch014

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Chapter 15

A Global Partnership Advancing Biopharmaceutical Development: Summary and Future Perspectives

John E. Schiel,*,[1] Michael J. Tarlov,[1] Karen W. Phinney,[1] Oleg V. Borisov,[2] and Darryl L. Davis[3]

[1]Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States
[2]Novavax, Inc., Gaithersburg, Maryland 20878, United States
[3]Janssen Research and Development, LLC, Spring House, Pennsylvania 19002, United States
***E-mail: john.schiel@nist.gov**

Upon reading the preceding book chapters, it is clear that comprehensive analysis of monoclonal antibody therapeutics is no easy task. These molecules embody various complex attributes, the characterization of which is a long and arduous process, yet monoclonal antibody therapeutics have taken residence as perhaps one of the most influential therapeutic classes of our time. The intent of this book series was to provide a comprehensive overview of monoclonal antibody therapeutics, using the NISTmAb as a vehicle for highlighting the characterization stages of product development. The preceding chapters represent a collaborative effort among biopharmaceutical professionals rising to this challenge. Contributors utilized the NISTmAb throughout, demonstrated the potential utility of class-specific reference materials as a means to facilitate open innovation, and identified a number of emerging research areas for future development. Conclusion of the series is therefore met with eager anticipation of continued biopharmaceutical advancement through industry-focused partnerships.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Summation and Perspective

Volume 1, Chapter 1 was written with the mindset of introducing the concept of a class-specific reference material and begins to highlight the potential utility of such a material. Reflecting upon this chapter nearly a year after its initial writing, the potential utility of these materials is even more evident after completing this series. The authors of the chapters contained herein have embraced this concept of widely available reference materials and have helped to guide both industry best practices and a new paradigm in thinking about collaborative characterization efforts around similar molecules. When looking at the series in its entirety, it is evident that each author brought his or her unique perspective to monoclonal antibody development, characterization, and regulatory challenges. It also is evident how each author embraced the book series concept in its entirety and used their expertise and individual chapter to fuel a cohesive, comprehensive description. When reading the series, the authors have worked together not only within chapters but between chapters to tie inter-related concepts together.

The molecular properties and unique functional activity by which immunoglobulin Gs (IgGs) convey immunity is described in Volume 1, Chapter 2. The intricate heterogeneity of these molecules is reviewed, and their unique mechanism of action is demonstrated. The attributes of these molecules are then described in Volume 1, Chapter 3, and the full heterogeneity of the molecule appreciated. The authors also introduce the concept that proper development of a monoclonal antibody must include more than initial characterization; instead, various factors throughout the life cycle must be considered as to how they may affect molecular properties and, ultimately, the fitness of the drug product. These further tie directly into Volume 1, Chapter 4, which highlights assays capable of monitoring the level of heterogeneity described and a detailed summary of current state-of-the-art techniques one may consider during monoclonal antibody characterization. Volume 1 is rounded out by reminding the reader that consideration of molecular properties is not only an analytically driven process; full understanding of the biology, measurement capability, and clinical impact also come into play.

It is clear from Volume 1 that successful therapeutic development is a summation of all stages from process to product to patient. One critical component to development of a successful product is detailed characterization of its various attributes, including identity, purity, stability, and activity. Physicochemical, biophysical, and bioactivity data feed upstream to determine what process/product changes might improve various attributes of the product. These data also feed forward into formulation, container closure, the in-house reference standard plan, and so forth. Prior to regulatory approval, the totality of these data must be submitted to regulatory agencies to demonstrate that the product can be reproducibly produced to within defined specifications and ensure its continued safety and efficacy. A large portion of this characterization data is therefore presented to regulatory agencies, as would be done in a Biologics License Application (BLA). Volume 2 of this series encompasses a great deal of what would be expected under the S3.2 Characterization section of the BLA. Although the current book series is not intended to be a representative guideline

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

for regulatory approval, it does include a detailed discussion of many of the various methodologies and resultant data that contribute to demonstration of a well-characterized product. Table 1 is a reproduction of a table from Volume 1, Chapter 4, which lists a summary of techniques that often are used for characterization studies. NISTmAb analysis methodology and data for the studies highlighted in bold are contained throughout Volume 2, and associated chapters listed for reference of each technique. The collective analysis throughout Volume 2 therefore provides a very holistic picture of the NISTmAb reference material (NISTmAb RM), as well as the current state of monoclonal antibody characterization.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Table 1. Example Characterization Method Summary

Category	*Purpose*	*Technique*[a]	*Volume 2 Locations*[b]
Primary Structure	**Mass confirmation based on sequence, assessment of mass variants**	**Native, deglycosylated, and reduced protein electrospray ionization-mass spectrometry (ESI-MS)**	V2:C1
	Amino acid sequence confirmation	**Tandem mass spectrometry (MS/MS) sequencing of peptide maps**	V2:C1, 2, 3
	N- and C-terminal variants	**MS/MS sequencing of peptide maps**	V2:C1, 3, 5
	Other post-translational modifications	**MS/MS sequencing of peptide maps**	V2:C3, 5
Glycosylation	***N*-linked glycosylation**	**Peptide or glycan map with mass spectrometry (MS) or MS/MS, permethylation with MS^n**	V2:C1, 3, 4
	***O*-linked glycosylation**	**Peptide map with MS**	V2: C3
	Attribute criticality	Enzymatic digest, bioassay	
Disulfide Structure	**Identification of disulfide-linked peptides**	**Non-reduced and reduced peptide map with MS**	V2: C1
	Free Free sulfhydryls	Ellman's assay (or similar)	

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Example Characterization Method Summary

Category	*Purpose*	*Technique*[a]	*Volume 2 Locations*[b]
Charge Variants	**Charge heterogeneity**	**Cation exchange high-performance liquid chromatography (CEX-HPLC)**	V2:C5
	Charge heterogeneity	**Capillary isoelectric focusing (cIEF)**	V2:C5, 7
	Attribute criticality	Purification and characterization of charge variants	
Size Variants	**Size heterogeneity under native conditions**	**Size exclusion chromatography (SEC)**	V2:C5, 6, 7
	Size heterogeneity of reducible forms	**Reduced capillary sodium dodecyl sulfate electrophoresis (rcSDS); reduced, denatured SEC (rdSEC)**	V2:C5, 7
	Size heterogeneity of covalent forms	**Non-reduced capillary sodium dodecyl sulfate electrophoresis (nrcSDS), denatured SEC (dSEC)**	
	Determination of monomer, dimer, and submicron aggregates	**Resolution by SEC with static light scattering detection (SEC-SLS)**	V2:C5, 8
	Size heterogeneity and sedimentation coefficient determination	**Sedimentation velocity analytical ultracentrifugation (SV-AUC)**	V2:C6
	Attribute criticality	Purification and characterization of size variants	
Biophysical Characterization	**Secondary structure**	**Fourier transform infrared (FTIR) spectroscopy**	V2:C6
	Tertiary structure	**Near-UV circular dichroism (near-UV CD) spectroscopy**	V2:C6
	Thermal stability	**Differential scanning calorimetry (DSC)**	V2:C6,7

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Table 1. (Continued). Example Characterization Method Summary

Category	*Purpose*	*Technique*[a]	*Volume 2 Locations*[b]
Biological Characterization	Justification of bioassay relevance to mechanism of action (MOA)	Bioassay(s)	
	Probe relevant functional domains	Receptor binding assays	
	Assess other relevant biological functions (e.g., complement-dependent cytotoxicity [CDC], antibody-dependent cellular cytotoxicity [ADCC])	Other characterization biological assays	

[a] cSDS also referred to as CE-SDS in V2:C5; r, reduced; nr, non-reduced; d, denatured. [b] Book locations are listed in Volume X: Chapter Y format (VX: CY).

The data contained herein therefore represents a historical set of baseline data, collected using state-of-the art platforms, which can be repeated in the reader's laboratory on the same material. Each of the methods presented in Table 1 entail critical steps and multiple variables that may alter the method performance. The methods used in this series are predominantly initial platform methods that each participating laboratory would perform on any new product. This is quite advantageous in that it gives a realistic cross section across multiple laboratories as to the various steps that may differ in sample preparation, analysis, and data interpretation. The authors of each individual chapter have discussed these significant factors and have provided discussions of the various intricacies of each analysis that may ultimately be optimized in a product-specific manner.

