

**State-of-the-Art and Emerging  
Technologies for Therapeutic  
Monoclonal Antibody Characterization  
Volume 2. Biopharmaceutical Characterization:  
The NISTmAb Case Study**



ACS SYMPOSIUM SERIES **1201**

**State-of-the-Art and Emerging  
Technologies for Therapeutic  
Monoclonal Antibody Characterization**  
**Volume 2. Biopharmaceutical Characterization:  
The NISTmAb Case Study**

**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*



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 ; 1201)

Includes bibliographical references and index.

Contents: v. 2. biopharmaceutical characterization: the NISTmAb case study

ISBN 978-0-8412-3029-3 (v.2)

I. 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.

Copyright © 2015 American Chemical Society

Distributed in print by Oxford University Press

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of \$40.25 plus \$0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

# 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**

# Preface

The focus of this volume is the characterization of monoclonal antibodies (mAbs). A significant body of work has been written both within this book series (see Volume 1) and elsewhere on the importance of the subject; however, several unique aspects of the current series are highlighted in this volume. The book focuses both on general aspects, techniques, and regulatory concerns common to any recombinant protein, as well as specific analytical results, in that it deals primarily with one mAb, the NISTmAb. Volume 2 therefore serves as both a foundational body of NISTmAb product knowledge as well as an evaluation of its suitability as an industry-appropriate reference material (RM). Volume 2 contains representative methods and associated data for the NISTmAb. The extent and quality of the data collected is comparable to that in a Biologics License Application (BLA). The number and quality of the researchers who have provided data on the NIST RM and perspective on the development of mAbs in general is unparalleled. The material is already being used throughout industry and is driving the adoption of new techniques. The need for the book and material go hand in hand, and therefore the NISTmAb will be made widely available to the community as NIST Reference Material 8671 shortly after publication of this series. The combination of an industry-driven book series and a material available for open innovation of analytical technologies represents a significant stride forward in the attempt to increase standardization and the use of standards within the biopharmaceutical industry.

In Volume 1, a framework was presented, detailing what potential critical quality attributes (PCQAs) should be measured when developing mAbs; offering a risk-based approach to determining and quantifying PCQAs as “nice to have” versus when they are “must have”; specifying to what level PCQAs must be measured; and ultimately, explaining why it all matters. In Volume 2, we attempt to use that framework as a guide to deduce key biochemical and biophysical parameters of the NIST RM. The ultimate goal of Volume 2 is helping researchers have the clearest picture possible of the NIST RM specifically and mAb characterization in general; describing what assays have been helpful in determining and quantifying which biochemical and biophysical attributes; and hopefully, highlighting where this knowledge can help researchers in their daily practice. Using the data found in Volume 2, an attempt is made to align the key attributes or PCQAs with the outcome of the analytical testing regime. In Volume 3, we address gaps identified either in Volume 1 or 2, utilizing emerging technologies that are just now finding their way into laboratories and gaining general acceptance.

A typical laboratory within a biopharmaceutical organization is a mixed bag of instrumentation that reflects the toolbox approach used by most companies to address wide-ranging questions. Each question generally has several levels of answers that depend on the stage or phase that the potential therapeutic has reached. In an attempt to provide a structure that is generally nonexistent experimentally, the nine chapters that comprise Volume 2 are organized in a logical flow going from primary structure (the paramount attribute) to modifications of the primary structure (post-translational modifications [PTMs]) and glycosylation, a specific type of PTM. Multifarious separation techniques, which are designed to provide orthogonal information of the purity, and a spectrum of biophysical techniques, which are used to probe higher order structure, are introduced next. Finally, technologies for detection and characterization of particulates and process-related impurities are discussed. Several of the techniques, such as peptide mapping or size exclusion chromatography (SEC), may be used throughout several of the chapters, but the focus of the techniques change to fit the question.

Chapter 1 of Volume 2 deals with primary structure determination. It uses several techniques common in the field, including intact protein analysis; top-down sequencing; subunit and IDeS (partial digestion) followed by middle-down sequencing and peptide mapping using various enzymes; and targeting to obtain reliable information on the entire protein, its constituent domains, and peptides as building blocks comprising the primary structure. The totality of evidence from various data sets provides a confidence in the primary structure of the NIST RM. In addition, presented data suggests that the NIST RM is low in variants such as fragments and other PCQAs (Table 1, Heterogeneity chapter/Volume 1, Chapter 3). The assays used cover those outlined for primary structure in Table 1 of the Well Characterized chapter/Volume 1, Chapter 4, as well as others. The primary structure and any even minor variants thereof must be ascertained to the highest degree possible. If any potential modifications exist that could impact the efficacy (antigen binding and other mechanisms of action) or safety (potential immunogenicity or aberrant clearance), they must be characterized and quantified to a level appropriate for the phase the product is in.

Chapters 2, 3, and 4 of Volume 2 build upon the primary structure by focusing on sequence variants of the primary structure, PTMs of the primary structure, and glycosylation (a key PTM), respectively. Chapter 3 was an interlaboratory study directed at identifying, characterizing, and quantifying PTMs ranging from deamidation to oxidation (Table 1). The chapter highlights that while most laboratories may identify a peptide with a PTM correctly, localization and quantification of the modification calls for additional criteria and should be done with caution. Chapter 2 deals with the identification and quantification of sequence variants (amino acid substitutions, omissions, or insertions not encoded by the DNA vector). It is important to note that whereas Chapter 1 calls out a sequence that is predicted from the DNA, Chapter 2 highlights that there are alternate sequences based upon modifications to the DNA or due to protein synthesis by living organisms. Sequence variant analysis (SVA) is a complex analysis that typically deals with very low-level species, but it is critical in order to establish a robust cell line and ensure its stability suitable for production of

large quantities of a mAb. Chapter 4 is an in-depth and exhaustive analysis of glycosylation, which is a key attribute of mAbs due to its importance for stability, receptor binding, and clearance (Mechanisms of Action chapter/Volume 1, Chapter 2).

Chapter 5 of Volume 2 covers separations and orthogonal techniques. The assays used in this chapter (cation exchange-high-performance liquid chromatography [CEX-HPLC], SEC, reversed-phase HPLC [RP-HPLC], hydrophobic interaction chromatography [HIC], reduced and non-reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis [SDS-PAGE], reduced and non-reduced capillary electrophoresis with SDS [CE-SDS], microchip electrophoresis with SDS [MCE-SDS], capillary zone electrophoresis [CZE] and capillary isoelectric focusing [cIEF]) cover many of the stability-indicating assays used to demonstrate process and product consistency (Table 1, Heterogeneity chapter/Volume 1, Chapter 3; Table 1, Well Characterized chapter/Volume 1, Chapter 4).

Chapters 6, 7, 8, and 9 of Volume 2 transition from the characterization and stability-indicating assay focus of the earlier chapters into higher order structure, aggregation, and process-related impurities. Biophysical techniques used in determining higher order structure is the subject matter of Chapter 6. Chapter 7 covers developability assessment of therapeutic proteins using a variety of orthogonal technologies. Chapter 8 focusses on protein particulates ranging from 0.1  $\mu\text{m}$  to 100  $\mu\text{m}$ , categorized as sub-visible/visible particles. Chapter 9 is directed toward non-product-related quality attributes: process impurities (host cell proteins, residual protein A from purification, and DNA).

Although each chapter can and does stand on its own, it is clear that the use of the NIST RM as a standard in biochemical and biophysical assays is enhanced by looking at the data holistically. In this sense, one can look for supporting data between chapters or where the data from several assays are not consistent. For instance, charge-based assays were used in Volume 2, Chapter 5, to quantify the total acidic peak variants as 14.4% and 24.1% using CEX and cIEF, respectively. A variety of acidic species were identified (and in some cases quantified) in other chapters, and potential acidic species contributing to the sum include sialylation (1.8% from Chapter 4, Glycosylation) as well as low-level deamidation and glycation (Chapter 1, Primary Structure, and Chapter 3, PTMs). It is clear that further explanation can be gleaned from considering the data in its totality, and additional experiments may then be warranted based on the resultant comparison. It also is evident that the overlap in assays between certain chapters provides an opportunity to look at correlations between assays even when no attempt is made at aligning methodologies. For example, the sialylation reported between chapters is consistently within 1 to 2%, all chapters report low-level fragments, and so forth.

Taken together, the totality of the data demonstrates that the NIST RM is a stable monoclonal IgG1 molecule that can be used as a standard and system suitability control in a wide variety of characterization assays. In addition, the NIST RM also has enough heterogeneity to be useful in demonstrating non-product-specific aspects of stability-indicating assays. We look forward to the feedback from the biopharmaceutical and regulatory community as to their



intention on using the material in their assays and the potential need to extend the IgG1 standard into other isotypes.

**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, Pennsylvania 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)

# 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 IgG1 $\kappa$  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.

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.

## Chapter 1

# Determination of the NISTmAb Primary Structure

**Trina Formolo,<sup>1</sup> Mellisa Ly,<sup>2</sup> Michaella Levy,<sup>4</sup> Lisa Kilpatrick,<sup>1</sup> Scott Lute,<sup>5</sup> Karen Phinney,<sup>1</sup> Lisa Marzilli,<sup>2</sup> Kurt Brorson,<sup>5</sup> Michael Boyne,<sup>4</sup> Darryl Davis,<sup>3</sup> and John Schiel\*,<sup>1</sup>**

<sup>1</sup>Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

<sup>2</sup>Mass Spectrometry and Biophysical Characterization, Analytical Research and Development, BioTherapeutics Pharmaceutical Sciences, Pfizer, Inc., Andover, Massachusetts 01810, United States

<sup>3</sup>Janssen Research and Development, LLC, Spring House, Pennsylvania 19002, United States

<sup>4</sup>Center for Drug Evaluation and Research, Office of Testing and Research, Division of Pharmaceutical Analysis, U.S. Food and Drug Administration, Saint Louis, Missouri 63110, United States

<sup>5</sup>Center for Drug Evaluation and Research, Office of Biotechnology Products, Division of Monoclonal Antibodies, U.S. Food and Drug Administration, Silver Spring, Maryland 20993, United States

\*E-mail: [john.schiel@nist.gov](mailto:john.schiel@nist.gov)

The primary structure of a protein, including therapeutic monoclonal antibodies (mAbs), is a critical quality attribute that determines a great deal of its structure, function, and stability. Significant effort is devoted to determining the complete amino acid sequence of recombinant proteins of potential therapeutic benefit. The focus of this chapter is to demonstrate current, state-of-the-art, mass spectrometry-based primary structure confirmation using the recombinant human IgG1κ NISTmAb as a representative example. A combination of intact mass analysis, top-down sequencing, IdeS fragment mass analysis,

IdeS fragment (or “middle-down”) sequencing, and peptide mapping are discussed with respect to their orthogonality and limitations. A historical review of the methods employed to determine the primary structure of proteins is also included for completeness.

## Introduction

A protein’s primary structure consists of the linear sequence of its constituent amino acid residues. Although the primary structure of a protein can be predicted from the encoding gene sequence, the possibility of transcriptional, co-translational, and post-translational events may alter the expressed protein, including its amino acid sequence. This heterogeneity has been encompassed by terms such as “protein isoform,” “protein variant,” and more recently, “proteoform” has been proposed to describe the variability observed at the expressed protein level (1). It is important to characterize the inherent heterogeneity of a therapeutic product’s primary structure. This exercise of primary structure confirmation has evolved through the years to include a variety of methodologies, each of which in some way exploits the unique properties of the peptide bond.

Proteins are a linear sequence of amino acid residues formed through a condensation reaction that produces an alpha carbon chiral center. The amide linkage (peptide bond) formed between the carboxyl and amino group of two amino acids is the key aspect of polypeptide synthesis and stability. This peptide stability is imparted by a resonant structure yielding partial double bond character.

The key experiment for determining the primary structure of a protein was first performed by Frederick Sanger in 1951. Sanger used chemical derivatization to selectively label the N-terminus of bovine insulin with the yellow dye fluorodinitrobenzene, followed by hydrolysis, and electrophoretic or chromatographic separation of the labeled N-terminal amino acid residue. Through the use of multiple rounds of partial protein hydrolysis, fractionation, and terminal amino acid determination, the entire sequence of the B chain of insulin was determined (2). At the same time, the related technique of Edman degradation was also being developed. In this technique, the Edman reagent (phenylisothiocyanate) is used to react with the N-terminal amino acid of a protein under weakly basic conditions. Addition of trifluoroacetic acid (TFA) causes an intramolecular rearrangement that releases the labelled N-terminal amino acid, which can then be identified by chromatographic methods (3). Important to this technique is that partial hydrolysis of the protein is not required, such that sequential cycles can be used to sequence the protein for up to 50 residues from the N-terminus with automated sequencers (4) provided that the N-terminus was not modified.

Early methods for complete protein amino acid sequencing combined enzymatic protein cleavage followed by wet chemistry techniques such as Edman sequencing and variations thereof. These methods include treatment with an endopeptidase, a protease that cleaves between non-terminal amino acids of

polypeptide chains, to first digest a protein into peptides. Early on, peptides were resolved from these complex mixtures prior to sequencing using electrophoretic techniques (5, 6). However, as high-performance liquid chromatography (HPLC) methods were developed for the separation of peptides, they became the standard means of fractionation applied to peptides intended for amino acid analysis or Edman sequencing (7–11). Due to their smaller size, the complete amino acid sequence of peptides can be determined using terminal sequencing techniques such as Edman sequencing or related variations and later pieced together to determine the complete primary structure of the protein. However, obtaining the complete sequence of a protein using such methods is time consuming and thus not feasible for the routine sequencing of larger proteins such as monoclonal antibodies (mAbs).

Fortunately, advances in HPLC and more recently ultrahigh-pressure liquid chromatography (UHPLC), mass spectrometry (MS), and their combination, liquid chromatography-mass spectrometry (LC-MS), have provided a high-throughput means of sequencing peptides generated by endopeptidase digestion. Chromatographically separated peptides may be fragmented in the mass spectrometer (tandem mass spectrometry, MS/MS or MS<sup>2</sup>) to provide information on the linear sequence of the amino acids forming the peptide. Although primary structure determination has largely moved to the more high-throughput LC-MS- and liquid chromatography-tandem mass spectrometry (LC-MS/MS) based methods, Edman sequencing can still play a critical role as a confirmatory technique, particularly in the case of a peptide containing one of the isobaric residues, Leu or Ile.