Considering the widespread future availability of the NISTmAb, it is expected that various characterization activities will undoubtedly continue, and an even larger repository of data and methods will be generated in the literature to follow this series. For example, novel instrument platforms will likely lead to optimized methods for evaluating NISTmAb-specific attributes. Many of the methods described above also may be further qualified for their intended use in assessing the identity (e.g., peptide mapping), purity (e.g., capillary isoelectric focusing [cIEF]), monomeric purity (e.g., size exclusion chromatography [SEC]), and stability (dependent on attributes) of the NISTmAb. Forced degradation studies may be performed in order to further elucidate potential degradation pathways and production of product-related impurities relevant for challenging methods during qualification exercises. One could imagine full validation of identity-, stability-, and purity-indicating methods for the NISTmAb according to International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guidelines. Such

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

validation is necessary in the regulatory environment to ensure consistency of product quality, and therefore robust performance criteria along with metrics for this critical reference material may be of great use to the end-user. Although not intended to be a regulatory exercise, method development endeavors at this level would operate under the same controls in order to simulate industrial challenges and thereby produce qualified and/or validated methods for the NISTmAb. Such assays are undoubtedly necessary for assuring the stability and quality of the material over time. As an impact to the industry, a well-characterized NISTmAb RM may further assist in establishing/harmonizing platform strategies for this class of molecules and streamlining the development and application of emerging technologies by providing data that co-evolves with industry state-of-the-art.

Although the current book series has covered analytical technology for monoclonal antibody characterization very comprehensively, not every orthogonal method for elucidation of various attributes of protein therapeutic was applied, nor would this have been a realistic or particularly fruitful endeavor. Rather, the methods which are arguably most informative and routinely used by industry were utilized, and an overwhelming majority of NISTmAb attributes were covered. Table 1 does, however, highlight two very important and inter-related components of characterization that deserve mention that were not attainable and/or relevant to the focus of the book series. The first is determination of attribute criticality. Although developability and forced degradation studies were performed, a more intensive set would likely be conducted not only to evaluate stability but also to link critical attributes to biological consequence through in vitro assays. This second important characteristic of biological behavior of such a protein therapeutic is of course its ultimate application; however, the NISTmAb shall not be used for any human or animal use, and therefore, related in vivo assays that may otherwise be performed with a true therapeutic candidate are not optional. Therefore, in vitro target binding affinity, effector function activity, and immunogenicity may be of particular interest in determining attribute criticality. This realm of product characterization is undoubtedly equally as important, and intrinsically intertwined, with physicochemical and biophysical product characterization.

It is of course the ultimate pursuit of analytical and biophysical characterization methods to produce more information-rich data with fewer assays and greater accuracy and precision. This driver spurs development of novel methodology and technology that continue to emerge in supplement of the already expansive characterization toolbox. It is true that an all-encompassing method would improve time to market and give the developing company an edge in the fast-paced development world. However, when discussing the technology with innovating experts in the field, the driving force remains consistent with the aspirations that once led them to the health-care field initially: the improvement of health care through safe and effective treatments, as ensured to the highest degree possible with the most innovative characterization tools available.

This driving force brings us to Volume 3, where experts describe cutting-edge technology being developed for use in the biopharmaceutical industry. It is clear from Volume 3, Chapters 2, 3, and 4 that higher order structure is of great interest in the characterization of biopharmaceuticals. The primary structure of

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

a protein is of course the foundation upon which higher order structure is built. The same nominal primary structure, however, may theoretically adopt a range of secondary, tertiary, and even quaternary structural formats. Proteins also are dynamic species in vivo and may adapt their overall conformation. Technology has now advanced to a level where the intermediary effects of a given change (whether it be an environmental stress or compositional change) can be evaluated in relation to changes in higher order structure. It is this three-dimensional information that truly may provide the lock-and-key understanding of structure and function and ultimately bear insight into improved therapeutics designed to have higher specificity and activity. Volume 3, Chapters 2, 3, and 4 are therefore dedicated to describing these emerging technologies, including nuclear magnetic resonance (NMR) spectroscopy, hydrogen-deuterium exchange mass spectrometry (HDX-MS), covalent labeling mass spectrometry (MS), and ion mobility MS. The evaluation of a variety of higher order structure methods in tandem with molecular modeling is proving to be an effective tool for providing a very high level of information related to product structure/function relationships, which can lead to improved control strategies or formulation design.

The principles of higher order structural methodology and the capabilities of these techniques provide a segue into discussion of protein aggregation. Each of the higher order structure technologies described provide analysis of and eventual insight into characterizing aggregated species. Additional emerging technologies that target aggregation are described in Volume 3, Chapter 5, including transmission electron microscopy (TEM), atomic force microscopy (AFM), and X-ray and neutron scattering. The impressive toolbox of emerging technologies for studying aggregation has provided a detailed understanding of the various mechanisms associated with aggregation; however, the risk associated with their formation continues to provide the impetus for more technological developments. In addition to a detailed understanding of aggregation mechanisms and aggregate properties, the rapid identification and screening of species prone to aggregation is important. Light scattering techniques are a particularly useful method for this analysis, and a novel technology for rapid, independent conditional screening is described in Volume 3, Chapter 6.

Considering the exhausting quantity of data described to this point from both state-of-the-art and emerging technologies, the interpretation and organization of such a wealth of information is of course no small task. As a representative example, MS is one technique that has played a significant role throughout the characterization assays described in this series. Although not a new technique, the level of information, novel capabilities and application, and pace at which data can now be collected are representative of the entirety of the characterization toolbox. As a representative discussion of the emerging informatics needs and technologies on the industry, a biopharmaceutical-focused discussion on MS-related informatics is provided in Volume 3, Chapter 7.

The informatics and technology for collecting all of the data are often overlooked when discussing the characterization of therapeutic proteins. Intricately linked to the biopharmaceutical manufacturing industry is the instrument and technology manufacturing industry that provides the robust tools for protein characterization. These tools cover a great operational space, including

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

raw materials and novel testing applications, advanced instrumentation, and data interpretation and analysis. The partnership between instrument vendors and biopharmaceutical developers is not only demonstrated by the commercialization of robust and directed instrumentation but also present through the dedicated research efforts of their application scientists. The final chapters of Volume 3 are representative of the high-caliber research that goes into instrumentation development and highlights a series of emerging microfluidic technologies, mass spectrometers and their applications, and laboratory automation. This industry-targeted effort greatly facilitates biomanufacturing and health care through improving the functionality, informational content, and organization of data relevant to product understanding and regulatory approval.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Reference Material and Characterization Technologies: Selected Lessons Learned

The NISTmAb RM used as the topic of this book series has proven useful in providing a representative sampling of the current technologies. The editors are hopeful that each contributing author took something unique away from the collaborative effort among individual chapter authors. Throughout the book, it has been clearly demonstrated that the current characterization toolbox can provide a very high-quality assessment of a biological protein's characteristics. The intent is that the material will reside as a commodity for discussion when developing improvements to our current state of characterization technologies. It is intended to be a medium of innovation by which various companies, academic researchers, instrument vendors, and regulators alike can devise strategies to best suit the needs of the task at hand. With that goal in mind, a select number of potential areas where continued investigation may benefit the biopharmaceutical community are described below, along with potential avenues for future studies.

Heterogeneity Assessment and Quantification

One of the most challenging aspects of product characterization is identifying as many product-related impurities as possible. A complete assessment of all of the heterogeneity of the material clearly requires the use of a variety of orthogonal techniques. A detailed assessment of amino acid side chain modifications/alterations is one of the main focuses of this effort. Biophysical and separation science assays are often correlated to some degree with peptide mapping results. Peptide mapping is an extremely powerful method that is to date irreplaceable in biopharmaceutical development. However, data from Volume 2 highlights the susceptibility of resultant data to minor changes in sample processing, analytical method, or data processing. For example, a sample processing issue may result in creation or removal of a post-translational modification (PTM) that may otherwise be detected upon optimization of digestion conditions tuned to measure that modification. As discussed to some extent in the PTMs chapter/Volume 2, Chapter 3, the NISTmAb will serve as

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

a perfect test sample for future round-robin studies to specifically differentiate sample preparation from method-related artifacts.

Beyond detection, quantification using extracted ion chromatograms (XICs) provided an interesting discussion not only in data collection but also in interpretation. It was found that different analysts, as well as different software manufacturers, interpreted MS data in various ways. This included using multiple isotopes, multiple charge states, missed cleavages, and so forth. The issue is further complicated depending on the enzyme chosen and the particular modification being discussed. Again, XIC quantification is an indispensable technique but one with several nuances that must be considered by the user and the reader of a particular data set. A series of interlaboratory studies targeted at quantification would be useful in not only further identifying these nuances but also clearly reporting them in the literature as they relate to biopharmaceutical development. This is important not only for instructional purposes but also to ensure complete reporting of method details to provide the reader with the necessary tools to adequately interpret the results. Further development of "smart" informatics tools to automate PTM quantitation beyond building XICs and integrating areas under the curve to produce tools that will be able to distinguish between true PTMs and method-induced artifacts such as on-column or in-source induced oxidation will undoubtedly benefit and simplify PTM analysis using peptide mapping with liquid chromatography-mass spectrometry (LC-MS).

The planning of studies to evaluate PTM identification and quantitation provide a very unique metrological challenge. Approaching such an endeavor would require material with known quantitative values or at least some with consistent levels of PTMs. Options for such studies may include forced degradation samples provided along with highly detailed analysis procedures and/or predigested samples with spiked peptides of known degradation quantity. Ultimately, certification of such materials with absolute values would be of the most interest; however, this is also a difficult task that would likely only be achievable with associated validated sample preparation protocols. It should be noted that even without certified concentrations, such degraded and/or altered materials may be useful for challenging methods with a common material to evaluate inter-method performance

Although peptide mapping is discussed above as an example, it should be noted that all assays in Volume 2 could benefit from highly defined interlaboratory studies targeting a unique aspect of a given method. Such studies are very difficult and often require years of planning to ensure methods are carried out appropriately in the testing laboratories. However, these results can provide metrological understanding of the most important factors in an analysis and therefore more sound measurement science in the future.

Higher Order Structure

As mentioned above, advances in studying higher order structure continue to be of critical importance to improving our understanding of the true behavior of protein molecule in situ. An understanding of the higher order structure of these molecules and dynamics would benefit our understanding of the interactions with

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

binding partners (e.g., antigen or effector function protein) and allow design of more optimized products. Improved higher order structural techniques also would allow for identification and evaluation of protein misfolding. An increase in the sensitivity and resolution of such methods would allow identification of small and localized changes that may affect function but may not be readily identifiable with bulk methods. Related to protein misfolding is of course aggregation, both of which are conditions that can lead to unwanted immunogenicity and/or a decrease in functional activity. Similar to binding activity, higher order structure techniques can assist in identifying changes that are associated with aggregation, can be used to identify problematic species when present, and further can be used to identify methods to engineer products or formulations to prevent their formation. The importance of the effect of protein form on protein function throughout the drug development life cycle will undoubtedly continue to fuel development of novel higher order structure techniques. Maturation of in silico computational algorithms for prediction of aggregation-prone sequences will likely also occur along with improvements in high-resolution techniques for examination of higher order structure of proteins.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Host Cell Proteins (HCPs) and Other Impurities

The Bioinformatics chapter/Volume 3, Chapter 7 highlighted the concept that it is much simpler to look for what one knows is present rather than find a complete "unknown," especially when that unknown is a minute component of the mixture. It is inherently impossible to screen specifically for all potential impurities simultaneously, and therefore targeted strategies based on appropriate risk assessment have proven to avoid contamination issues. The emerging technologies section highlighted two areas, HCPs and adventitious agents, where the researcher must search for an unknown in a very high quantity of known material and provide emerging pathways for continued improvement in analytical technologies. In both cases, very robust and reliable methods are available; however, they both have defects that limit the extent to which they can provide information, potentially outgrown by the modern biotechnology field due to the advances in science and increases in regulatory expectations.

ELISA-based HCP assays, for example, are robust and sensitive assays for the detection of the total content of residual proteins. However, they are based on polyclonal antigen binding to the target and are hence inherently limited by the quality and specificity of the animal-derived anti-HCP reagent used, which is tightly linked to the method of its production. This results in difficultly in comparing results even when seemingly similar reagents and methods are utilized. Estimation of the coverage of these species is also a difficult task, as many proteins may not elicit a response in the initial sensitization and development of the anti-HCP lot in the host. For this reason, MS has emerged as novel technology to identify and quantify HCPs. The selectivity of MS can be much higher, and MS offers an orthogonal measure. The robust performance of such an assay must continue to be evaluated, however, and challenged with real-world samples. These methods have advanced to a significant degree, as demonstrated in this series, and may one day become a more routine component of HCP analysis.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Challenges with a Longitudinally Available Material

The previous series and discussion of this chapter gives a summary of the good that can result from the availability of a widely available, class-specific biopharmaceutical reference material. A variety of potential uses were described in Chapter 1 of Volume 1, the class-specific and representative nature of the NISTmAb is expounded in Volume 1, and a significant level of product understanding is presented throughout the remainder of the series by way of biophysical and analytical characterization data. The characterization data presented is at an extremely high level; considering the various opportunities provided by such a material, however, it goes without saying that it also represents unchartered territory. It is evident that there are many subtle variation in each of the methods described that may in some cases lead to related, but ultimately differing, conclusions with respect to absolute values or specific attributes. This challenge is and will become evident with reports that inevitably will conflict—albeit moderately—on the NISTmAb properties. At such times, it will be critical to report to the greatest degree possible all experimental methods and data used to draw the said conclusions so as to adequately discuss the method-related aspects separately from molecular attributes. It is expected that such initially conflicting data sets may provide for healthy discussions as to the viability of various test methods, and ultimately, areas for improvement and further metrological evaluation/improvement will be identified. For safety purposes, it is the consistency in a measurement and the relation to the clinical outcome that is important. The question that remains unanswered is which methods can (or should) be refined to more absolute rather than relative measurements. The hope is that the availability of reference materials such as the NISTmAb will allow a means for more intricate analysis of these possibilities.

Looking Back and Looking Ahead: A NIST Perspective

The Biomanufacturing Program at the National Institute of Standards and Technology (NIST) was formally initiated in 2012. The aim of the Program was, and continues to be, to develop fundamental measurement science, reference data, standards, and new technologies to support the development, manufacturing, and regulatory approval of protein therapeutics. These types of infrastructural measurement tools are expected to promote development of new, innovative protein therapeutics; facilitate the development of biosimilars; and scientifically underpin regulatory decisions. A hallmark of the Program has been outreach to the biopharmaceutical industry, the Food and Drug Administration (FDA), and other stakeholders to help identify the key measurement issues involving characterization of protein therapeutics. The scientific and technical activities of the Program are focused on industry-identified measurement problems and are grouped into three broad areas: protein stability, protein structure, and production cell variability. The scale and scope of the Program has grown considerably over the past 3 years and now includes the participation of more than 30 scientists from across NIST. The activities of the Program take advantage of the diverse range of scientific and technical expertise and facilities available at NIST and include the

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.
Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

use of neutron scattering methods to understand the conformation, stability, and early stages of aggregation of protein therapeutics; the development of protein particle reference materials to calibrate and understand the bias of different protein particle measurement technologies; the development of two-dimensional (2D) NMR methods to characterize protein therapeutics with atomic resolution; and the invention of a high-throughput, microfluidic-based device that can rapidly measure the viscosity of liquid formulated protein therapeutics. Three years into the Biomanufacturing Program, we believe that a firm foundation has been established at NIST for a long lasting core competency in measurements and standards for biologic medicines.

In the fall of the same year that the NIST Biomanufacturing Program was established, a serendipitous confluence was set into motion, ultimately leading to the current book project. Mike Tarlov was contacted by Darryl Davis of the Janssen Pharmaceutical Companies and Oleg Borisov of Novavax with regard to NIST's interest in participating in the development of an American Chemical Society (ACS) book on therapeutic monoclonal antibody characterization. The invitation could not have come at a more opportune time because NIST was about to receive a significant quantity of an IgG1 monoclonal antibody material as a result of a kind contribution by MedImmune LLC. The intent was to develop this IgG1 material, now known as the NISTmAb, into a NIST reference material, an effort being led by NIST scientist John Schiel. In further discussions with Darryl and Oleg, marriage of the two aligned missions was decided on as an ideal means forward. Distribution of the NISTmAb material to book participants began in late 2013, the analytical characterization of which would be the basis for much of the book series. The publication of this third and final volume marks the culmination of this remarkable collaborative effort, involving nearly 100 scientists from the biopharmaceutical industry, academia, FDA, NIST, and other organizations. We are especially grateful to the co-editors and co-authors and for the privilege of being involved in the book series. The experience has taught us much and has profoundly shaped NIST's thinking about the development of protein reference materials. Many valuable lessons and best practices have been learned, and together they constitute a model for NIST in developing future reference materials not only for biopharmaceutical products but also for other application areas such as clinical diagnostics.