LC-MS and LC-MS/MS are used in the field of proteomics to identify a large number of proteins within a complex mixture. In a technique commonly referred to as “bottom-up” proteomics analysis (12, 13), the detection of a limited number of peptides matching the known sequence of an individual protein is considered sufficient to confidently identify that protein within the mix. Similar instrumentation and analytical methods are used for bottom-up approaches in the biopharmaceutical industry; in this setting, however, LC-MS peptide mapping techniques are used with the goal of obtaining full amino acid sequence coverage and establishing a unique chromatographic peptide trace for a particular purified protein. This method provides confirmation of the primary structure as well as identification of post-translational modifications (PTMs). Common proteolytic enzymes such as trypsin and lysyl endopeptidase (Lys-C) are typically employed for peptide mapping in the biopharmaceutical industry. The use of peptide mapping as a tool for protein sequencing is discussed later in this chapter.

In addition to peptide mapping, additional MS-based techniques that leverage instrument improvements in resolving power, mass accuracy, and sensitivity have been implemented in the biopharmaceutical setting. Larger proteins, as well as individual subunits, antibody fragments, and domains, can be analyzed in much greater detail with state-of-the-art MS. For example, the average mass of an intact mAb can be measured by high-resolution mass spectrometers, resulting in a mass accuracy below 50 ppm (14–16). Isotopic resolution can be achieved after chemical and/or enzymatic treatments to cleave a mAb into constituent components of smaller size and complexity. For example, a highly specific

protease (IdeS) can be used to cleave the protein in the hinge region to produce antibody fragments for accurate mass analysis (17–20). Alternatively, reduction of a mAb into its constituent heavy (H) and light (L) chain subunits can be performed to selectively monitor their accurate mass (21).

In addition to a direct mass analysis, MS/MS fragmentation of IgG fragments and/or subunits can also provide sequence information in what has been termed “middle-down” sequence analysis (19). Current sequencing efforts involve fragmentation techniques such as collision-induced dissociation (CID), electron transfer dissociation (ETD), higher energy collision dissociation (HCD), and infrared multiphoton dissociation (IRMPD) (22–26). These techniques require little to no sample preparation such as enzymatic digestion and can be used as orthogonal approaches to bottom-up methods. Intact antibodies may also be subjected to MS/MS fragmentation for top-down sequencing (27, 28). Due to their size, however, the large number and variety of fragment ions produce complex spectra that can be difficult to interpret (27). The pursuit of top-down sequencing is therefore an ongoing effort that holds promise considering the rapid development of modern bioinformatics platforms designed to automate the identification of amino acid sequences from gas-phase fragmentation reactions. Such advances will prove important not only for top-down sequencing of intact mAbs, but for peptide mapping and middle-down applications as well.

In this chapter, data generated from analyses of the intact NISTmAb, IdeS digestion fragments, and peptides will be used to confirm that its primary amino acid sequence is consistent with the intended sequence. The techniques are presented in order of the level of detail each technique is currently capable of elucidating. Intact mass analysis provides confirmation of the protein molecular weight and major heterogeneity (e.g., glycoforms); subunit analysis of IdeS fragments provides additional localization of modifications; and peptide mapping remains the confirmatory technique to provide up to 100% amino acid sequence coverage of the protein. A discussion of the complementarity of intact and subunit/fragment analyses to peptide mapping for primary structure verification is also included. Some regulatory aspects, although not a focus of the chapter, will also be discussed. Various methods and available techniques from multiple contributors will be shown using different available technologies (columns, data analysis software, and MS instruments).

We observed that a variety of primary structure analyses were required for sequence confirmation, PTM analysis, and sequence variant analysis, each of which may require specialized expertise and a combination of technologies. For this reason, the NISTmAb reference material presented herein will be useful for internal comparison of methods and for choosing appropriate platforms as analytical technology evolves. The reference material may also be useful in assay qualification exercises to demonstrate that multiple peptide mass mapping approaches—different instruments, columns, and ionization techniques—can be used to obtain high-quality data suitable for regulatory review.

## NISTmAb Primary Structure

The basic definition of primary structure is the linear sequence of amino acid residues covalently linked by peptide bonds, which is read from the N-terminus to the C-terminus (i.e., first to last translated amino acid) (29). Although this definition describes a sequential linear arrangement, there is indeed “structure” to this configuration of amino acids because many proteins are “branched” (disulfide bonded thru intra- and inter-chain branches). Higher-order structure also inevitably arises as a result of amino acid sequence due in part to the interaction of atoms in space. Because only certain angles of alpha carbons and backbone nitrogens are permissible, the side chain of a particular amino acid within a sequence can interact with the functional groups of other residues within a limited physical proximity. Such interacting residues may be many “linear” amino acids away within the primary amino acid sequence and brought together in space due to the sum of interactions between other residues within the sequence to yield a final conformation. The term “primary structure,” therefore, is often substituted for the term “primary sequence.” The rationale for this replacement is that the sequence is the fundamental determinant of which distinct regions of the protein will interact, thereby influencing the final three-dimensional structure (Higher Order Structure chapter/Volume 3, Chapter 2).

Figure 1 shows a schematic representation of the NISTmAb IgG1 $\kappa$  reference material, with the expected disulfide bond linkages and N-glycosylation sites labeled. The corresponding primary amino acid sequence to be confirmed in the current chapter is provided in Figure 2, with the domain structure color coding as described in Figure 1. It should be noted that a variety of homology numbering schemes have been developed for IgGs due to the high level of sequence similarity (e.g., EU, Kabat) as discussed in the Mechanisms of Action chapter/Volume 1, Chapter 2(30). For well-defined mAbs, however, the actual sequence number is often used as described in Figure 1 and throughout this chapter for the NISTmAb. A combination of intact, subunit/fragment, and peptide mapping analyses will be sequentially described in the following sections to confirm the primary structure of the humanized IgG1 $\kappa$  NISTmAb.



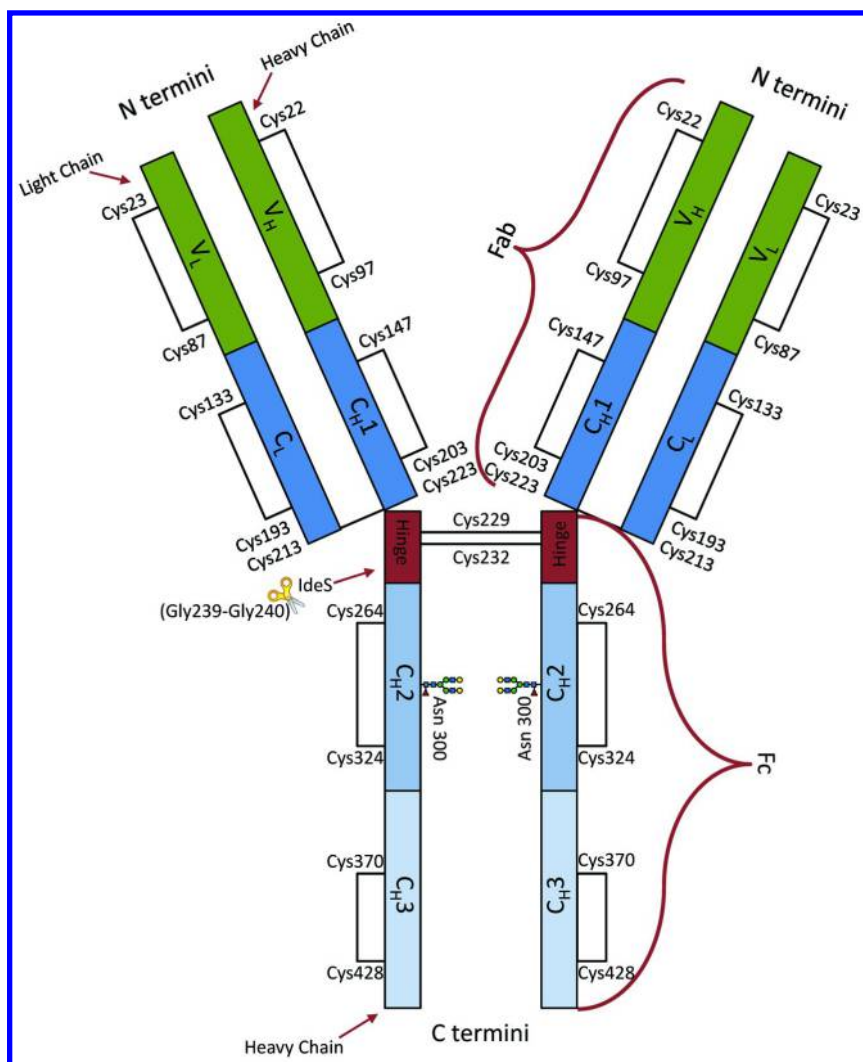


Figure 1. Schematic representation of the NISTmAb. Humanized IgG1 $\kappa$  domains are marked as variable (V) or constant (C) with subscripts for light (L) or heavy (H) chain and distinct heavy chain domains numbered 1, 2, or 3 starting from the N-terminus. Black lines indicate intra- and inter-chain disulfide bonds, with cysteine residues numbered as they appear in the light or heavy chain sequence starting from the N-terminus. The IdeS cleavage site is indicated with scissors, and the N-glycosylation site is noted with a G2F N-glycan structure. (see color insert)

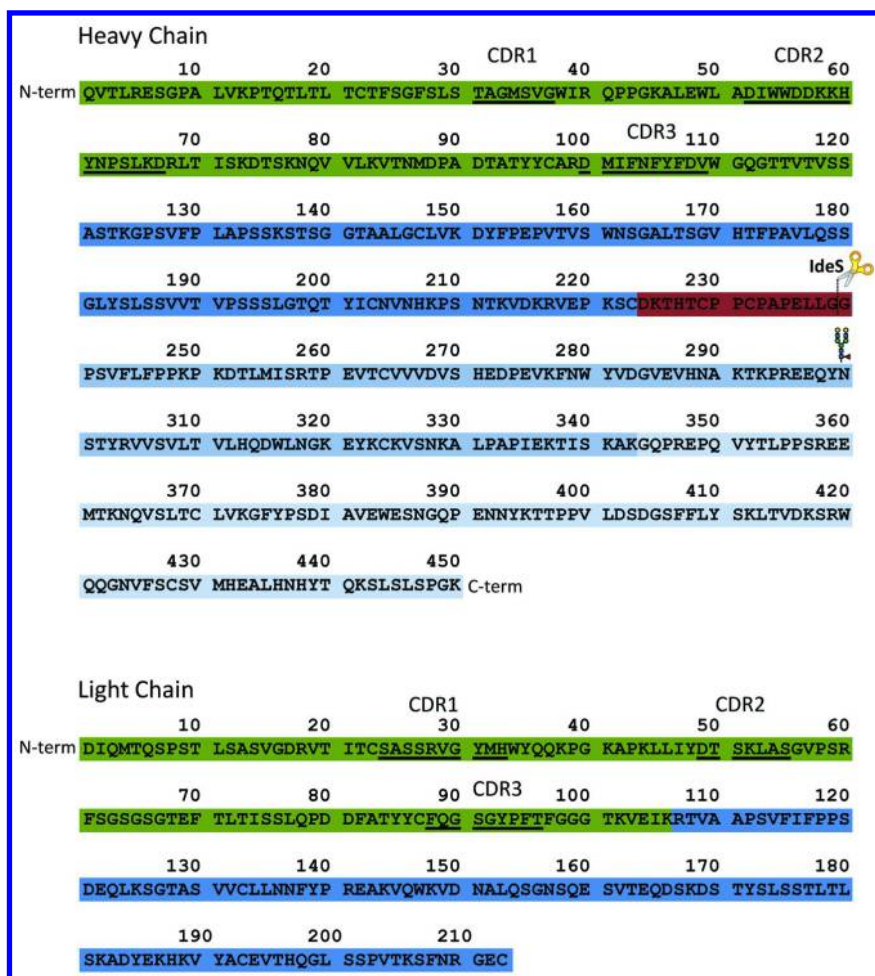


Figure 2. NISTmAb amino acid sequence. Antibody domains of the heavy and light chains are distinguished by green (variable), dark blue (constant 1), red (hinge), medium blue (constant 2), or light blue (constant 3), and complementarity-determining regions (CDRs) are underlined. The IdeS cleavage site is indicated with scissors, and the N-glycosylation site is noted with a G2F N-glycan structure. (see color insert)

## Molecular Mass Analysis of Intact mAb

The advent of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) made the analysis of large protein molecules possible by MS. MALDI is a robust ionization technique that results in predominantly singly charged ions (31), whereas ESI produces an envelope of charge states for a given protein (12). The multiple charging of intact proteins with ESI allows analysis of proteins with molecular weights greater than the mass range of the mass spectrometer. Deconvolution of the resulting multiple charge states with ESI has become the method of choice for intact protein molecular weight determination and can differentiate mAb glycoforms and other PTMs when used with high-resolution mass spectrometers (32).

A range of low- and high-resolution mass analyzers have been applied to the mass determination of intact mAbs. The highest resolution can be achieved with Fourier transform ion cyclotron resonance (FTICR) mass spectrometers. Recently, optimized FTICR conditions were developed to allow sufficiently long transient lifetimes to achieve baseline unit resolution for an intact mAb (33). This achievement has since been reproduced on additional FTICR instruments (34, 35); however, these instruments have yet to gain widespread application in the biopharmaceutical industry.

Mass analyzers more routinely used for intact mAb analysis include the quadrupole time-of-flight (QTOF) and Orbitrap mass spectrometers. The results from the analysis of the intact mAb are used to rapidly provide information on primary structure, some PTMs, and the multichain architecture of an intact antibody. Confirming that the experimental molecular mass agrees with its theoretical mass (within approximately 50 ppm or less) signifies that the antibody likely has the intended primary structure.

Chromatographic separation prior to introduction into the mass spectrometer can be used to improve throughput and sensitivity of MS-based intact measurements. Reversed-phase-high-performance chromatography (RP-HPLC) and size exclusion chromatography (SEC) both have been successfully implemented for intact mAb measurements (14, 18), and each is demonstrated in the current chapter for analysis of the NISTmAb. RP-HPLC is typically performed on a C4 to C8 column at elevated temperatures with solvents of high elutropic strength, often resulting in elution of a single peak (15). However, resolution of certain mAb variants such as C-terminal lysine, degradation products, and disulfide isoforms have been reported (36, 37). SEC for intact mass analysis is operated with organic solvents (e.g. acetonitrile [ACN]) and typically results in elution of a single peak. Using organic solvents, SEC can be used as an effective and rapid sample cleanup method for formulation buffers and salts prior to the mass spectrometer.

## Experimental Materials and Methods

### *Intact Protein Analysis*

The intact molecular mass of the candidate reference material (RM) 8670 NISTmAb, lot 3f1b, was analyzed by four different laboratories using their respective platform methodologies.