What are some of the things that NIST learned? First, the development of protein reference material, such as the NISTmAb, by NIST for use by the biopharmaceutical community represents an expansion in scope of attribute characterization relative to existing protein reference materials developed for other applications. The development of this three-volume book series is a testament to the fact that extensive, rigorous characterization of a single protein therapeutic is difficult, and indeed, the challenges in analytical characterization of protein therapeutics are well documented (*1*, *2*). The NISTmAb is arguably the most complex protein molecule in NIST's protein reference material portfolio. It is the largest by molecular weight and the first glycosylated protein reference material developed by NIST. The intent is to eventually certify the NISTmAb material for concentration, at which point it will become a NIST Standard Reference Material. The certified value for concentration will be metrologically traceable to the

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

International System of Units (SI) and will include an evaluation of measurement uncertainty (*3*). NIST has considerable expertise and experience in the process of developing reference materials that are certified for concentration. In working with scientists in the biopharmaceutical industry, however, it is clear that there would be considerable value for NIST to develop sound metrological strategies to establish identity for biomolecular species or attributes. For example, can a protein be certified for amino acid sequence or glycan composition, and can a level of confidence in the assignment be determined? There are increasing requests for NIST to develop such well-characterized biological materials for the biopharmaceutical industry and other biomedical sectors. NIST is addressing and making progress in this challenge. For example, a complete human genome reference material (RM 8398) was recently issued that was extensively sequenced by multiple techniques, with the results weighted and analyzed to eliminate as much variation and error as possible.

Second, there are many similarities between the development and life-cycle management of protein therapeutics by biopharmaceutical companies and protein reference materials by standards organizations like NIST. In development of the book series, NIST learned much from the biopharmaceutical industry about how to approach and manage these challenges. A reference material must be qualified in order to determine that it is fit for its intended purpose. For the NISTmAb, it is necessary to establish the identity, purity, homogeneity, and stability of the material, which is also a requirement in the development and manufacturing of protein therapeutics. Stability is of paramount importance in the development of a reference material like the NISTmAb. Although it is difficult to project demand for this material, we anticipate the NISTmAb will be a widely used product by the biopharmaceutical industry, instrument vendors, and other biomedical researchers. The benefit of using the NISTmAb is also that data generated from this material can be widely available and shared in the literature without facing a potential risk of disclosing proprietary information. In addition to characterization technology development, it may find eventual use in supporting intermediate product and final product testing for monoclonal antibody products. In such instances where the NISTmAb is specified in product licenses filed with regulatory agencies, a reliable, uninterrupted supply of the material will be essential. Because NIST has a considerable quantity of the NISTmAb, we believe it is likely that NIST will be able to satisfy most customer demands, provided that the material is stable under storage. As part of development of this book series, measurements conducted with the guidance and assistance of industry scientists who are experts in stability testing of protein therapeutics fortunately indicate that the NISTmAb is highly stable. An accelerated and real-time stability testing program is under development at NIST, and stability testing will continue for the lifetime of the NISTmAb.

Third, we discovered that a crowd-sourcing approach as employed in the development of this book series is a highly effective strategy for development of a protein reference material. Several benefits emerged from this approach, which represents a first for NIST in developing a protein reference material:

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

- Through collaboration, extensive and rigorous physicochemical characterization of the NISTmAb material was accomplished to an extent that would not have been possible by NIST alone. Moreover, NIST was able to leverage industry experts and capabilities not available at NIST.
- By working directly with the biopharmaceutical industry in this endeavor, NIST was able to obtain a better understanding of how customers might use the NISTmAb material, what structural attributes of the material were therefore important, and other potential protein reference materials to consider for future development.
- The simultaneous development of the book series and characterization of the NISTmAb fostered collaboration across the biopharmaceutical industry, instrument vendors, academia, and government agencies. This project demonstrated that the community of biopharmaceutical scientists is capable of coming together to discuss and move forward the field of analytical characterization in a way where everyone can benefit.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Critical to the success of the crowd-sourcing approach was the availability of a high purity, stable monoclonal antibody material of sufficient amount that it could be shared with multiple participants. This material not only served as a unifying focal point for many NIST measurement activities but also for the entire biopharmaceutical community. We will continue to use the crowd-sourcing approach when appropriate and in coordination with the industrial community and other stakeholders. It should be noted that several measurement intercomparisons exploiting the intact NISTmAb or fragments thereof are currently underway or being planned, including interlaboratory studies involving HDX-MS, NMR spectroscopy, and glycosylation analysis.

Looking ahead, we foresee the development of additional non-product-specific reference materials, such as the NISTmAb, that would be of potential benefit to the biopharmaceutical industry. There is a recognized need for additional globally acceptable reference materials to support the qualification and validation of physicochemical analytical methods (*4*). Such reference materials would assist in demonstrating that an analytical method has the requisite specificity, resolution, and precision to detect a meaningful difference in a quality attribute that might impact product safety and efficacy. In addition, these materials would support comparison of various analytical methods within and between different organizations and would be useful in assessing the performance of new analytical technologies. It should be noted that because NIST does not have the capability or resources to make biologic materials, progress in the development of some reference material types may hinge on the contribution of high-quality candidate materials from the biopharmaceutical industry.

These reference materials, such as the NISTmAb, are not meant to replace product-specific reference standards. The use of product reference standards is currently and will continue to be an essential practice in the development of protein therapeutics. Product reference standards support routine testing of product lots for quality control purposes, such as biological assays and physicochemical testing. As noted by Mire-Sluis (*5*), however, for some physicochemical analytical techniques, non-product-specific analytical method performance standards are

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

increasingly being used to support validation of method performance and being incorporated into method system suitability criteria. Below, different types of reference materials are considered that NIST could potentially develop in coordination with stakeholders of the biopharmaceutical community to support characterization of biologics, along with emerging trends that could influence their use and adaptation:

- **Additional uses of the NISTmAb material—**It may be feasible to transform physically or chemically the NISTmAb, thereby altering the amount or profile of a specific quality attribute. For example, the distribution of aggregates could be changed to enhance the population of oligomers through chemical cross-linking. If the identity and amount of oligomers could be determined and the preparation is stable, the preparation could be offered as a reference material for demonstrating method suitability and performance of size-based methodology for characterization of protein aggregates. Alternatively, if it is possible to isolate the aggregates and they are sufficiently stable, they could be used as spike-in materials for a similar purpose. Other potential uses of the NISTmAb include enzymatic treatment of the NISTmAb to alter its glycosylation profile to support qualification and validation of glycosylation analysis methods or enzymatic cleavage of the NISTmAb and separation of Fab or Fc fragments that could then be used as reference materials for ascertaining performance of HDX-MS or other analytical methods.
- **Other class monoclonal antibodies reference materials—**Although IgG1 monoclonal antibodies are currently the dominant class of antibodies, IgG2 and IgG4 antibody classes have been approved or are in development. In addition, it is common industry practice to concurrently develop IgG1, IgG2, and IgG4 candidates in order to examine potential differences in effector function. The analytical characterization and test methods developed for the IgG1, IgG2, and IgG4 classes can differ (e.g., determination of disulfide bond linkages), and the availability of IgG2 and IgG4 class reference materials would support method validation and development of system suitability criteria. Some members of the biopharmaceutical industry have requested that NIST consider developing IgG2 and IgG4 class reference materials, and NIST will continue discussions about the potential uses of these materials and explore potential sourcing options.
- **Antibody biosimilars—**Many blockbuster monoclonal antibody therapeutics are going off-patent, and numerous biosimilar versions of these products are in development. A foundational aspect of biosimilar development is physicochemical and biological comparison of the originator reference product with the biosimilar product. Nonetheless, there is a need to develop reference materials to support the development of antibody biosimilar products (*4*). Non-product-specific reference materials, such as the NISTmAb, could be used in the development of biosimilars by demonstrating system suitability of physicochemical

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

assays and ensuring that assays are reproducible, sensitive, and capable of detecting differences between products.