Lab 1 used reversed phase LC-MS on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Sunnyvale, CA) to directly measure the mass of the NISTmAb. The NISTmAb (10 mg/mL neat) sample was injected (5  $\mu$ g) onto a C5 column (1  $\times$  50 mm, Bio Wide Pore, Sigma Aldrich, St. Louis, MO) coupled to an Orbitrap Elite mass spectrometer and analyzed in the high mass range (2000  $m/z$  to 4000  $m/z$ ) with a resolving power of 30,000. Calibration in this mass range was performed with a solution of 3.5  $\mu$ g/ $\mu$ L poly(propylene glycol) 2700 (PPG 2700) and 7 mmol/L sodium acetate according to the manufacturer's protocol. Protein was eluted at 50  $\mu$ L/min (mobile phase A = 0.1% formic acid in water and mobile phase B = 0.1% formic acid in ACN) with a gradient of 10% to 80% mobile phase B over 30 min. Deconvolution of the mass spectra was performed with Protein Deconvolution v 2.0 (Thermo Fisher Scientific).

Lab 2 performed LC/MS using SEC coupled to a Bruker Daltonics (Billerica, MA) maXis QTOF mass spectrometer (resolving power of 30,000) for the analysis of the intact mAb. The NISTmAb was diluted to 1 mg/mL, and a 5  $\mu$ L aliquot was injected onto a BEH200 SEC column, 1.7  $\mu$ m, 4.6  $\times$  150 mm (Waters, Milford, MA). A mobile phase of 70% deionized water (diH<sub>2</sub>O)/30% ACN (v/v) with 0.05% TFA (w/v) was used to isocratically elute (and desalt) the mAb. A run time of 15 minutes, flow rate of 0.2 mL/min, column temperature of 30  $^{\circ}$ C, and detection at 214 and 280 nm were used. Mass analysis of the protein was performed using a mass range of  $m/z$  700 to  $m/z$  5000 and the following instrument parameters: nebulizer 1.6 bar, dry gas 8 L/min, capillary voltage 4500 V, iscid 160 eV, CE 15 eV, and dry temperature 220  $^{\circ}$ C. Instrument calibration was performed with Agilent ESI-L tune mix (Agilent, Santa Clara, CA) followed by lock mass correction using the ion at  $m/z$  1221.9906. Observed masses were determined from the zero-charge mass spectra following deconvolution of the multiply charged mass spectra with the Maximum Entropy algorithm in DataAnalysis software. Deconvolution parameters consisted of the following: deconvolution mass range: 140,000–170,000 Da, deconvolution data point spacing: 0.1  $m/z$ , sum peak, mass list signal to noise (S/N) threshold: 1, mass list absolute intensity threshold: 250.

Lab 3 injected 5  $\mu$ g of the mAb, which was then eluted from a phenyl MassPREP micro desalting column (Waters, Milford MA), 20  $\mu$ m, 2.1  $\times$  5 mm, and analyzed with a Waters Xevo G2 TOF mass spectrometer (operated at 20,000 resolution) in the mass range of 600  $m/z$  to 4600  $m/z$  (Waters, Milford, MA). The molecular masses of the observed proteoforms of the intact mAb were determined using the MaxEnt 1 deconvolution software (MaxEnt Solutions, Ltd., Cambridge, MA). Lab 3 experimental masses were calculated using the five most abundant charge states.

Lab 4 analyzed the NISTmAb via direct infusion into an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, Sunnyvale, CA) for intact mass analysis. Prior to analysis, the protein was desalted using a Zeba spin desalting column into Optima grade water (Thermo Fisher Scientific, Rockford, IL) and diluted to a final concentration of  $\sim 1 \mu\text{mol/L}$  in 10% ACN, 1% formic acid in water. The mAb was infused directly ( $5 \mu\text{L/min}$ ) into an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, Sunnyvale, CA) and analyzed in the high mass range ( $1800 m/z$  to  $4000 m/z$ ) with resolving power 15,000 at  $200 m/z$ . Calibration of the Orbitrap Fusion Tribrid mass spectrometer in the high mass range was performed using an enfuvirtide solution ( $44 \mu\text{g/mL}$  in 50:50 ACN:H<sub>2</sub>O with 0.1% formic acid). Approximately 125 scans were summed prior to spectra deconvolution.

Labs 1, 2, and 4 also analyzed de-N-glycosylated NISTmAb, prepared using an overnight digestion with the enzyme peptide-N-glycosidase F (PNGase F). The same instrumental platforms and analysis parameters were used as discussed above for the untreated sample.

Lab 4 performed top-down fragmentation of the de-N-glycosylated NISTmAb by direct infusion using the Thermo Fusion Tribrid mass spectrometer. After de-N-glycosylation overnight, the protein was buffer exchanged into Optima grade water using Amicon Ultra-0.5 mL centrifugal filters with 10,000 molecular weight cutoff (MWCO) membranes (Millipore, Darmstadt, Germany). The protein was diluted to  $\sim 3 \mu\text{mol/L}$  in 25% ACN, 1% formic acid prior to direct infusion at  $5 \mu\text{L/min}$  and a source fragmentation of 50 eV. The resolving power was increased to 60,000 or 120,000 at  $200 m/z$ . The 46-51+ charges states were isolated in the ion trap with a  $400 m/z$  window prior to fragmentation by CID at normalized collision energy (NCE) of 30–100% and ETD with a reaction time of 5–25 ms was performed independently. Approximately 500 scans were summed with each fragmentation method. Fragments with monoisotopic molecular masses were calculated from the raw data by the Protein Deconvolution v. 3.0 software (Thermo Fisher) in addition to being manually interpreted. The fragment coverage was matched to the L chain or H chain of the molecule using ProSight Lite (Northwestern University, Evanston, IL) in the single protein search mode with a mass tolerance of 50 ppm.

The NIST Mass and Fragment Calculator (<http://www.nist.gov/mml/bmd/bioanalytical/massfragcalc.cfm>) (38) was used to calculate theoretical masses using either International Union of Pure and Applied Chemistry (IUPAC) average atomic masses for the elements or average masses estimated from organic sources (15, 39). Theoretical masses were also calculated with PAWS (2000.06.08, Genomic Solutions, Ann Arbor, MI), a commercially available program, for comparison. The molecular masses of the proteoforms were also calculated with Protein Deconvolution v. 3.0 (Thermo Fisher Scientific, Waltham, MA) for use in the top- and middle-down sequencing sections.

## Results

Analysis of intact proteins using MS is a rapid method for empirically determining the molecular mass of a target protein along with some information on PTMs. The measured molecular mass is compared to a theoretically calculated molecular mass based on the expected amino acid composition. It is important to consider how the theoretical calculation is performed because the resultant value may differ depending on the source of constituent mass values. For example, rounding masses or using average versus monoisotopic masses for the elements will affect the final calculation. Small changes in the elemental atomic masses used magnify and cause significant shifts in the final theoretical mass for large proteins. A comparison of theoretical calculations for the NISTmAb is shown in Table 1 between IUPAC average atomic masses for the elements, average values estimated from organic sources, and mass values optimized for proteins from a commercially available program PAWS. Although perhaps only small differences are found between the calculations, it is important to understand how the theoretical values are calculated and to retain this information as the accuracy and resolving power of mass spectrometers evolve over time. Throughout this chapter, all theoretical values used are derived from the NIST Mass and Fragment Calculator software using average values of the elements estimated from natural sources unless otherwise noted (38) (e.g., Table 1, Theoretical Mass 2) as these appear to be the most appropriate values as discussed by Zhang et al. (15).

The NISTmAb was analyzed by three different online LC-MS platforms and compared to theoretical mass values. The data collected using SEC coupled to a QTOF instrument (Lab 2) are depicted in Figure 3A (raw data) and 3B (deconvoluted data). Figure 3A shows a representative raw intact mAb spectrum as multiple charge states; the fine structure corresponding to various proteoforms can be seen in Figure 3A inset. The deconvoluted data from Lab 2 identified three major glycoforms (>40% relative abundance based on peak intensity) as G0F/G0F, G0F/G1F, and G1F/G1F (Figure 3). It should be noted that glycan assignments were based on putative structures common to mAbs, which can be further verified using techniques described in the Glycosylation chapter/Volume 2, Chapter 4. Each of the major proteoforms identified are consistent with 2 L chains, 2 H chains, the expected 16 disulfide bonds, pyroglutamic acid at both H chain N-termini, and no C-terminal lysine on either H chain as depicted in Table 2. The latter two modifications are commonly observed in mAbs due to chemical or enzymatic processing during production (40–44).

Similarly, the intact protein masses were determined online using RP-HPLC coupled to an Orbitrap Elite mass spectrometer (Lab 1, Figure 3C) and a QTOF mass spectrometer (Lab 3, data not shown). The same three major glycoforms were identified and are consistent with the predicted NISTmAb structure (Table 2). All three platform methodologies recognized the same three major glycoforms, and the experimentally determined masses were within 50 ppm or less of their theoretical masses. All the LC-MS experiments also confidently identified a fourth glycoform (a minor species defined as between 3% and 40% relative abundance) as G1F/G2F. It should be noted that even for the major glycoforms identified, there may be isobaric glycoform compositions present in the sample.

For example, both G0F/G2F and G1F/G1F have the same theoretical mass and cannot be distinguished using this platform alone. Assignment of glycoforms at the mass level is based on composition and putative structure; more detailed analysis is required to confirm glycan structure as discussed in the Glycosylation chapter/Volume 2, Chapter 4.

Beyond the three most abundant peaks, differences in species observed and identification of minor and trace level species varied between the individual laboratories. The minor and trace level proteoforms identified using LC-MS included the expected glycoforms (e.g., G2F/G2F), the addition of one and/or two hexoses (e.g., from glycation and/or N-glycans containing non-human alpha-galactose linkages, gal- $\alpha$ -gal), C-terminal lysine, and singly N-glycosylated mAb. In all cases other than peak 10 in Lab 3, all identifications were made with error  $\leq 50$  ppm.

**Table 1. Theoretical Mass Calculations of Intact NISTmAb**

<i>Glycosylation Type</i>	<i>Theoretical Mass 1 (Da)</i>	<i>Theoretical Mass 2 (Da)</i>	<i>Theoretical Mass 3 (Da)</i>
G0F/G0F	148036.4	148037.2	148038.0
G0F/G1F	148198.6	148199.3	148200.1
G1F/G1F	148360.7	148361.4	148362.3
G1F/G2F	148522.8	148523.6	148524.4
G2F/G2F	148685.0	148685.7	148686.6

Theoretical Mass 1 was calculated using the NIST Mass and Fragment Calculator (<http://www.nist.gov/mml/bmd/bioanalytical/massfragcalc.cfm>) with International Union of Pure and Applied Chemistry (IUPAC) average atomic masses for the elements; Theoretical Mass 2 was calculated using the NIST Calculator with average values estimated from natural sources. Theoretical Mass 3 was calculated using PAWS (2000.06.08, Genomic Solutions, Ann Arbor, MI). Values include 16 disulfide bonds, two N-terminal pyroglutamic acids, and no C-terminal Lys.

Peaks 17 and 18 represent compositions modified by glycation and/or N-glycoforms with gal- $\alpha$ -gal structures. Both proteoforms were confirmed to be present on the NISTmAb as discussed below and in the Glycosylation chapter/Volume 2, Chapter 4, using complementary approaches. N-glycan structures corresponding to compositions with fewer N-acetyl-glucosamine (GlcNAc; -GlcNAc) residues, such as those detected in peaks 4, 5, and 6, are also identified in the intact mass spectra. These glycoforms were confirmed using orthogonal techniques such as N-glycan analysis (Glycosylation chapter/Volume 2, Chapter 4).

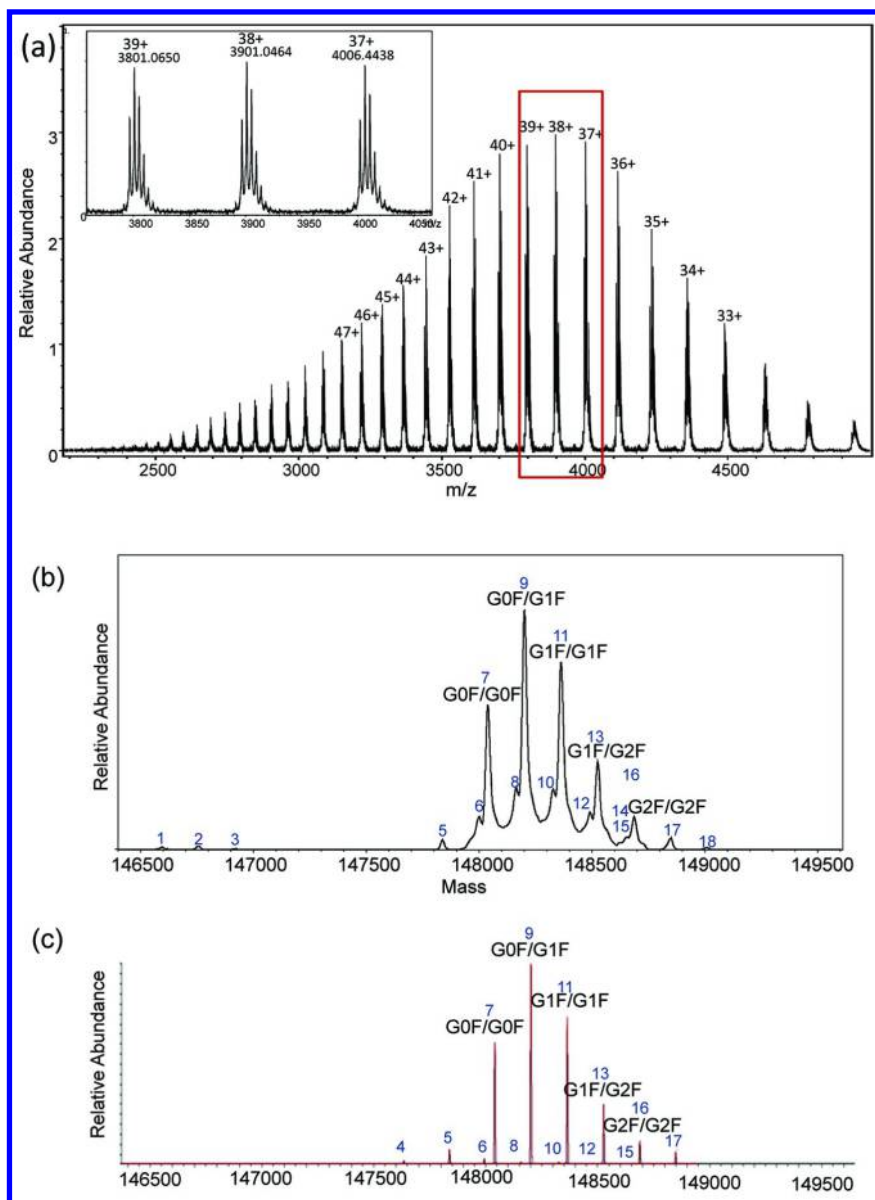


Figure 3. Intact mass measurements of the NISTmAb using liquid chromatography-mass spectrometry (LC-MS). (a) Raw mass spectrum of intact NISTmAb using an ultrahigh-resolution quadrupole time-of-flight (QTOF) mass spectrometer (Lab 2). (b) Deconvoluted zero-charge mass spectrum of intact NISTmAb using an ultrahigh-resolution QTOF mass spectrometer (Lab 2). Assignments of the trace and minor level peaks, labeled 1–18, are listed in Table 2. (c) Spectrum collected with an Orbitrap Elite mass spectrometer and deconvoluted with Protein Deconvolution. Masses of the proteoforms and assignments are shown in Table 2.



**Table 2. Average Masses of Intact NISTmAb Measured Using Orbitrap Elite (Lab 1), Bruker maXis Quadrupole Time-Of-Flight (QTOF) (Lab 2), and Waters QTOF (Lab 3) Mass Spectrometers**

<i>Proteoform</i>	<i>Theoretical Mass (Da)</i>	<i>Observed Mass (Da)</i>		
		<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>
1. G0F/Aglycosylated	146591.8	ND	146595.3 (t)	ND
2. G1F/Aglycosylated	146754.0	ND	146756.0 (t)	ND
3. G2F/Aglycosylated	146916.1	ND	146915.8 (t)	ND
4. G0F/G0F – 2GlcNAc	147630.8	147633.1 (t)	ND	ND
5. G0F/G0F – GlcNAc	147834.0	147837.9 (mn)	147837.4 (mn)	ND
6. G0F/G1F – GlcNAc	147996.1	147993.2 (t)	148000.6 (mn)	148000.0 (mn)
7. G0F/G0F	148037.2	148040.5 (M)	148039.3 (M)	148045.0 (M)
8. G0F/G0F + K	148165.3	148154.6 (t)	148164.9 (mn)	148163.0 (mn)
9. G0F/G1F	148199.3	148201.9 (M)	148201.7 (M)	148206.5 (M)
10. G0F/G1F + K	148327.5	148325.5 (t)	148328.1 (mn)	148302.0 (mn)
11. G1F/G1F	148361.4	148364.3 (M)	148363.8 (M)	148368.5 (M)
12. G1F/G1F + K	148489.6	148483.7 (t)	148492.4 (mn)	ND
13. G1F/G2F	148523.6	148526.7 (mn)	148526.1 (mn)	148531.0 (mn)
14. G1F/G1F + 2K	148617.8	ND	148624.1 (t)	ND
15. G1F/G2F + K	148651.8	148643.1 (t)	148654.5 (mn)	ND
16. G2F/G2F	148685.7	148688.5 (mn)	148687.5 (mn)	ND

*Continued on next page.*

**Table 2. (Continued). Average Masses of Intact NISTmAb Measured Using Orbitrap Elite (Lab 1), Bruker maXis Quadrupole Time-Of-Flight (QTOF) (Lab 2), and Waters QTOF (Lab 3) Mass Spectrometers**

Proteoform	Theoretical Mass (Da)	Observed Mass (Da)		
		Lab 1	Lab 2	Lab 3
17. G2F/G2F + Hex	148847.7	148847.9 (mn)	148849.4 (mn)	ND
18. G2F/G2F + 2Hex	149010.0	ND	149008.2 (t)	ND

Theoretical values include 16 disulfide bonds, two N-terminal pyroglutamic acids, and no C-terminal Lys (K), unless noted otherwise. Numbers in proteoform column refer to peak designation in Figure 3. Symbols M, mn, and t represent major (>40% of maximum peak height), minor (3% to 40% of maximum peak height), and trace (<3% of maximum peak height) relative abundance based on MS peak intensity, respectively. Putative N-glycan structures are listed. Although not listed here, alternative isobaric structures are possible in many cases.

Another common modification observed in Figure 3 is the presence of one or two C-terminal lysine residues (peaks 8, 10, 12, 14, and 15 in Table 2). Observed masses for several of the lysine-containing peaks may have alternative assignments based solely on mass accuracy ( $\leq 50$  ppm). For example, peak 8 may have been designated as a G1F/G1F-GlcNAc proteoform using intact mass data alone. Therefore, the use of orthogonal data (e.g., subunit analysis, peptide mapping, N-glycan analysis) is critical in determining the actual proteoform present in the intact analysis, allowing for more detailed structural information about the protein.

The mAb was treated with PNGase F to eliminate heterogeneity resulting from N-glycans and detect additional PTMs. The deconvoluted spectra for the de-N-glycosylated mAb are shown in Figure 4 using SEC/MS (ultrahigh-resolution-QTOF) (Lab 2) and RP-HPLC/MS (Orbitrap Elite) (Lab 1), with observed and theoretical masses listed in Table 3. Following de-N-glycosylation, the most abundant isoform is the fully de-N-glycosylated mAb. Glycated proteoforms (+162 Da and +324 Da, peaks c and f, respectively) were observed at low levels, as well as the presence of one or two C-terminal lysine residues (+128 Da and +256 Da, peaks b and d, respectively), and the combination of the two modifications (peak e).

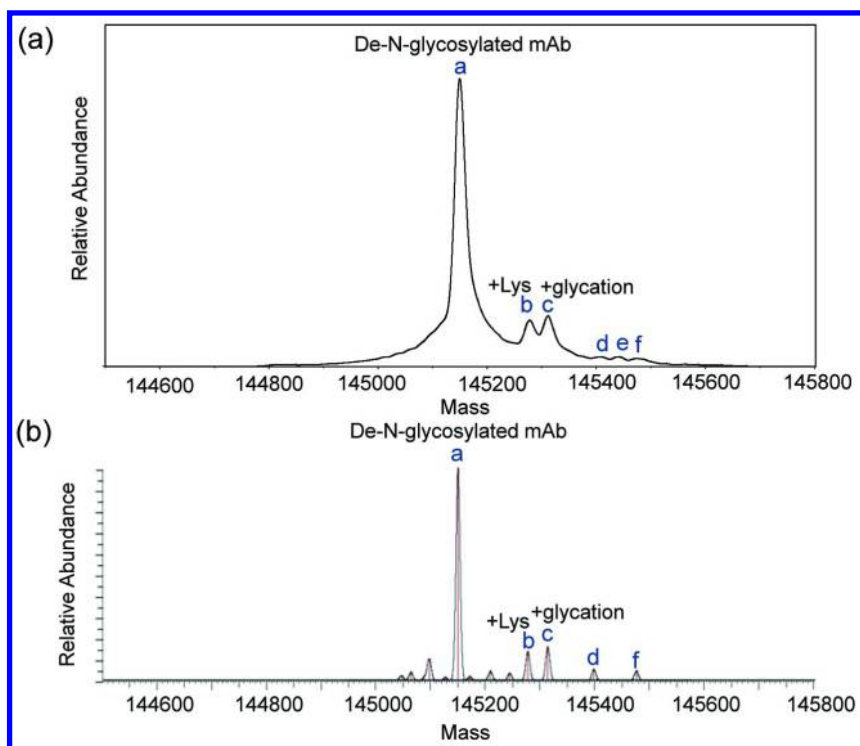


Figure 4. Intact mass measurements of de-N-glycosylated NISTmAb analyzed by liquid chromatography-mass spectrometry (LC-MS) on (a) an Orbitrap Elite (Lab 1), and (b) a maXis quadrupole time-of-flight (QTOF) (Lab 2). Masses of the identified proteoforms are shown in Table 3.

**Table 3. Average Masses of De-N-glycosylated NISTmAb**

Proteoform	Theoretical Mass (Da)	Observed Mass (Da)	
		Lab 1	Lab 2
a. De-N-glycosylated	145148.5	145150.3 (M)	145150.4 (M)
b. + Lysine (+Lys)	145276.6	145277.8 (t)	145278.1 (t)
c. + Hexose (+Hex)	145310.6	145314.3 (mn)	145312.2 (mn)
d. +2 Lys	145404.8	145398.5 (t)	145406.7 (t)

Continued on next page.

**Table 3. (Continued). Average Masses of De-N-glycosylated NISTmAb**

Proteoform	Theoretical Mass (Da)	Observed Mass (Da)	
		Lab 1	Lab 2
e. +Lys and +Hex	145438.6	ND	145440.1 (t)
f. +2 Hex	145472.7	145476.3 (t)	145478.1 (t)

Letters in Proteoform column refer to peak designation in Figure 4. Symbols M, mn, and t represent major (>40% of maximum peak height), minor (3% to 40% of maximum peak height), and trace (<3% of maximum peak height) relative abundance based on mass spectrometry (MS) peak intensity.

### Top-Down Sequencing Results

The fourth laboratory in this study designed their intact mass measurement experiments with the intention of further evaluating the NISTmAb structure using top-down fragmentation. For this experiment, the sample was buffer exchanged to a MS-compatible solvent and directly infused. The entire charge state envelope was deconvoluted to determine the molecular mass for the intact and de-N-glycosylated NISTmAb, shown in Figure 5A and 5B, respectively. For both the intact and the de-N-glycosylated NISTmAb, adducts were observed that likely were retained during the protein purification process (denoted by asterisks).

Deconvolution of NISTmAb spectra collected at 30,000 resolving power allowed detection of the five most abundant glycoforms (>20 ppm mass error, Figure 5A). Interestingly, no species corresponding to -GlcNAc were observed, indicating ionization conditions and/or instrument configuration provided softer ionization not conducive to in-source fragmentation. The deconvoluted de-N-glycosylated NISTmAb spectra are also in agreement with the two most dominant forms observed in the LC-MS experiments discussed above. In the infusion mode, it was noted that a higher propensity for adduct formation was observed (Figure 5). The relatively large quantity of adduction formed during the infusion experiment has been reported previously (45), and the approximately +98 Da shift is commensurate with a phosphate adduct remaining as an artifact from sample preparation. The higher quantity of adduct formation explains the identification of fewer low-abundance proteoforms during the infusion experiment. Alternate purification strategies may be employed to minimize the appearance of these species.

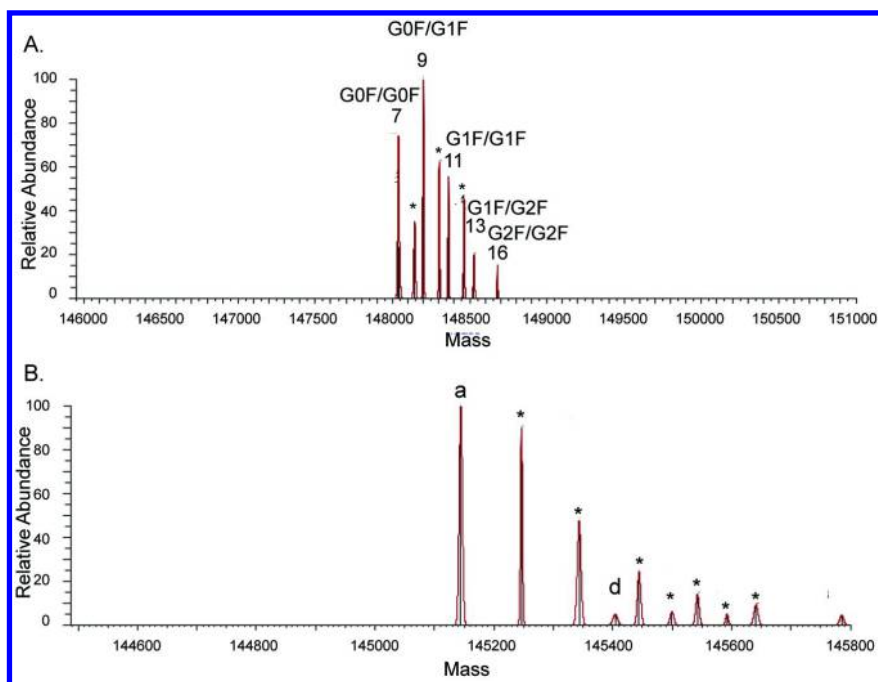


Figure 5. Intact mass measurement of the NISTmAb by direct infusion. The intact molecular mass of the NISTmAb was determined using a Fusion Tribrid mass spectrometer for the (a) intact NISTmAb with numbering or lettering consistent with that listed in Table 2, and (b) de-N-glycosylated NISTmAb with the labels consistent with those listed in Table 3. The asterisks (\*) denote adducts remaining from the purification process.

Despite the appearance of adducts during infusion experiments, this mode of analysis affords the opportunity to sum multiple fragmentation spectra of the intact parent ions over an extended time range to provide MS/MS spectra. Top-down analysis of the de-N-glycosylated NISTmAb was performed on the Orbitrap Fusion Tribrid mass spectrometer following direct infusion of the protein. Source fragmentation energy was applied at 50 eV, and the 46–51+ charge states were isolated in the ion trap with a 400  $m/z$  window and then fragmented by CID (30–100% NCE) or ETD (5–25 ms reaction times). Approximately 500 scans were summed to generate a single spectrum for deconvolution. The monoisotopic masses of the antibody fragments were extracted by Xtract within the Protein Deconvolution v. 3.0 software and assigned to either the L or H chain in ProSight Lite with a 50 ppm error tolerance. The known modifications, such as N-terminal pyroglutamic acid and C-terminal lysine cleavage on the H chain, were included in the ProSight Lite parameters as well as the expected intra-strand disulfide bonds. The loss of one disulfide near the N- or C-terminus of both the L and H chain was allowed to account for disulfide cleavage during fragmentation. These data were also fit using a 10 ppm error tolerance in ProSight Lite. The sequence coverage of the intact, non-reduced NISTmAb is shown below (Figure 6). For the H chain, 85 unique fragments corresponding to 68 intramolecular bonds

broken were identified, covering 15.1% of the protein sequence with the majority of the fragmentation occurring near the N and C termini. There were 44 unique fragments identified corresponding to the L chain with 33 intra-chain bonds broken, covering 15.5% of the protein sequence. The top-down approach taken here is relatively automated and quickly identified the fragments corresponding to the termini of the H and L chains. Despite the relatively low sequence coverage obtained as compared to peptide mapping, the automated nature of this combined analysis yields orthogonal data and high confidence in the predicted protein sequence. Continued improvement in both instrument and software capabilities will likely result in top-down methods playing a more prominent role in future biopharmaceutical platform workflows.