- **Next-generation antibody constructs—**Next-generation antibody therapeutics with novel structural motifs are an active area of research and development (*6*). Examples of these novel architectural constructs include antibody drug conjugates (ADCs), where a cytotoxic agent is covalently linked to an antibody; engineered antibodies with optimized Fc functionality, where approaches include glycosylation engineering or mutation of amino acids of the Fc region; and multispecific antibodies capable of targeting multiple different epitopes with a single recombinant molecule. Many of these molecules present new challenges in analytical characterization. For example, the drug-to-antibody ratio (DAR) is a critical attribute of ADCs that can impact stability and efficacy of the therapeutic. The highly heterogeneous nature of ADCs presents new analytical challenges in accurately and reliably determining the DAR. Members of the biopharmaceutical industry have indicated to NIST that a well-characterized ADC reference material to support method validation would be of value.
- **Precision antibody medicines—**Another potential driver for the increased use of non-product-specific standards is the trend toward development of personalized, or precision, biologic products. With the development of antibody products targeted for smaller patient populations, product batch sizes may potentially decrease from tens of kilograms to tens of grams. Therefore, the amount of material that is appropriate to serve as a primary and secondary in-house reference material will be limited, and manufacturers will likely increase their reliance on non-product-specific, analytical method performance standards. This use of non-product-specific reference materials will be more likely for structurally class-similar monoclonal antibody therapeutics that have many product quality attributes in common.

This book series, produced through a novel, extensive, public-private partnership, represents a first of its kind exposition of an important protein therapeutic class. Key to this effort was the use of a common, shared test material, the NISTmAb, through which multiple, independent laboratories were able to generate and compare representative physicochemical and bioassay characterization data. We hope that the NISTmAb will eventually be widely used throughout the biopharmaceutical industry and elsewhere. The need for additional antibody-based reference materials to support the development of new therapeutic monoclonal antibodies and antibody biosimilars will certainly grow. NIST looks forward to working closely with the biopharmaceutical industry, regulators, pharmacopeias, and other standards organizations to assess needs, coordinate efforts, and guide NIST reference material and other activities for maximum impact.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

References

1. Chirino, A. J.; Mire-Sluis, A. *Nat. Biotechnol.* **2004**, *22*, 1383–1391.
2. Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. *Nat. Rev. Drug Discovery* **2012**, *11*, 527540.
3. May, W.; Parris, R.; Beck II, C.; Fassett, J.; Greenberg, R.; Guenther, F.; Kramer, G.; Wise, S.; Gills, T.; Colbert, J.; Gettings, R.; MacDonald, B.; *NIST Special Publication*, 2000, pp 260−136. http://www.nist.gov/srm/publications.cfm (accessed October 2015).
4. Robinson, J.; Bristow, A. *CMC Strategy Forum*, July 15−16, 2013, Gaithersburg, MD.
5. Mire-Sluis, A.; Ritter, N.; Cherney, B.; Schmalzing, D.; Blumel, M. *Bioprocess Int.* **2014**, *12*, 12–31.
6. Anderson, J. B.; Syed, B. A. *Nat. Rev. Drug Discovery.* **2014**, *13*, 413–414.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ch015

Appendix

Table 1. Acronyms

Acronym	*Definition*	*Source*
1D	one-dimensional	HOS, Preface
2D	two-dimensional	HOS, Preface, Summary
3D	three-dimensional	HOS
ACN	acetonitrile	Covalent HOS
ACS	American Chemical Society	Summary
ADC	antibody drug conjugate	IM, Preface, Summary
ADCC	antibody-dependent cellular cytotoxicity	Summary
ADH	alcohol dehydrogenase	IM
AFM	atomic force microscopy	Aggregation, Summary
API	atmospheric pressure ionization	IM
API	active pharmaceutical ingredient	none of the chapters
ATD	arrival time distribution	IM
AUC	analytical ultracentrifugation	Aggregation
BIRD	blackbody infrared radiative dissociation	Preface
BLA	Biologics License Application	Aggregation, Summary
BME	β-mercaptoethanol	Preface
BS3	bis(sulfosuccin-imidyl)suberate	Aggregation
BSA	bovine serum albumin	IM
BSC	biosafety cabinet	Adventitious
CCD	charge-coupled device	Aggregation, Adventitious, SMSLS

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
CD	circular dichroism	Covalent HOS, HOS, Preface
CDC	complement-dependent cytotoxicity	Summary
cDNA	complementary DNA	Adventitious
CDR	complementarity-determining region	Aggregation, HOS
CE	capillary electrophoresis	
CEX-HPLC	cation exchange high-performance liquid chromatography	Aggregation, Preface Summary
CFU	colony-forming unit	Adventitious
CGS	centimeter, gram, second	SMSLS
CHO	Chinese hamster ovary	HOS, Adventitious
CID	collision-induced dissociation	Aggregation, Bioinformatics, HOS, Preface
cIEF	capillary isoelectric focusing	Summary
CIU	collision-induced unfolding	IM
CL	covalent labeling	Covalent HOS
Cp	crossing point	Adventitious
CPE	cytopathic effect	Adventitious
cryoprobe	cryogenically cooled probe	HOS
Ct	threshold cycle	Adventitious
CTD	charge transfer dissociation	Preface
dAb	domain antibody	Aggregation
DAR	drug-to-antibody ratio	Summary
DC	direct current	IM
DHSS	diffuse hard sphere scattering	IM
DLS	dynamic light scattering	SMSLS
dNTP	deoxynucleotide triphosphate	Adventitious
DR	dose response	Covalent HOS
DSC	differential scanning calorimetry	Aggregation, HOS, Preface, Summary
dSEC	denatured SEC	Summary

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
DTT	dithiothreitol	Covalent HOS, Preface
E. coli	*Escherichia coli*	HOS
ECD	electron capture dissociation	Bioinformatics, HOS, Preface
ECHOS-NMR	Easy Comparability of Higher Order Structure by NMR	HOS
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide	Covalent HOS
eFT	enhanced Fourier transformation	IM
EHSS	elastic/exact hard sphere scattering	IM
ELISA	enzyme-linked immunosorbent assay	Preface
EMR	extended mass range	IM
ER	endoplasmic recticulum	Aggregation
ESI	electrospray ionization	IM, Preface
ESI-MS	electrospray ionization mass spectrometry	Aggregation, Summary
ETD	electron transfer dissociation	Bioinformatics, HOS, Preface
ETS	Error Tolerance Search	Bioinformatics
Fab	antigen binding fragment	HOS, IM
Fc	crystallizable fragment	HOS, IM
FDA	Food and Drug Administration	Adventitious, Preface, Summary
FDR	false discovery rate	Bioinformatics
fSASA	fraction of the solvent-accessible side chain surface area	Covalent HOS
FT-ICR	Fourier transform ion cyclotron resonance	Bioinformatics, HOS, Preface
FTIR	Fourier transform infrared	Aggregation, HOS, Preface, Summary
Fv	variable fragment	HOS
FWHM	full width at half maximum	Preface

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
GCSF	granulocyte colony-stimulating factor	Covalent HOS
GDH	glutamate dehydrogenase	IM
GEE	glycine ethyl ester	Covalent HOS
GLP	good laboratory practice	Covalent HOS
H	heavy	Covalent HOS
HC	heavy chain	HOS
HCCF	harvested cell culture fluid	Adventitious
HCD	higher energy collision-induced dissociation	Preface
HCD	higher energy collisional dissociation	IM
HCID	higher-energy collision-induced dissociation	Bioinformatics
HCP	host cell protein	Bioinformatics, Preface, Summary
HDMS	high definition mass spectrometer	IM
HDX	hydrogen-deuterium exchange hydrogen-deuterium exchange mass	Covalent HOS, Preface
HDX-MS	spectrometry hydrophilic interaction liquid	HOS, Summary
HILIC	chromatography heteronuclear multiple quantum coherence	IM
HMQC	spectroscopy	HOS
HPLC	high-performance liquid chromatography	Aggregation
HPLC	high pressure liquid chromatography	Covalent HOS, Preface
HRF	hydroxyl radical-based footprinting	Covalent HOS

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
HSQC	heteronuclear single quantum correlation heteronuclear single quantum coherence	Aggregation
HSQC	spectroscopy	HOS
IAA	iodoacetic acid International Conference on Harmonization of Technical Requirements for Registration of	Covalent HOS
ICH	Pharmaceuticals for Human Use	Adventitious, Summary
IFN	Interferon Alpha-2	HOS
Ig	immunoglobulin	HOS
IgG	immunoglobulin G	Summary
IM	ion mobility	IM
IMS	ion mobility system	IM
IRMPD	infrared multiphoton dissociation	Preface
ISD	in-source decay	Preface
L	light	Covalent HOS
LAL	Limulus amoebocyte lysate	Adventitious
LC	liquid chromatography	Bioinformatics, IM, Preface
LC	light chain	HOS
LC-MS	liquid chromatography-mass spectrometry	Aggregation, Covalent HOS, Preface, Summary
LC-UV	liquid chromatography-UV LRR and Ig domain-containing, Nogo	Preface
LINGO	receptor-interacting protein time-dependent total intensity light	Aggregation
LS	scattering	SMSLS
LTQ	linear trap quadrupole	Bioinformatics
MAA	Marketing Authorization Application	Aggregation
mAb	monoclonal antibody	Aggregation, Bioinformatics, Covalent HOS, HOS,

Continued on next page.