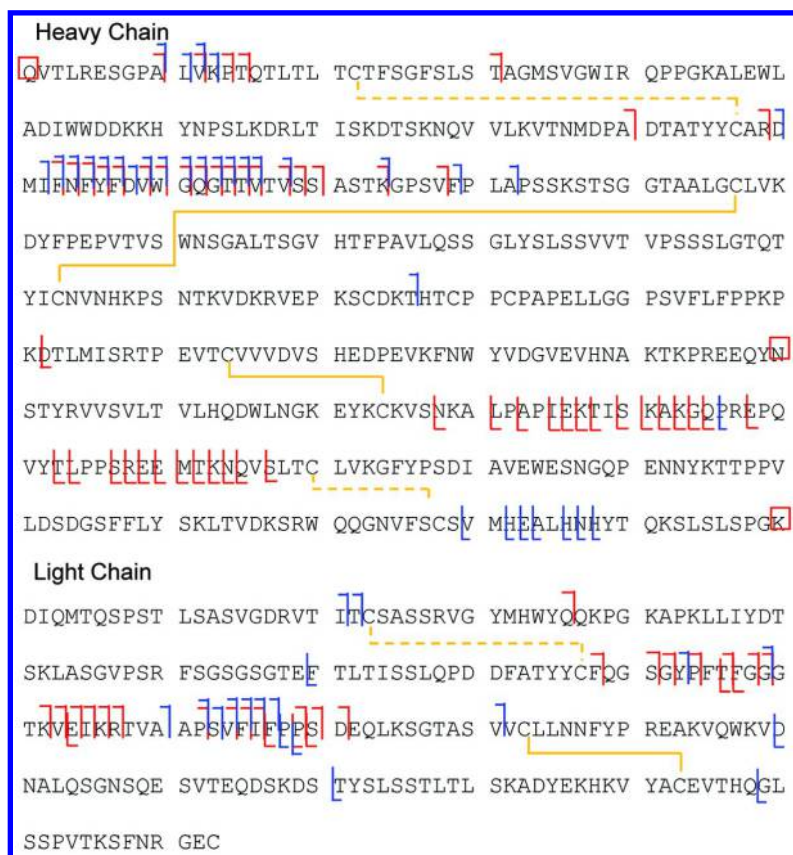


Figure 6. Sequence coverage of the NISTmAb by top-down analysis. The sequence coverage from direct infusion of the NISTmAb of the heavy (top) and light (bottom) chains are shown with observed fragments from collision-induced dissociation (CID) and higher energy collision dissociation (HCD) fragmentation (blue), as well as electron transfer dissociation (ETD) fragmentation (red). Disulfide bonds are shown in yellow, with those cleaved during fragmentation shown as dashed lines. The fragments were assigned using ProSightLite and confirmed manually. (see color insert)

## Discussion

Intact mAb mass measurements provide valuable information about the protein, require little to no sample preparation, and can be performed quickly. The high resolving power of TOF and Orbitrap mass analyzers allows direct observation of mAb heterogeneity, including N-glycosylation and glycation, as well as N- and C-terminal modifications for known and unknown protein sequences. The use of these techniques for intact structural analysis is gaining momentum for the initial analysis of mAb primary structure. Improvement in mass spectrometer instrumentation continues to be an active research area (46) and may one day provide a routine platform for even greater structural information.

Intact mass analysis provides preliminary assurance of the amino acid composition (along with some PTMs) within the limitations of the instrumentation. The putative matches reported for the intact measurements in the current chapter resulted in average mass error (based on data in Table 2) of 26.6, 16.4, and 59 ppm for Lab 1, Lab 2, and Lab 3, respectively. This corresponds to a mass error of approximately 2 to 9 Da. An absolute differentiation of isobaric amino acids of course cannot be made (e.g., Ile vs. Leu), but it is also worth noting that near-isobaric compositions cannot be unequivocally differentiated at this level (e.g., Lys vs. Glu, Leu/Ile vs. Asn or Asp). Unequivocal assignment of composition becomes increasingly complex with additional size; for example, amino acid pairs EG, DA, VS, and single residue W all have a nominal mass of 186 Da. Despite this apparent limitation, one must consider that the recombinant product is expressed through a highly controlled and engineered cell line intended to produce a protein of known sequence. Observation of a mass that closely conforms to the expectation therefore provides confidence in the putative assignment, which can be further verified with orthogonal techniques.

The examples shown in this chapter highlight the ability to observe proteoforms with differences in glycosylation, N- and C-terminal modifications, and glycation of a mAb across multiple mass spectrometer platforms. The intact mass analysis does not allow for the identification of modifications such as deamidation, which has an overall smaller change in the intact mass than can be resolved in a molecule of this size, or inversion of amino acids within the sequence, which does not change the overall molecular weight of the molecule. Furthermore, proteoforms with lower overall abundance were not consistently identified by all of the laboratories (unique identifications at the trace level). The presence of PTMs such as intrinsic deamidation present at low levels may therefore not be observable regardless of the instrument resolving power. Although a threshold of intensity necessary for identification is beyond the scope of this chapter, observation of an expected composition (e.g., +162 is indicative of a hexose) provides reasonable assurance of PTM assignments, which can then be further verified.

For those proteoforms that are present in relatively high abundance, top-down sequencing can be used to fragment the peptide backbone of the protein to gain further confidence that the protein sequence is accurate (Figure 6). Significant advances in top-down sequencing of mAbs have been made using in-source decay

with pseudo MS<sup>3</sup> of product ions (26), ETD (23), and ECD (24, 25). However, complete sequence coverage of intact proteins of this size using these methods is limited, making this a developing strategy with great potential for rapid sequence confirmation.

Ultimately, intact mass analysis is a very powerful technique capable of evaluating mAb primary structure with minimal sample handling. However, it cannot unequivocally differentiate all heterogeneity intrinsic to a mAb. The confidence of intact mass assignments is related to process knowledge and experience and can therefore yield actionable data. Ultimately, analysis of smaller species (e.g., subunits, fragments, peptides) allows for higher resolution of the individual components and can lead to improved mass accuracy and better confidence in mass assignments. Subunit analyses and peptide mapping of the NISTmAb are described below to provide complementary primary structure assignment by enzymatic digestion and additional MS analysis.

## Subunit and IdeS Fragment Analysis

The analysis of mAb H and L chain subunits has been used to support primary structure confirmation and detect PTMs (47, 48). These analyses can be performed by chemical reduction (dithiothreitol [DTT] or *tris*(2-carboxyethyl)phosphine [TCEP]) followed by LC-MS analysis. Upon chemical reduction of the antibody, the H and L chains are formed and can be chromatographically separated during LC-MS analysis. In addition, alkylation of the resulting sulfhydryl residues can be performed to minimize reformation of disulfide bonds. A promising new approach to subunit analysis uses online electrochemical reduction for direct infusion of subunits (34). This technique resulted in a relatively complex data set because only partial reduction was achieved; however, it shows promise for a completely automated subunit analysis workflow.

Enzymes such as papain and pepsin can cleave IgG molecules in the hinge region to produce large fragments, although digestion efficiency varies between IgG subclasses due to differences in disulfide linkages within their hinge regions (49, 50). Pepsin cleaves the IgG H chain below the hinge region, resulting in a F(ab')<sub>2</sub> fragment (i.e., two Fab fragments that remain linked via the two hinge disulfide bonds); however, the Fc is further digested into many smaller pieces (51, 52). IgG3 and IgG4 antibodies are highly sensitive to pepsin fragmentation, whereas IgG1 and IgG2 molecules are resistant (53). IgG molecules also show subtype specificity in regards to digestion with papain (54, 55). IgG1 and IgG3 antibodies are both sensitive to the cysteine protease papain, which digests the H chain above the hinge region, resulting in two Fab fragments and the Fc fragment (56–58). IgG4 molecules are resistant to papain cleavage in the absence of cysteine, but can be cleaved if cysteine is provided for the reaction. IgG2 remains fairly resistant to papain cleavage under either condition. Both pepsin and papain have been reported to lack reliable specificity and often produce heterogeneous cleavages (59). As a result, alternative enzymes for subunit analysis are gaining in popularity.



The protease IdeS (originally isolated from *Streptococcus pyogenes* and now available in recombinant form) selectively cleaves between the G-G residues (NISTmAb Gly239-Gly240) in the hinge region of the antibody to produce the  $F(ab')_2$  and single chain Fc (scFc) fragment ( $C_{H2}$  and  $C_{H3}$  domains) (Figure 7A) (60). The  $F(ab')_2$  can be further reduced (e.g., TCEP or DTT), resulting in the generation of the L chain (which also happens to be the complete subunit) and  $Fd'$  fragments ( $V_H$ ,  $C_{H1}$ , and hinge regions) (Figure 7B). The three antibody components (scFc,  $Fd'$ , and L chain), each approximately 25 kDa, can be chromatographically separated when analyzed by LC-MS. The combination of the antibody fragment's size with current methodology and instrumentation provides a convenient platform for (1) highly accurate mass measurements; and (2) the possibility of middle-down sequencing with a variety of MS/MS fragmentation modes (17–20).

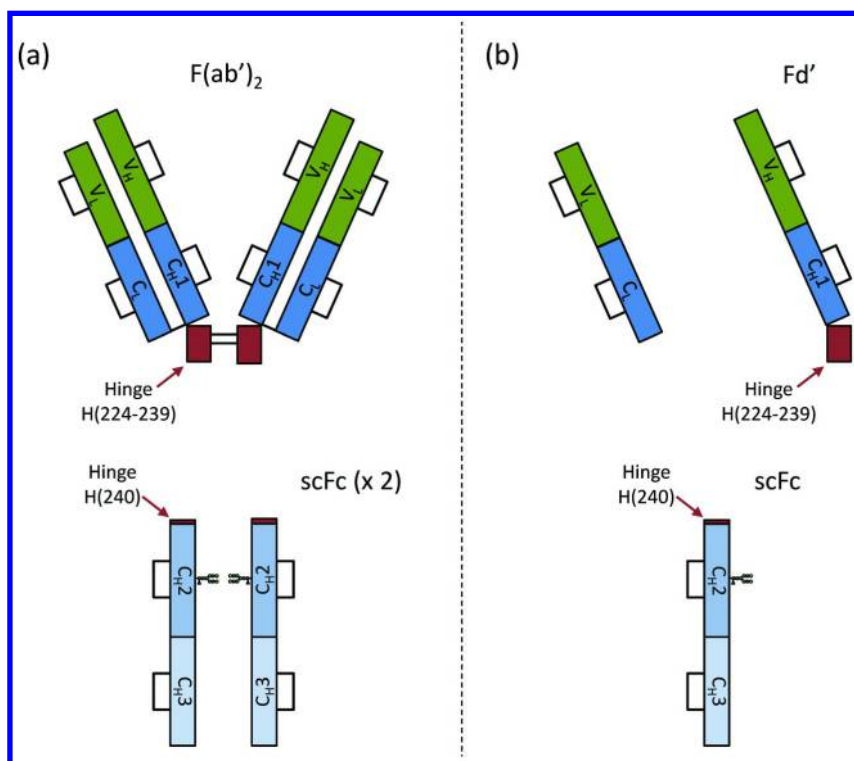


Figure 7. NISTmAb fragments after IdeS digestion. Digestion with IdeS results in (a) one  $F(ab')_2$  fragment and two scFc fragments. Reduction of disulfide bonds results in (b) two light chain subunits, two  $Fd'$  fragments (1-239), and the two previously formed scFc fragments (240-449); for which one of each is shown in (b). Domains comprising each fragment are labeled as described for Figure 1. (see color insert)

## Experimental Materials and Methods

Labs 1, 2, and 4 performed subunit/fragment analysis on the NISTmAb. Lab 1 diluted the mAb to 2 mg/mL in 50 mmol/L phosphate buffer containing 150 mmol/L NaCl at pH 6.6. Samples were digested with IdeS (FabRICATOR® IgG protease, Genovis AB, Cambridge, MA) following the manufacturer's protocol. IgG fragments were then reduced by adjusting to 1 mol/L guanidine and 50 mmol/L TCEP (pH 2.0) and incubating at 80 °C for 30 min. Analysis of the digests was performed on an Orbitrap Elite mass spectrometer coupled to a C8 column (2.1 × 100 mm, Bio Wide Pore, Sigma Aldrich). Protein subunits/fragments (10 µg) were eluted (mobile phase A = 0.02% TFA in water, mobile phase B = 0.02% TFA in ACN) at 200 µL/min starting at 10% mobile phase B and increasing by 1% per min over 40 min. MS spectra were deconvoluted with Xtract (Thermo Fisher Scientific).

Lab 2 also performed an IdeS digestion (FabRICATOR® IgG protease, Genovis AB) prior to LC/MS analysis. The NISTmAb was diluted to 2 mg/mL (50 µL total) in IdeS digestion buffer (pH 6.6). Two microliters of IdeS enzyme was added, gently vortexed, and incubated at 37 °C. Denaturation and reduction involved adding 100 µL of 8M guanidine, 25 µL of IdeS digestion buffer, and 25 µL of 1M DTT to the mAb sample and incubating at 37 °C for 45 minutes. The resulting fragments (10 µL or 5 µg load) were injected on a C4 reversed phase column (Waters BEH300 C4, 1.7 µm, 2.1 × 100 mm) at a column temperature of 65 °C. RP-UHPLC/ESI-QTOF MS was performed on a Waters H-Class UHPLC coupled to a Bruker Daltonics (Billerica, MA) maXis QTOF mass spectrometer. RP-UHPLC conditions also included flow rate: 0.2 mL/min, run time: 70 min, and a detection wavelength of 214 nm. The chromatographic gradient used three mobile phases: mobile phase A: 95% water/5% ACN (v/v), mobile phase B: 2% TFA (w/v), and mobile phase C: 50% ACN/50% isopropanol. The gradient was generated using a constant 5% B while C was raised from 3 to 40% in 45 min. A 10 minute cleaning period (90% C) was used followed by a 15 minute re-equilibration period at initial conditions. Mass spectrometer settings include a mass range of  $m/z$  600 to  $m/z$  3000, nebulizer: 1.6 bar, dry gas: 8 L/min, iscid: 45 eV, collision energy: 15 eV, and dry temperature: 200 °C. Instrument calibration was performed with Agilent ESI-L tune mix (Agilent, Santa Clara, CA) followed by lock mass correction using the ion at  $m/z$  1221.9906. Observed masses were determined from the zero-charge mass spectra following deconvolution of the multiply charged mass spectra with the Maximum Entropy algorithm in DataAnalysis software. Deconvolution parameters consisted of the following: deconvolution mass range: 15,000 to 60,000 Da, deconvolution data point spacing: 0.05  $m/z$ , sum peak, mass list S/N threshold: 0, mass list absolute intensity threshold: 0.

For the de-N-glycosylated subunit analysis, the intact NISTmAb was first diluted to 2.0 mg/mL (50 µL total) in IdeS digestion buffer (pH 6.6) and 2 µL of PNGase F was added. The sample was then incubated at 37 °C for 18 hours. The sample was then denatured and reduced as stated above.

Lab 4 performed IdeS digestion according to the manufacturer's protocol on the de-N-glycosylated NISTmAb. The digested mAb was reduced with 45

mmol/L TCEP and 3 mol/L guanidine HCl (final) prior to being desalted using 10,000 MWCO spin columns (Amicon, Darmstadt, Germany) and Zeba desalt spin columns (Thermo Fisher Scientific). The fragment mixture was diluted to approximately 0.5 mg/mL based on a starting concentration in 40% ACN, 1% formic acid (Optima grade) and directly infused into an Orbitrap Tribrid Fusion mass spectrometer.

## Results and Discussion

### *IdeS Fragment Mass Analysis*

Further analysis of the NISTmAb primary structure was performed in Labs 1 and 2 by IdeS fragment mass analysis using IdeS digestion followed by reduction of the disulfide linkages. LC MS was used to separate the resulting fragments (see Figure 8) and determine their accurate masses. Theoretical and experimentally determined monoisotopic mass assignments are summarized in Table 4.

The observed masses of the predominant forms of the scFc fragment were consistent with the expected primary amino acid sequence plus the G0F or G1F N-glycans. Due to the increased sensitivity of the IdeS fragment analysis, several additional trace level glycoforms also were detected (see Table 4). Moreover, the chromatographic resolution combined with accurate mass assignment of a fragment bearing C-terminal Lys (Figure 8) allowed for confident assignment of this trace-level species indicated by the intact mAb measurements above. Low levels of the aglycosylated scFc also were detected eluting after the N-glycosylated scFc form.

The observed masses of the predominant L chain subunit and Fd' fragment were also consistent with their calculated theoretical masses (see Table 4 and Figure 9). The major L chain species has the expected amino acid sequence with only a trace level of glycation (+162 Da). The Fd' fragment was observed with pyroglutamic acid (-17 Da) at the N-terminus. In addition, low levels of oxidation (+16 Da) and glycation were detected on the Fd' fragment.

Minor differences in species observed between the two laboratories are readily observed in the UV traces shown in Figure 8. The chromatogram from Lab 1 showed higher levels of cleavage between Asp-Pro residues (e.g., D(P) clipping), presumably due to higher temperature or lower buffer pH during reduction of disulfide bonds (61). Cleavage at Asp-Pro residues has been well documented and linked to the basicity of the proline residue nitrogen (61). As with other sample preparation techniques, the optimization of reaction conditions to minimize sample artifacts (e.g., D(P) clipping) while allowing reaction completion is ultimately desired.

Both laboratories were able to detect oxidized forms of the Fd' fragment by MS; however, chromatographic resolution of the oxidized species was obtained only by Lab 2. A number of parameters may be responsible for the variations in the UV profiles obtained between the two laboratories, including column format, chemistry, and selectivity resulting from the different mobile phase compositions used. The chromatographic method used by Lab 2 may be preferable for

characterization and stability testing of this mAb as it allows a greater number of variants to be visually observed and identified.

An expanded view of the deconvoluted, de-isotoped mass spectral data for the L chain, determined by Lab 1 and Lab 2, is shown in Figure 10. Simulated isotopic profiles are also included as calculated by the respective instrument platform software. It should be noted that the data for Lab 1 were collected in profile mode; however, the Xtract software automatically centroids the data during processing.

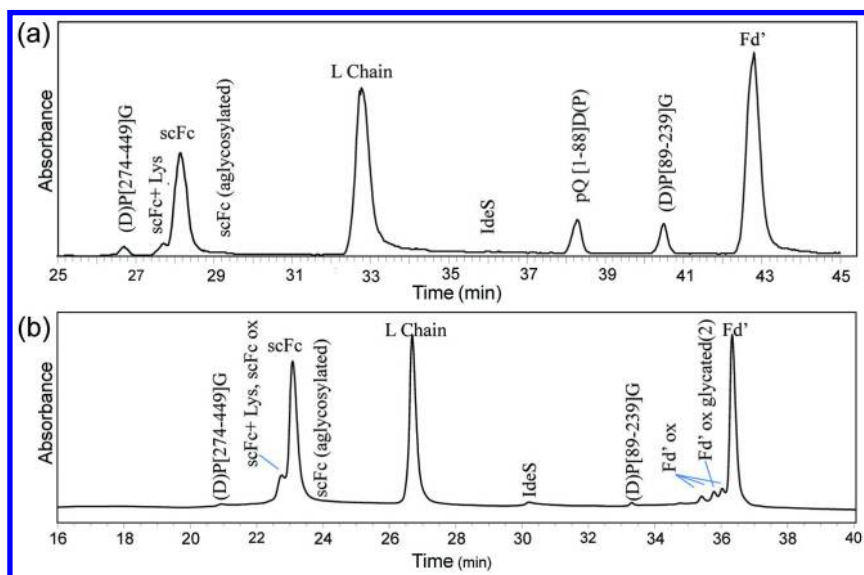


Figure 8. Liquid chromatography-mass spectrometry (LC-MS) fragment analysis following IdeS digestion and reduction. UV chromatograms of the IdeS fragments of the NISTmAb are shown (masses compiled in Table 4) for experiments performed by (a) Lab 1 and (b) Lab 2. Abbreviations: scFc represents the single-chain Fc fragment (heavy [H] chain residues 240-449); scFc + Lys represents the presence of C terminal lysine (H chain residues 240-450); L represents light chain (L chain residues 1-213); Fd' represents the  $V_H$ ,  $C_{H1}$ , and partial hinge (H chain residues 1-239); and IdeS represents the IdeS enzyme. Other fragments of the heavy chain such as (D)P[274-449]G are shown with residue numbers in brackets from the N- to C terminal residues (cleavage occurs C terminal to residue in parentheses).

**Table 4. Accurate Mass Assignments (Monoisotopic) for the NISTmAb IdeS Fragments**

<i>IdeS</i> Fragment	<i>Proteoform</i>	<i>Theoretical Mass (Da)</i>	<i>Observed Mass (Da) <sup>a</sup></i>	
			<i>Lab 1</i>	<i>Lab 2</i>
scFc (H(240-449)) <sup>b</sup>	G0F	25220.463	25220.464 (M)	25220.477 (M)
	G1F	25382.516	25382.554 (M)	25382.530 (M)
	G2F	25544.569	25544.556 (mn)	25544.580 (mn)
	G2F + G2F + Hex	25706.622	25706.457 (t)	25706.624 (t)
	G0FG0F + Lys	25348.558	25348.555 (mn)	25348.561 (mn)
	G1FG1F + Lys	25510.611	25510.691 (mn)	25510.615 (mn)
	G2FG2F + Lys	25672.664	25672.710 (t)	25672.678 (t)
	aglycosylated	23775.930	23775.985 (t)	23775.934 (t)
	(D)P[274-449]G + G0F	21499.574	21499.633 (mn)	21499.599 (mn)
	(D)P[274-449]G + G1F	21661.627	21661.640 (mn)	21661.654 (mn)
	(D)P[274-449]G + G2F	21823.680	21823.679 (mn)	21823.707 (mn)
	G1	25236.458	ND	25236.477 (t)
	G2	25398.511	ND	25398.521 (t)
	G0F – GlcNAc	25017.384	ND	25017.397 (t)
	G1FG1F – GlcNAc	25179.437	ND	25179.451 (t)
	G2F + 2G2F + 2 Hex	25868.675	ND	25868.676 (t)
	G1FG1F + GlcNAc	25585.596	ND	25585.604 (t)

<i>IdeS</i> Fragment	<i>Proteoform</i>	<i>Theoretical Mass (Da)</i>	<i>Observed Mass (Da)</i> <sup>a</sup>	
			<i>Lab 1</i>	<i>Lab 2</i>
	G2FG2F + GlcNAc	25747.648	ND	25747.663 (t)
L chain	unmodified	23113.304	23113.329 (M)	23113.324 (M)
	+ Hex	23275.357	23275.354 (t)	23275.374 (t)
Fd' (H(1-239)) <sup>c</sup>	pQ	25672.807	25672.835 (M)	25672.82 (M)
	pQ + Hex	25834.807	ND	25834.816 (t)
	pQ, oxidized <sup>d</sup>	25688.806	25688.879 (t)	25688.807 (t)
	pQ, oxidized + 2 Hex	26012.912	26012.989 (t)	26012.989 (t)
	(D)P[89-239]G	15864.738	15864.741 (mn)	15864.775 (mn)
	pQ[1-88]D(P)	9826.079	9826.128 (t)	ND

Theoretical monoisotopic masses were calculated with internal software and include N-terminal pyroglutamic acid and lack of C-terminal Lys unless otherwise noted. <sup>a</sup> Relative abundance categories M: major (>40% of maximum peak height), mn: minor (3 to 40% of maximum peak height), t: trace (<3% of maximum peak height). <sup>b</sup> The presence of Lys450 is denoted as + Lys. <sup>c</sup> pQ = pyroglutamic acid. <sup>d</sup> Lab 2 observed three chromatographically resolved species of this composition.

It is clear that increased resolution is obtained for lower mass fragments as compared to intact mAb measurements. It should be noted, however, that although isotopic resolution is achievable in this mass range, the isotopic distribution is such that the monoisotopic peak is not detected. The monoisotopic peak at this mass range is in fact usually not detected in such experiments due to its low relative abundance compared to isotope-containing species. Despite this, calculation of the unmodified L chain monoisotopic mass is based on experimental data and is calculated using instrument manufacturer software algorithms. In both cases, it is within 2 ppm of the theoretical value as shown in Table 4. In addition, the isotopic distribution, as well as the relative abundance of each isotopic peak, is consistent with that predicted in the simulation for both experiments. Collectively, the increased resolution provides higher confidence for primary structure confirmation.

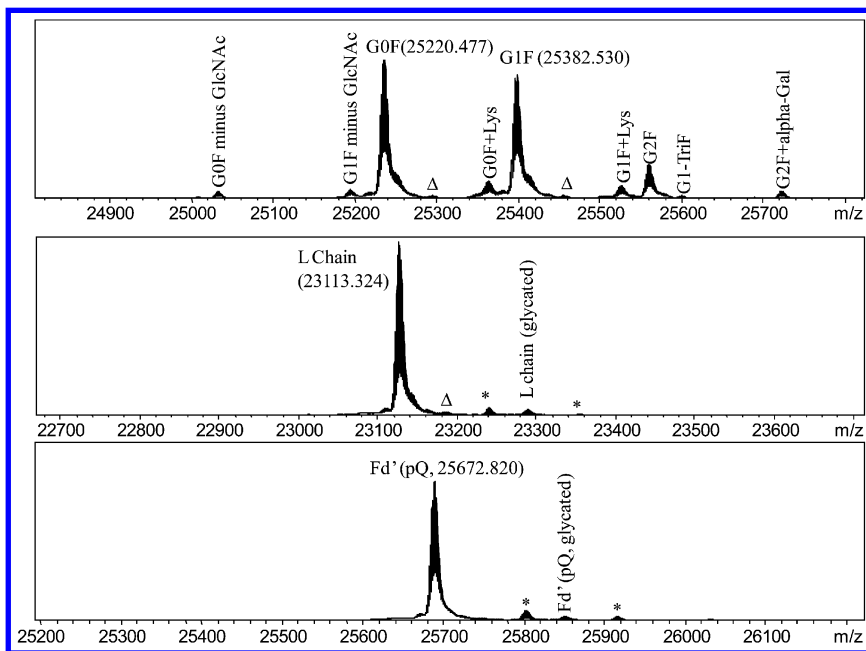


Figure 9. Lab 2 deconvoluted mass spectra of NISTmAb fragments following IdeS digestion and reduction. Zero-charge mass spectra of the scFc fragment (top), light (L) chain subunit (middle), and Fd' fragment (bottom) of the NISTmAb. The abbreviation “pQ” represents pyroglutamic acid formed from the N-terminal amino acid glutamine. Observed masses of major, minor, and trace-level species are found in Table 4. The “\*” represents a trace level of trifluoroacetic acid (TFA) adduction (+114 Da), and the “Δ” represents guanidine adduction (+59 Da); both are method-related artifacts.

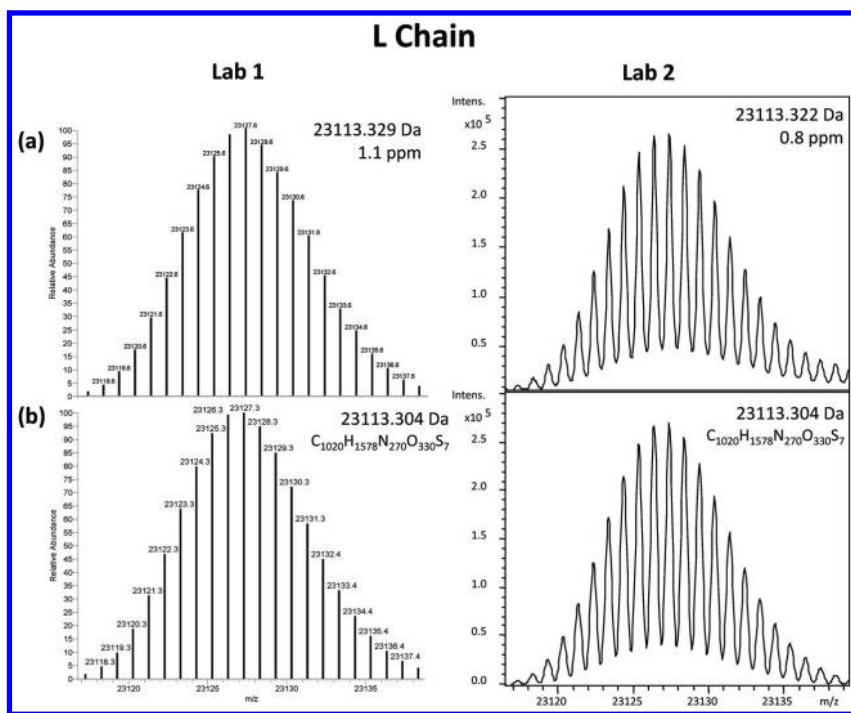


Figure 10. Experimental (a) and theoretical (b) light (L) chain subunit mass spectra of the NISTmAb following IdeS digestion and reduction. Note that data for Lab 1 were collected in profile mode; however, Xtract software automatically centroids data for output.