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
MALDI	matrix-assisted laser desorption/ionization	Bioinformatics, IM, Preface
MALDI-TOF-MS	matrix assisted laser desorption/ionization-time of flight-mass spectrometry	Aggregation
MALS	multi-angle light scattering	SMSLS
MCO	metal-catalyzed oxidation	Aggregation
MD	molecular dynamics	IM
MKSA	meter, kilogram, second, ampere	SMSLS
MMV	minute virus of mice	Adventitious
MOA	mechanism of action	Summary
MS	mass spectrometry	Bioinformatics, Covalent HOS, IM, Summary
MS/MS	tandem mass spectrometry	Bioinformatics, Covalent HOS, Summary
Mw	molecular weight	IM
MWCO	molecular-weight cutoff	Covalent HOS, IM
NAT	Nucleic Acid Amplification Techniques	Adventitious
NC	negative control	Adventitious
NDF	neutral density filter	SMSLS
near-UV CD	Near-UV circular dichroism	Summary
NEC	negative extraction control	Adventitious
nESI	nano-electrospray ionistion	IM
NET	Normalized Elution Time	Bioinformatics
NHS	N-hydroxysuccinimide National Institute of Standards and	Aggregation
NIST	Technology	HOS, Preface, Summary
NISTmAb RM	NISTmAb reference material	Summary
NMR	nuclear magnetic resonance Nuclear Overhauser Enhancement	Covalent HOS, HOS, IM, Preface, Summary

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
NOESY	Spectroscopy	HOS
nrcSDS	non-reduced capillary sodium dodecyl sulfate electrophoresis	Summary
oaToF	orthogonal acceleration time-of-flight	IM
PA	projected area approximation	IM
PAGE	polyacrylamide gel electrophoresis	Adventitious
PC	positive control	Adventitious
PCR	polymerase chain reaction	Adventitious, Preface
PDB	protein data bank	Aggregation, Covalent HOS, HOS, IM
PF	protection factor	Covalent HOS
PFS	prefilled syringes	Aggregation
pI	isoelectric point	Aggregation
PK	pyruvate kinase	IM
PMF	peptide mass fingerprinting	Bioinformatics
PNGase F	Peptide-N-Glycosidase F	Covalent HOS
PQA	product quality attribute PROtein FIngerprint by Line shape	Bioinformatics
PROFILE	Enhancement	HOS
PSA	prostate-specific antigen	Bioinformatics
PTFE	polytetrafluoroethylene	SMSLS
PTM	post-translational modification	Bioinformatics, HOS, Preface, Summary
QbD	quality by design	Bioinformatics
q-PCR	quantitative PCR	Adventitious
QToF	quadrupole-time-of-flight	Bioinformatics, IM
R&D	research and development	Aggregation
RC	rate constant	Covalent HOS

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
rCGE	reduced capillary gel electrophoresis reduced capillary sodium dodecyl sulfate	Aggregation
rcSDS	electrophoresis reduced, denatured size exclusion	Summary
rdSEC	chromatography	Summary
RF	radio frequency human granulocyte macrophage-colony	IM
rhGM-CSF	stimulating factor	HOS
RM	reference material	Adventitious
ROA	Raman optical activity	HOS
ROC	receiver operating characteristic	Bioinformatics
RP	reversed-phase reversed phase-high-performance liquid	Bioinformatics
RP-HPLC	chromatography	Aggregation
RPM	revolutions per minute	SMSLS
S/N	signal to noise ratio	Bioinformatics, HOS
SANS	small-angle neutron scattering	Aggregation
SAP	spatial aggregation propensity	Aggregation
SAP	serum amyloid protein	IM
SASA	solvent-accessible side chain surface area	Covalent HOS
SAXS	small-angle X-ray scattering	Aggregation
scFv	single chain variable fragment	Aggregation
SDS	sodium dodecyl sulfate sodium dodecyl sulfate-polyacrylamide gel	Preface
SDS-PAGE	electrophoresis	Aggregation

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
SEC	size exclusion chromatography	Aggregation, Preface, Summary
SEC-MALLS	size exclusion chromatography-multi-angle laser light scattering	Aggregation
SEC-SLS	size exclusion chromatography with static light scattering detection	Summary
SI	International System of Units	HOS, SMSLS, Summary
SIC	selected ion chromatogram	Covalent HOS
SID	surface-induced dissociation	Aggregation, Bioinformatics, Preface
SLS	static light scattering	SMSLS
SMSLS	simultaneous multiple sample light scattering	SMSLS
SNase	staphylococcal nuclease	HOS
SV-AUC	sedimentation velocity analytical ultracentrifugation	Summary
TCEP	tris(2-carboxyethyl)phosphine	Covalent HOS, Preface
TEM	transmission electron microscopy	Summary
TM	trajectory method	IM
TOF	time of flight	Bioinformatics, Preface
TOP	take-off point	Adventitious
TROSY	Transverse Relaxation-Optimized Spectroscopy	HOS
TSE	transmissible spongiform encephalopathy	Adventitious
TWIM	travelling wave ion mobility	IM
UHV	ultra high vacuum	IM
UPLC	ultrahigh performance liquid chromatography	Bioinformatics, HOS, IM

Continued on next page.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Table 1. (Continued). Acronyms

Acronym	*Definition*	*Source*
USP	U.S. Pharmacopeial Convention	Adventitious
VCD	vibrational circular dichroism	HOS
XIC	extracted ion chromatogram	Bioinformatics, Summary

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ot002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Subject Index

B

Biologicals, adventitious agent testing, 227
  available testing strategies, 230
    adventitious agent detection, design of PCR methods, 234
    biopharmaceutical manufacturing process, illustration, 231*f*
    multi-plex real-time PCR sample detection method, examples, 236
    PCR, application, 239
    PCR implementation, contamination, 238
    PCR-based detection, cell-based detection, 238
    PCR-based viral detection, contamination experience, 233
    viral detection methods, 230*f*
    in vitro virus detection, 231
    in vitro virus detection, limitations, 232
  future perspectives
    one assay, many uses, 241
    sensitivity, suggestions to improve, 240

C

Characterization, monoclonal antibody, 113
  antibodies affecting aggregation propensity, physico-chemical properties
    electrostatic properties, 125
    Fv and Fc domain structure, 128
    glycosylation, 126
    spatial aggregation propensity (SAP) modelling, glycosylation, 127
    surface hydrophobicity, 126
  antibody aggregates, conventional approaches
    characterize aggregates, conventional assays, 131
    circular dichroism (CD), 135
    NISTmAb monomer, far-UV, 136*f*
    NISTmAb monomer, fluorescence intensity, 138*f*
    NISTmAb monomer, fourier transform infrared spectroscopy (FTIR) spectra, 137*f*
    process characterization, 129
    production, aggregate control, 130
    sedimentation coefficient, distribution, 134*f*
    size exclusion chromatography–multi-angle laser light scattering (SEC-MALLS) analysis, 133*f*
  antibody aggregation, common causes and mechanisms, 114
    chemical modifications, 117
    deamidation, isomerization, modifications, 119
    disulfide scrambling, modifications, 118
    fragmentation, modifications, 119
    interfaces, protein adsorption, 120
    leachables, 121
    native protein-protein self-association, 115
    oxidation, modifications, 118
    possible antibody aggregation pathways, illustration, 115*f*
    protein unfolding and misfolding, 116
  conclusions, 149
  emerging technologies
    atomic force microscopy (AFM), 141
    high resolution technologies, 145
    imaging technologies, 139
    limited proteolysis and cross-linking, 144
    macroscopic technologies, 141
    negative stain transmission electron microscopy (TEM), 140*f*
    nuclear magnetic resonance (NMR) spectroscopy, 146
    ribbon structure, IgG1 Fc, 147*f*
    in silico aggregation prediction, 148
    spatial aggregation propensity score, 149*f*
    X-ray and neutron scattering, 142
  kinetics and formulation section, 122

G

Global partnership advancing biopharmaceutical development, 415
  NIST perspective, 425

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

additional non-productspecific reference materials, development, 428
antibody biosimilars, 429
crowd-sourcing, 427
next-generation antibody constructs, 430
NIST Biomanufacturing Program, 426
NISTmAb material, additional uses, 429
other class monoclonal antibodies reference materials, 429
precision antibody medicines, 430
product reference standards, 428
protein therapeutics, life-cycle management, 427
two-dimensional (2D) NMR methods, 425
reference material and characterization technologies
heterogeneity assessment, 422
higher order structure, 423
host cell proteins (HCPs), 424
longitudinally available material, challenges, 425
summation and perspective, 416
analytical and biophysical characterization methods, 420
covalent labeling mass spectrometry (MS), 420
example characterization method summary, 417*t*
higher order structural methodology, principles, 421
monoclonal antibody characterization, analytical technology, 420

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

H

Host cell proteins (HCPs), identification and quantification, 357
bioprocess development, application of 2D-LC-MS^E methods, 374
case I, identifying HCPs
analysis, sample amount, 378
2-dimensional chromatography, fluidic configuration, 377*f*
experimental conditions, 376
generic HCP ELISA assay, 383
HCP identification, data processing, 378
host cell proteins (HCPs), 382*t*
low-abundance peptide (LLEELEEGQK), identification, 381*f*
MS system and MS settings, 377
NISTmAb, total ion chromatograms, 380*f*
results, case I, 379
sample - NISTmAb, 375
sample preparation, 375
Case II, HCP removal during purification, 384
Case III, process changes, effects, 386
HCP analysis, common methodologies, 359
enzyme-linked immunosorbent assays (ELISA), 360
sodium-dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), 362
western blotting method, 364
HCPs, LC-MS analysis, 366
data collection mode, 372
HCP proteome, sample loading capability, 370
HCP proteomes LC separations and analysis, performance, 367
summary and outlook, 388
Humanized IgGk NIST monoclonal antibody, ion mobility and mass spectrometry measurements, 75
ion mobility, 78
drift tube instrumentation, 80
RF-confining drift cell instrumentation, 83
second-generation travelling wave ion mobility mass spectrometer, schematic, 82*f*
travelling wave instrumentation, 81
waters synapt G1 HDMS instrument, schematic, 84*f*
results
collision-induced unfolding, 97
experimental and theoretical Ω_{N2}, comparison, 96*f*
IM analysis, 91
MD simulations, 93
native MS analysis, 88
NISTmAb, collision-induced unfolding, 98*f*
NISTmAb mass spectrometry (MS)-spectra, comparison, 89*f*
NISTmAb theoretical structure, temporal evolution, 94*f*
RF-confining drift cell ion mobility (IM) plot, 91*f*
sample cone voltages, NISTmAb, 90*f*

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

theoretical Ω_{N2} and Ω_{He} value calculations, 95*t*
travelling wave ion mobility (TWIM), 92*t*
sample preparation
extended mass range (EMR), schematic, 87*f*
MD simulations, IMoS mobility software, 88
orbitrap exactive plus EMR sample preparation, 85
orbitrap exactive plus-EMR voltages and pressures, 87
RF-confining drift cell instrument voltages and pressures, 86
synapt G2 HDMS instrument voltages, 85

I

Informatics, mass spectrometry-based protein characterization, 189
concluding remarks, 218
bioinformatics, annual number of publications, 219*f*
MS-based proteomics, informatics
accurate mass and time (AMT) tags, 194
database searching, evaluation of algorithms, 204
database searching programs, examples, 199*t*
descriptive scoring algorithms, 199
interpretative scoring algorithms, 200
Mascot, scoring models, 202
MassMatrix, scoring models, 203
MS-based proteomics, 192
partial three-amino-acid sequence tag VAL, simplified representation, 200*f*
peptide fragmentation, 195
peptide fragmentation, example product ions, 196*f*
peptide mass fingerprinting workflow, schematic diagram, 193*f*
proteome and proteomics, 191
protonated peptide ions, nomenclature for fragmentation, 196*f*
representative scoring models, 201
ROC curves, representation, 205*f*
schematic diagram, 198*f*
SEQUEST, scoring models, 201
statistical/probabilistic scoring algorithms, 200
stochastic scoring algorithms, 200
tandem LC-MS/MS experiment, schematic diagram, 195*f*
tandem mass spectrometric, 197
tandem mass spectrometric data analysis, automated database searching, 198
tandem MS, 194
peptide mapping, informatics
background ions, 209
HCP analysis, FDR, 218
identified and unidentified ions, peak area distribution, 210*f*
modified peptides, 208
MS-based proteomics and peptide mapping, difference, 205
peptide mapping data, challenges in analysis, 207
peptide mapping software, 214
peptide mapping software, BiopharmaLynx, 216
peptide mapping software, Byonic, 215
peptide mapping software, Mascot Error Tolerant Search, 215
peptide mapping software, MassAnalyzer, 215
peptide mapping software, MassHunter BioConfirm, 216
peptide mapping software, proteomic search engines, 214
processing peptide mapping data, differential profiling, 213
processing peptide mapping data, identification, 211
processing peptide mapping data, quantitation, 212
processing peptide mapping data, software, 210
product impurities, 209
search results, confidence, 216
sequence variant analysis, FDR, 217
sequence variants, 209
tasks, peptide mapping with LC-MS/MS, 206
unmodified peptides, 208

L

LC-MS analyses, automated and online sample preparation
affinity capture, 350
NIST monoclonal antibody, affinity capture, 352*f*

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

other workflows and applications, 352
digestion
biology and chemistry, 337
digestion automation, 340
digestion products, acquisition, 347
LCMS-8050 triple quadrupole mass spectrometer, 344
NISTmAb by high mass accuracy LC-MS/MS, peptide mapping, 349*f*
online automation, improvements, 343
overnight manual digestion and Workstation digestion, comparison, 345*f*
peptide mapping, 347
peptides, quantitative analysis, 347*f*
Perfinity Workstation – The Perfinity Workstation, configuration and valve diagram, 341*f*
quantification and linearity, 345
quantitative measurements, 346
trypsin immobilized enzyme reactor (IMER), 342
future directions, 353
summary and conclusion, 354

M

MALDI in-source decay MS
introduction, 317
materials and methods
data collection, 320
sample preparation, 319
results and discussion, 320
antibody heavy and light chains, sequence coverage, 328*f*
Fab heavy chain, sequence assignment, 330*f*
heavy chain, sequence modifications, 325
IdeS protease, MALDI-ISD spectrum, 329*f*
intact subunit and intact fragment mass analysis, 321
light chain and heavy chain, comparison of sequence assignments, 327*f*
linear mode of mAb, MALDI-TOF spectra, 322*f*
mAb, heavy chain MALDI-ISD spectrum, 326*f*
mAb, MALDI-ISD spectrum, 323*f*
MALDI-ISD sequence coverage, 324*t*
middle-down heavy chain domain sequencing, 328
subunit analysis workflow, 321*f*
subunit sequencing, 323
T_3 spectra, light chain of antibody, 331*f*
T_3 spectra, sequence confirmation, 330
Monoclonal antibodies, covalent labeling techniques for higher order structure, 45
challenges
preserving structural integrity, 68
quality control and software solutions, 69
conclusions, 70
future applications
epitope/paratope mapping, 66
protein quality measures and biosimilars, 67
future directions
alternative enzymes, higher resolution, 69
intact and top-down MS, 70
methods
data analysis, 51
deglycosylation and digestion, 50
mass spectrometry, 50
NISTmAb, structural model, 51
protein labeling, 49
results
carboxyl group footprinting experiments, results, 60*t*
carboxyl group labeling, results, 59
high oxidation rates, 58
homology model, 63
HRF and carboxyl group labeling results, comparison, 62
HRF and GEE, comparison, 63*t*
HRF experiment, dose-response plots, 54*f*
hydroxyl radical footprinting, 52
hydroxyl radical footprinting experiments, results, 54*t*
labeling and deglycosylation, 52
low oxidation rates, 58
NIST monoclonal antibody (mAb), homology model, 64*f*
outliers, 58
protection factor (PF) mapping, 65*f*
structural prediction, 65
tryptic peptides, summary, 53
Monoclonal antibodies, emerging technologies for higher order structure, 17
HDX-MS, 32