This IdeS fragment analysis with high mass accuracy is in good agreement with the intact protein measurements presented in Tables 2 and 3. However, IdeS fragment mass analysis was shown to have improved mass accuracy and resolution compared to intact mass measurements and to provide more confident assignments. Chromatographic resolution of the L chain subunit and H chain fragments also allowed identification of additional mAb heterogeneity. IdeS fragment analysis therefore offers an orthogonal measurement, confirming primary structure heterogeneity. Although these complementary data are highly informative, the methods require additional sample processing as compared to intact mAb analysis and may require optimization.

Finally, the mAb was treated with PNGase F to reduce the structural protein heterogeneity resulting from N-linked glycosylation in the Fc region and then analyzed by LC-MS analysis (Lab 2, data not shown). The deconvoluted mass spectra obtained for the L chain subunit and Fd' fragment following de-N-glycosylation remain unchanged, indicating no N-glycosylation sites on the NISTmAb L chain or the C<sub>H</sub>1, variable, or hinge domains of the H chain. In contrast, the scFc fragment contains less structural heterogeneity following de-N-glycosylation. As a result, one major scFc species was detected and was



consistent with its theoretical de-N-glycosylated mass, which includes all four cysteines in the reduced form, no C-terminal Lys, and the conversion of the de-N-glycosylated asparagine residue (Asn) to an aspartic acid residue (Asp). Trace level glycation and oxidation were also identified in the scFc (Table 5).

**Table 5. Accurate Mass Assignments for the De-N-glycosylated IdeS Fragments of NISTmAb**

<i>IdeS Fragment</i>	<i>Proteoform</i>	<i>Theoretical Monoisotopic Mass (Da)</i>	<i>Lab 2 Observed Monoisotopic Mass (Da)<sup>a,b</sup></i>
De-N-glycosylated scFc (H(240-449)) <sup>c</sup>	unmodified	23776.914	23776.928 (M)
	oxidized	23792.909	23792.919 (t)
	oxidized	23792.909	23792.922 (t)
	oxidized (2)	23808.904	23808.894 (t)
	glycated	23938.967	23938.972 (t)
	+Lys	23905.009	23905.013 (mn)
	+Lys, oxidized	23921.004	23920.993 (t)
	aglycosylated	23775.930	23775.911 (t)
	(D)P[274-449]G	20056.024	20055.999 (t)
	(D)P[274-449]G, oxidized	20072.019	20071.992 (t)

<sup>a</sup> Monoisotopic masses were determined from the zero-charge mass spectra after deconvolution and de-isotoping of the multiply charged data and accounting for Asn300 to Asp300 conversion due to de-N-glycosylation. All observed masses agree with theoretical masses to within 2 ppm for major and minor proteoforms and within 3 ppm for trace-level species, which is consistent with the accuracy of contemporary mass spectrometers for isotopically resolved proteins. <sup>b</sup> Relative abundance categories: M: major (>40% of maximum peak height), minor: mn (3 to 40% of maximum peak height), T: trace (<3% of maximum peak height). <sup>c</sup> The presence of Lys450 is denoted as “+Lys.”

### *Middle-Down Sequencing*

Middle-down sequencing analysis was performed by Lab 4 on the de-N-glycosylated NISTmAb following IdeS digestion and reduction. Direct infusion was used for analysis of the digested sample as opposed to LC separation of the IdeS fragments prior to MS analysis. The mass spectrum shows resolution of the three main components, and their deconvoluted molecular masses are in good agreement with the theoretical masses (Figure 11, Table 6) of the partially reduced LC subunit (containing one disulfide bond), partially reduced Fd' (containing one disulfide bond), and fully reduced and de-N-glycosylated scFc fragment.

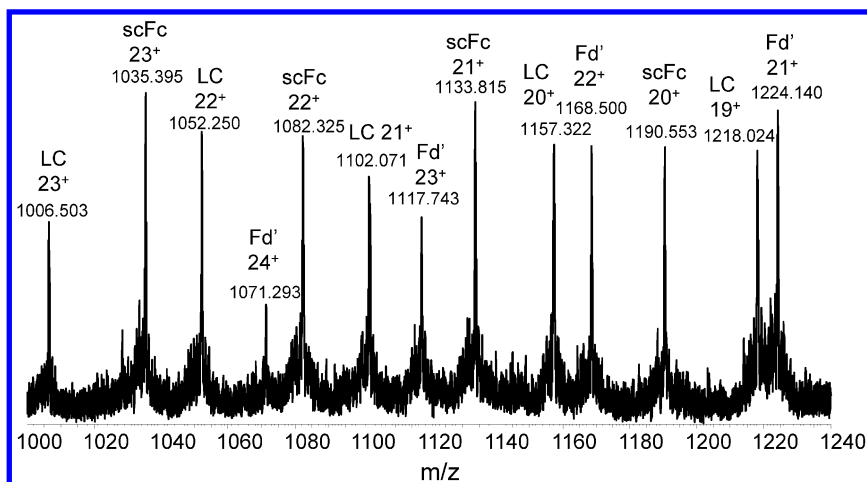


Figure 11. Raw mass spectrum of the de-N-glycosylated and IdeS digested NISTmAb by direct infusion. Mass spectrum of the digested NISTmAb shows three distinct charge state distributions corresponding to the light chain (LC), N-terminal fragment of the heavy chain (Fd'), and the C-terminal fragment of the heavy chain (scFc). The masses calculated in Xtract (Thermo) are shown in Table 6.

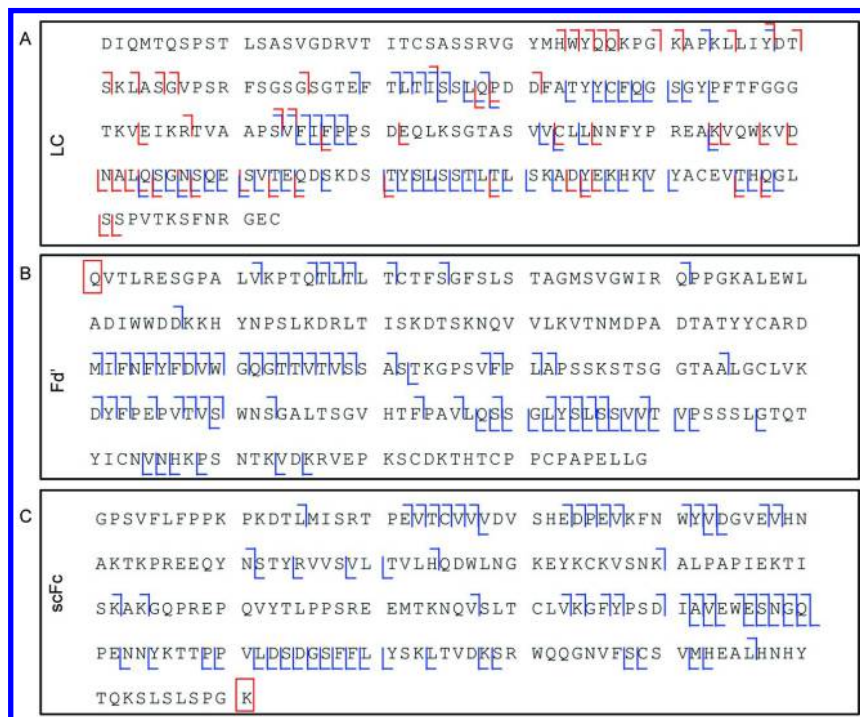
**Table 6. Middle-Down Mass Measurements Observed via Direct Infusion Following De-N-glycosylation, IdeS Digestion, and Reduction**

IdeS Fragment	Theoretical Monoisotopic Molecular Mass (Da)	Lab 4 Observed Molecular Mass (Da)
L chain (LC) subunit (one disulfide)	23111.304	23111.319
Fd' fragment (N-terminal pQ, one disulfide)	25670.807	25670.744
scFc fragment (C-terminal Lys cleavage)	23776.914	23776.212

To further analyze the middle-down fragments of the NISTmAb, fragmentation by CID, HCD, and ETD (L chain only) was performed. For each component, multiple charge states were singly isolated and fragmented at various collision energies (CID and HCD) or reagent reaction times (ETD). The MS<sup>2</sup> fragment ions generated for each IdeS fragment at the various fragmentation energies were recorded, and all fragment ions were manually added to one sequence to generate a single fragmentation map for the L chain, Fd', and scFc. The MS<sup>2</sup> fragment ions were matched to the protein sequence in ProSight Lite with a mass tolerance set to 10 ppm (Figure 12) and accounting for the known modifications of N-terminal proglutamic acid and C-terminal Lys on the scFc

portion, as well as one disulfide bond on the Fd' portion because partial reduction was observed in middle-down mass measurements. For the L chain, the map was generated for both of the N-terminal disulfides intact or complete disulfide reduction corresponding to a mixed species observed in the MS<sup>1</sup> mass spectra.

Middle-down sequencing provides an intermediate approach to protein structure analysis, providing information overlapping with and orthogonal to peptide mass mapping and intact mass analysis approaches. Because the previous limitations of available fragmentation methods are now being lifted by advancing technology, middle-down sequence coverage is approaching the levels obtained from tryptic or Lys-C digests. The ease of sample processing and data integration offered by middle-down sequencing provides a straightforward approach to confirming primary structure with the potential to identify locations of PTMs such as disulfide linkages, oxidation, and deamidation, to name a few. Middle-down sequencing platforms therefore help to bridge gaps in sequence confirmation and provide a link between the measured experimental intact mass and the peptide mass maps commonly used to verify amino acid sequence.



*Figure 12. De-N-glycosylation and IdeS-treated fragment maps generated by middle-down mass spectrometry of the (A) light (L) chain, (B) Fd', and (C) scFc. Fragments corresponding to collision-induced dissociation (CID) or higher energy collision dissociation (HCD) are shown in blue, whereas the fragments generated by electron transfer dissociation (ETD), performed on light chain only, are shown in red. Boxed residues indicate pyroglutamic acid (B) and Lys loss (C). (see color insert)*

## Discussion

Chromatographic resolution of various proteoforms is greater for IdeS fragments than for intact species, facilitating detection of gross changes in product heterogeneity using methods conducive to the high-throughput demands of a screening environment. The smaller mAb fragments generated from IdeS digestion and disulfide bond reduction are also in a molecular mass range that allows current Orbitrap and QTOF mass spectrometers to provide isotopic resolution of ESI-produced multiply charged ions. In this ~25 kDa regime, the monoisotopic ion often is not directly observed due to its relatively low natural abundance. However, the isotopic resolution allows experimental monoisotopic mass determination based on the isotope pattern for each fragment, resulting in high mass accuracy. Therefore, confident composition assignments, including localization of PTMs to more specific regions (e.g., scFc, Fd', or L chain) of the antibody are possible.

It is important to keep in mind that inferring the primary structure of a mAb from the IdeS fragments (without isolation and gas-phase fragmentation) has similar limitations as with the intact molecule. The exact elemental composition cannot be determined with absolute certainty, and without gas-phase fragmentation, the linear amino acid sequence cannot be verified. However, in the case of a biotherapeutic protein, the DNA sequence and thus the intended amino acid sequence are known. Therefore, when combined with orthogonal historical data (e.g., middle-down sequencing and/or peptide mapping as discussed below), accurate monoisotopic masses of the IdeS fragments yield high-confidence data for primary structure confirmation.

Considering the sequencing of mAbs, subunit/fragment analysis is quite advantageous because the smaller units are more amenable to common gas-phase fragmentation modes. Thus, middle-down sequencing can be useful in providing increased confidence in primary structure verification (19). However, fragmentation is often incomplete for subunits and fragments such that additional methods are commonly needed to increase sequence coverage and localize PTMs. Often, this is achieved through peptide mapping platforms that utilize more extensive enzymatic digestion prior to LC-MS and gas-phase fragmentation.

## Peptide Mapping

Peptide mapping is the fundamental technique for primary sequence determination in which a protein is cleaved into peptides using a proteolytic enzyme and analyzed using LC-MS or LC-MS/MS. Sample preparation for a “reduced” peptide map involves protein denaturation followed by reduction of the disulfide bonds, often using DTT or TCEP. The free thiols are typically alkylated (e.g., iodoacetic acid or iodoacetamide) to prevent disulfides from reforming prior to digestion (e.g., trypsin or Lys-C). A “non-reduced” peptide map involves the same sample preparation without the reduction step, allowing identification of the native positions of disulfide bonds as discussed in more detail below.

The resulting mixture of proteolytic peptides is then analyzed by LC-MS using RP-HPLC with UV detection. The chromatographic profile from the protein digest can be used as an identity assay by matching the protein's unique set of peptide peaks to those generated by a previously characterized reference material. In addition, visual inspection of the UV profile allows easy detection of gross changes in a peptide, indicating a potential amino acid modification or change in the primary sequence.

All peaks in the peptide map can be further confirmed by LC-MS based on mass, MS/MS fragmentation pattern, and retention time. Experimental molecular masses and gas-phase fragmentation patterns are matched to a predicted theoretical digest to demonstrate that the antibody contains the correct amino acid sequence as predicted from the DNA sequence. Peptide mapping is also commonly used to detect storage-induced modifications (e.g., oxidation, deamidation, clipping) and PTMs (e.g., glycosylation).

## The History of Peptide Mapping

### *The 1960s*

The initial sequencing of peptides by MS was performed using electron impact (EI) of derivatized peptides. Peptides were derivatized to amino alcohols prior to analysis because production of gas-phase ions from inherently non-volatile or low-volatile species was challenging (62–66). Researchers were able to introduce manually collected peptide fractions into the mass spectrometer for analysis and thereby construct a protein sequence. Soon the process was facilitated by the coupling of GC and MS instruments, as well as the development of computer algorithms to interpret spectral data (66). Also during this time, MS/MS began to be developed (67–69). This technique allows one to select and isolate an ion within the mass spectrometer and then impart additional energy to fragment the selected ion. These experiments were the first to achieve the “sequencing” of peptides in the gas phase using MS/MS.

### *The 1970s and 1980s*

One of the greatest challenges during early peptide mapping was obtaining precision between chromatographic experiments, instruments, and analysts. This lack of reproducibility arose from aspects of the experimental parameters, including the impurity and activity of enzymes or impurity of mobile phases. The latter is of paramount importance for the mass spectrometer as it is common for lower purity solvents to foul the system and cause high background and suppressed ionization. LC pumps used in the 1980s were also unable to produce a dependable gradient without ripple and within a small percent of the expected gradient ( $\pm 0.1\%$ ), which is critical to reproducible retention times.