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

electron-transfer dissociation (ETD), 34
hydrogen-deuterium exchange (HDX) workflow, 33*f*
immunoglobulin, IgG schematic antibody, 19*f*
NISTmAb, ribbon diagram structures, 20*f*
mAb crystal structures, 20
crystallizable fragment (Fc), IgG antibody crystal structures, 21*f*
mAbs, HDX-MS footprinting, 35
deuterium uptake mapped, 39*f*
HDX-MS peptide map, 36
hydrogen-deuterium exchange (HDX) heat map, 37*f*
NISTmAb light chain, hydrogen-deuterium exchange (HDX) heat map, 38*f*
mAbs, NMR structural fingerprinting, 27
NISTmAb, NMR structural fingerprinting, 28
0.30 mM intact NISTmAb, two-dimensional (2D) spectral fingerprint, 31*f*
HSQC NMR spectral fingerprinting methodology, 28
one-dimensional (1D) water flip-back Watergate water suppression, overlay, 30*f*
unlabeled NISTmAb domains, two-dimensional (2D) spectral fingerprint, 31*f*
protein biologics, NMR structural fingerprinting, 22
2D NMR fingerprinting methodology, 26
fingerprinting protein biologic structure, HSQC spectroscopy, 25
NMR, structural characterization, 24
one-dimensional (1D) proton, 23
spectroscopic methods, 22
summary, 39
Monoclonal antibody analysis
mAbs, structural feature, 246
microfluidic technologies, results
2100 bioanalyzer, mAb sizing and QC, 279
2100 Bioanalyzer analysis, gel-like image, 280*f*
analysis of the NISTmAb RM, deconvoluted mass spectra, 275*f*
capillary isoelectric focusing, charge variant analysis, 281
CID MS/MS spectrum, example, 270*f*
deglycosylation, analysis of IdeS-digested NISTmAb RM, 277*f*
glycopeptide analysis, 268
glycoproteins digests, schematic, 268*f*
IdeS-generated mAb fragments, MS analysis, 276
intact and deglycosylated mAb, MS analysis, 274
intact NISTmAb RM, 2100 Bioanalyzer electropherogram, 281*f*
mab-glyco chip, N-glycan analysis, 270
mAb-Glyco Chip, schematic, 272*f*
mAbs, preparation using IdeS, 278
N-glycan quantification, histogram, 274*f*
N-glycans, automated preparation and analysis, 271*f*
N-glycans detected, chromatogram, 273*f*
N-glycopeptide quantification, histogram, 269*f*
NISTmAb RM, cIEF analysis, 282*f*
NISTmAb RM, ion chromatogram, 261*f*
NISTmAb RM, peptide map, 262*f*
NISTmAb RM tryptic digest, peptides detected, 263*t*
observed and calculated mass values, comparison, 276*t*
polaris-HR HPLC-chip, 260
reduced NISTmAb RM, 2100 Bioanalyzer electropherogram, 280*f*
monoclonal antibodies, microfluidics and analysis, 248
bioanalyzer chip, 254*f*
HPLC-chip for peptide separations, schematic, 251*f*
mAbs, Capillary isoelectric focusing (cIEF), 254
mAbs, microfluidic electrophoresis, 253
microfluidics-based LC/MS, application, 249
microfluidics-based surface plasmon resonance, mAb analysis, 249
porous graphitized carbon (PGC) enrichment and separation, 252
separation and quantification of ADCs, cIEF, 255
standard HPLC-chip microfluidic device, photograph, 250*f*

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

NISTmAb reference material, methods of analysis
bioanalyzer analysis, 259
capillary isoelectric focusing, 259
glycan nomenclature, 257
glycopeptide analysis, 257
LC/MS analysis, 256
peptide analysis, 256
PNGase F treatment, mAb fragments, 258
PNGase F treatment, MS analysis of intact mAb, 258
released glycan analysis, 257
tryptic digestion, 256

N

Next-generation biotherapeutic characterization tools, 1
biotherapeutics characterization, development of emerging technologies, 5
Q2, responses, 6*f*
development of emerging technologies, mass spectrometry instrument performance, 11
mass spectrometers, ability, 14*f*
mass spectrometers, resolution, 12
proteins by mass spectrometry, analysis, 13
Q6, responses, 11*f*
emerging technologies, additional development, 4
Q1, responses, 4*f*
process-related testing, application of mass spectrometry, 15
Q7, responses, 15*f*
product characterization, determination of higher order structure, 8
liquid chromatography, 9
Q4a and Q4b, responses, 8*f*
product characterization, methods and utility in respect to mass spectrometry, 10
Q5, responses, 10*f*
protein modifications identification, additional technological development required, 7
Q3, responses, 7*f*
summary, 15
NIST reference mAb, bioinformatic analysis
deamidation and ammonia loss, 403
deamidated and ammonia loss peptides, table, 404*f*
general peptide mapping, 410
NIST reference mAb, UV peptide map, 411*f*
peptide FNWYVDGVEVHNAK, candidate assignment, 412*f*
peptide map, zoomed in region, 413*f*
glycation, 404
XICs, comparison, 405*f*
intact glycosylated peptides, 405
ETD MS2 spectrum, 406*f*
introduction, 395
bottom-up data, NIST reference mAb, 396
Byologic, interactive dashboard, 399*f*
Byologic bioinformatics software, Protein Metrics, 398
therapeutic protein characterization, 397
oxidation, 403
oxidized peptides, table, 403*f*
sequence variant analysis, 400
MS2 annotation, side-by-side comparison, 402*f*
SVA peptide spectrum matches, table, 402*f*
top-down/middle-down mass spectrometry, 407
annotated high-resolution/high accuracy ETD MS2 spectrum, 408*f*
NIST reference mAb, fragmentation maps, 409*f*

O

Orbitrap mass spectrometry, 289
harvest cell culture screening, therapeutic mAb analysis, 308
automated 2D-LC system, fluidic configuration, 310*f*
mAb, characterization, 311*f*
NISTmAb, middle-down LC-MS/MS, 309*f*
intact antibody mass measurement, 292
antibody and related products, native MS, 296
antibody-antigen complex, native MS analysis, 298
average molecular mass, measurement, 294
mAb-Ag complex, native MS analysis, 299*f*

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.

monoisotopic molecular mass, measurement, 295
NISTmAb, molecular mass measurement sing Q exactive Orbitrap LC-MS, 295*f*
NISTmAb light chain and heavy chain, monoisotopic mass measurement, 297*f*
intact subunit and large fragment of mAb, sequencing, 300
c and z ions, number, 307*t*
eleven N-deglycosylated humanized antibodies, MS analysis, 301*f*
eleven N-deglycosylated humanized antibodies, theoretical masses, 302*t*
intact mAb subunit, sequencing, 303
mAb, middle-down sequencing, 307
NISTmAb intact light and heavy chains, LC-MS/MS, 305*f*
NISTmAb intact light and heavy chains, protein backbone fragmentation and sequence coverage, 306*f*

P

Protein aggregation, simultaneous multiple sample light scattering (SMSLS), 159
background, 160
equilibrium characterization, 165
NISTmAb, Rayleigh scattering ratio, 168*f*
three proteins, equilibrium properties, 169*t*
materials, 164
protein aggregation, kinetics
aggregation rate (AR), NISTmAb, 180*f*
air/liquid interface, effect, 181*f*
Arrhenius extrapolations, 177
Arrhenius plots, 175
different time-dependent aggregation signatures, 172*f*
equivalent weight average molecular weight ratio, 174*f*
high protein concentration, determining AR, 173
latex spheres, 184*f*
mAb, particulate population, 185*f*
monoclonal antibody, dimerization, 179*f*
monoclonal antibody C (mAbC), aggregation, 182*f*
NISTmAb, aggregation, 176*f*
NISTmAb, Arrhenius plot, 178*f*
NISTmAb, cloudiness, 181*f*
protein aggregation, particulates formation, 183
protein aggregation, stirring effects, 179
SMSLS, proteins studied, 178*t*
temperature, aggregation, 175
time-dependent signatures, 170
varying formulation conditions, 182
SMSLS instrumentation
16-cell simultaneous multiple sample light scattering (SMSLS) system, 163*f*
operation, modes, 164
sensitivity and minimum volume, 162
SMSLS hardware, 161
SMSLS software, 163

Publication Date (Web): October 15, 2015 | doi: 10.1021/bk-2015-1202.ix002

Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biopl
ACS Symposium Series; American Chemical Society: Washington, DC, 2015.