The inherent weakness of electron impact-gas chromatography-mass spectrometry (EI-GC-MS) was the lack of adequate production of gas-phase ions from peptides. This problem was addressed in the early 1980s with the

introduction of fast atom bombardment (FAB) ionization (70, 71). This method allowed for the generation of gas-phase peptide ions without derivatization while also increasing the size of the peptides that could be ionized. This ionization method, considered a direct ionization method, is often cited as the one of the first soft ionization techniques. FAB ionization sources used with tandem MS instrument platforms such as the hybrid sector-quadrupole or the triple quadrupole allowed the complete sequencing of native peptides using MS fragmentation (72–75).

### *The 1990s*

Although some laboratories had access to higher performance configurations prior to the 1990s (e.g., magnetic sectors and TOFs, more stable LCs, FAB sources), most laboratories were reliant upon quadrupole instruments. At that time, many laboratories were instead collecting chromatographic peaks in combination with N-terminal sequencing (Edman degradation) to confirm the amino acid sequences of proteins. Higher resolution hybrid instruments became more available in the 1990s along with purer, more specific enzymes and more efficient HPLC systems. With these improvements, rapid and reproducible sequence confirmation of a protein was made possible.

During the latter half of the decade, widespread adoption of MS for peptide sequencing and protein identification rose with the introduction of ESI and MALDI sources. ESI was particularly useful for peptide mapping because it could be directly coupled to high-performance/pressure liquid chromatography (HPLC) instrumentation. At the same time, two mass spectrometer platforms, the QTOF and ion trap, were adopted for peptide mapping and protein analysis. Using this instrumentation, peptide mapping by LC-MS became more common.

## **Current Peptide Mapping**

Present-day peptide mapping methodologies have vastly improved in speed, reliability, and reproducibility. The most common enzymes for the peptide mapping of mAbs are the proteases trypsin and Lys-C. The popularity of trypsin, a serine protease, was first driven by its availability, relative lack of contaminants, and specificity. Trypsin cleaves proteins on the carboxyl side of the amino acid residues arginine (R) and lysine (K) except when the following residue is a proline (76), whereas Lys-C, as the name implies, cleaves proteins on the carboxyl side of lysine residues (77–81). Enzymatic products from either enzyme are thus ensured a positively charged residue on both the N-terminal (free amine) and C-terminal (R or K residue side chain) of each peptide, thereby making them ideal for positive ion mode MS. The presence of a positive charge on both termini of the peptide also facilitates tandem MS/MS identifications via detection of both b- and y-type ions. Most proteins contain numerous arginine and lysine residues and thus produce a large number of peptides amenable to LC-MS upon tryptic or Lys-C digestion.

A number of proteolytic enzymes with alternate specificities are also available for use in peptide mapping analyses. These include Glu-C, which is capable of cleaving C-terminal to glutamic acid and aspartic acid residues (82–84); chymotrypsin, which cleaves C-terminal to Phe, Trp, and Tyr residues and sometimes Met and Leu (85); and Asp-N, which specifically cleaves N-terminal to aspartic acid (86). The use of alternate enzymes is typically reserved as a means to provide complementary data to tryptic or Lys-C peptide profiles, for example, when increased levels of sequence coverage are necessary (87–90). These enzymes may be used alone or applied to peptides already generated via tryptic or Lys-C digestion. This latter strategy, using mixed enzyme digestions, is often applied when peptides produced from single enzyme digests are not amenable to MS detection or fragmentation.

### *Types of Liquid Chromatography and Mass Spectrometry Instruments*

Modern instrument configurations for peptide mapping platforms generally include a UHPLC chromatographic system coupled to a hybrid mass spectrometer. Improvements in chromatography instrumentation have allowed higher pressure delivery of mobile phase gradients to the column and, therefore, the use of more efficient chromatographic supports. Reversed phase (C18) columns are currently available with sub-2  $\mu\text{m}$  porous particles that provide very high efficiency separations. Superficially porous silica particles have also been developed that yield high-efficiency separations with lower back pressures. A variety of stationary phase chemistries are now available that provide a wider array of options from which the most optimal separation methods can be determined for any given protein digest. Although nano and capillary pumps are not uncommon, it is more customary to use a 2.1 mm ID column with a 200  $\mu\text{L}/\text{min}$  flow rate. Mobile phase A is usually near 100%  $\text{H}_2\text{O}$  with an organic modifier of 0.1% to 0.2% TFA or formic acid plus a small fraction of ACN to prevent collapse or matting of the stationary phase. Mobile phase B is typically ACN with either some low percentage of  $\text{H}_2\text{O}$  and/or isopropyl alcohol.

The typical mass spectrometer found in the biopharmaceutical laboratory has a resolving power of  $\geq 10$  K and sensitivity down to the femtomole level. The two most commonly used high-resolution MS platforms for peptide mapping are the QTOF and Orbitrap mass analyzers. Both Orbitrap and TOF-type instruments are demonstrated in the current chapter to provide excellent platforms for peptide mapping with high mass accuracy. Perhaps second only to resolution and mass accuracy, the ability of these instruments to provide short scan times is a driving force allowing the dramatic improvement in sequence coverage as seen in recent years. Fast scan time is necessary to monitor eluting peptides, isolate peptides of interest, and subsequently analyze their ion fragments all within the ever-decreasing timescale of LC platforms. In addition, the latest mass spectrometers not only provide fast spectral scanning, resulting in an increased number of peptides that can be analyzed in a given time, but are also capable of providing high-resolution, high-quality data in a rather automated fashion. High-throughput peptide identification today relies heavily on the use of complex

software algorithms capable of interpreting these large data sets with high levels of confidence.

### *Data Acquisition Modes and Sequence Coverage*

The one hurdle that must be overcome to confirm a complete amino acid sequence via LC-MS/MS methods is the production of fragments representing complete b- and y- ion series (or c- and z- series for ETD) that cover the full length of each peptide. As seen in the intact and subunit mass analysis experiments discussed earlier, at times the composition of peptides and proteins may be inferred by mass alone. For peptides less than 400 Daltons, one might be able to determine the constituent amino acids (though not the order) based solely on mass. With increasing peptide size, however, the number of theoretical amino acid composition permutations isobaric to an observed mass will also increase (91, 92). Therefore, obtaining comprehensive peptide fragmentation patterns that rule out alternative identifications is the key to complete and confident amino acid sequencing. Currently, a number of methods are commonly available for peptide fragmentation, including threshold dissociation techniques like CID or HCD, or radical-based mechanisms such as ETD or ECD.

The criteria by which ions are selected for fragmentation during an automated analysis must also be considered. In a Data Independent Analysis (DIA) approach, many (or all) ions eluting at a given time are simultaneously fragmented. This global fragmentation approach can provide excellent protein coverage when combined with an optimized LC gradient that separates samples well enough so that only a limited number of peptides coelute at a given time. This ensures that the resulting MS/MS spectra are not too complex for confident interpretation. A second option is the use of a data-dependent acquisition (DDA) mode. In this case, an initial high-resolution MS scan is performed on the eluting peptide fraction. These data are assessed in real time to determine the most intense (or top N most intense) ions eluting at the particular time. The top N number of ions are then sequentially selected for individual fragmentation and the resulting ions measured at either high or low resolution, depending on the capabilities of the hybrid instrument used. Regardless of the ion selection mode or fragmentation method chosen, each MS/MS scan is later assessed (see the Automated Data Analysis Section below) for a fragmentation pattern that will elucidate the sequence of constituent amino acids. In general, DDA mode analyses performed with ion trap, low-resolution MS/MS will have increased sensitivity and speed compared to high-resolution methods. Although this increases the number of peptides fragmented for sequencing in a given amount of time, it slightly reduces the specificity by way of reduced mass accuracy. Conversely, because higher resolution data require longer scan times, fewer peaks are sampled in this platform. However, the increase in resolution allows for better mass accuracy, thereby providing more confident amino acid sequence assignments and peptide identifications (91, 92). Both low-resolution trap and high-resolution Orbitrap MS/MS acquisition are capable of producing satisfactory peptide maps.



Most popular search engines (e.g., Mascot, Sequest) use a combination of both accurate  $m/z$  and MS/MS fragment matching to score for peptide identification (Bioinformatics chapter/Volume 3, Chapter 7). However, sequence coverage may be reported as a composition match (peptide-level) or connectivity match (amino acid-level). Peptide-level coverage is often calculated based on matching the  $m/z$  of a predicted amino acid composition to an accurate  $m/z$  measurement of the intact peptide. In this case, the entirety of that peptide is deemed confirmed and that segment of the protein is “covered.” When composition matching is used to report sequence coverage, the percent coverage is calculated by determining the number of residues comprising each confirmed peptide per the total number of amino acids in the predicted protein sequence. It is important to note that although the number and type of residues can often be determined with a high level of confidence using MS, the sequential order of the amino acids in the peptide is inferred by matching the observed peptide mass to the theoretical mass derived from the intended sequence of the original DNA construct.

Connectivity matching is a more in-depth method of reporting sequence coverage in which each amino acid of a given peptide (and therefore the entire sequence) is considered independently. Here MS accurate mass data are combined with MS<sup>2</sup> fragmentation to provide not only compositional data but amino acid sequential connectivity information as well. For an amino acid to be considered confirmed or covered, the constituent peptide expected to contain that amino acid must be identified AND a fragment ion (e.g., y- and/or b-type ion for CID fragmentation) corresponding to the mass of that individual residue must be observed in one or more MS<sup>2</sup> scans to confirm its connectivity within the peptide sequence. Depending on the properties of the individual peptide analyzed, as well as the mass spectrometer settings and type of fragmentation employed, a complete ion series may not be detected for every peptide. Sequence coverage as determined by connectivity matching is calculated using the number of individual amino acids confirmed by MS/MS analysis as a percentage of the total number of amino acids in the predicted protein sequence.

In both peptide-level and amino acid-level coverage reporting, a variety of parameters must be considered when interpreting reported values. The acceptance criteria or metrics needed to confidently assign the MS and/or MS<sup>2</sup> data for peptide composition or amino acid sequence are not well defined. The scoring algorithms, mass accuracy tolerance settings, minimum acceptable S/N threshold levels, number and type of fragment ion(s) required for identification, and minimum number of scans in which the ion or fragment ion is identified are several examples of important factors that may affect percent coverage. It should be noted that collection of MS/MS data does not always imply that amino acid-level coverage is reported. Often protein coverage is reported at the compositional level but is only partially covered at the sequence, or connectivity, level. In this case, the MS/MS data are used to identify a given peptide to a pre-defined confidence level, and the results are reported at the peptide level.

The quantity and complexity of MS and MS/MS data generated in a peptide map can be significant. Initial platforms built for the automated analysis of peptide mapping data have been primarily leveraged from proteomic bioinformatics tools that focus on protein identification using database searching algorithms. In these experiments, sequence coverage is secondary to the number of peptides identified per protein because the greater the number of peptides identified, the higher the confidence in the identification of that protein. For biopharmaceutical protein characterization purposes, however, the importance of identifying and localizing PTMs and even single amino acid substitutions necessitates complete connectivity sequence coverage. Algorithms geared toward biopharmaceutical characterization should include a means of calculating sequence coverage at this level so that any ambiguous regions are easily identified for further evaluation. In addition, high-throughput identity testing would be facilitated by software capable of producing an annotated chromatogram mapping the identified peptides and tabulated masses to each UV peak. The end goal here would then be to automate the matching of UV chromatograms to an annotated reference trace so that any changes in the profile, even those at low levels, can be detected in an unbiased and reliable manner. A number of emerging software packages are available today geared toward such in-depth levels of protein characterization, and their various analytical strategies are discussed in the Bioinformatics chapter/Volume 3, Chapter 7.

## **Experimental Materials and Methods**

Labs 1 and 2 performed peptide mapping of the NISTmAb using a variety of enzymes and digest conditions as discussed below.

### *Lab 1*

#### *Enzyme Digestion for Peptide Mapping*

The mAb was denatured in buffer comprising 6 mol/L guanidine HCl in 1 mmol/L ethylenediaminetetraacetic acid (EDTA) and 0.25 mol/L Tris(hydroxymethyl)aminomethane/ Tris(hydroxymethyl)aminomethane HCl (Tris), pH 7.5. DTT was added to a final concentration of 10 mmol/L, and the IgG was reduced at 37 °C for 30 min. Iodoacetamide (IAM) was added to a final concentration of 20 mmol/L, and the sample was alkylated at room temperature for 30 min in the dark. Alkylation was quenched by the addition of DTT to a total, final concentration of 20 mmol/L. The sample was then buffer exchanged with solution comprising freshly made 2 mol/L urea in 0.1 mol/L Tris, pH 7.8. MS-grade trypsin, sequencing-grade chymotrypsin, or sequencing-grade Glu-C was added at a 1:35 (enzyme:IgG) mass ratio. Using a microwave hydrolysis system (Discover Proteomics model 908005, CEM Corp, Matthews, NC),

digestion proceeded in SPS mode for 30 min at 50 °C. Digests were brought to a concentration of 0.5 µg/µL with 0.1% TFA for MS analysis.

### Peptide Mapping MS Analysis

LC-MS/MS was performed using the Dionex UltiMate 3000 UHPLC system coupled to the Orbitrap Elite mass spectrometer with heated electrospray ionization source (HESI II) (Thermo Scientific, Waltham, MA). Peptides (10 µg) were loaded onto a C18 column (3 µm, 15 cm × 2.1 mm; Supelco Analytical Discovery BIO Wide Pore, Sigma-Aldrich, St. Louis, MO) via autosampler and washed for 10 min using 97% mobile phase A (volume fraction of 0.1% formic acid in water) and 3% mobile phase B (volume fraction of 0.1% formic acid in ACN) at a flow rate of 0.200 mL/min. Peptide elution was achieved over a 110 min linear gradient increasing from 3% to 45% mobile phase B. MS data were collected using DDA mode with one cycle of experiments consisting of one full MS scan followed by MS/MS of the ten most intense peaks, with dynamic exclusion enabled. CID fragmentation was performed using a normalized collision energy of 35, activation Q of 0.250 and activation time of 10 ms. MS/MS fragment ions were analyzed in the ion trap.

### Peptide Identification

MS data were interrogated using the Byonic software package (Protein Metrics, Inc., San Carlos, CA). Data were searched for peptides with a static modification of cysteine carbamidomethylation (+57.021464 Da) and variable modifications of methionine oxidation (+15.994915 Da), asparagine deamidation (+0.984016 Da), succinimide intermediate (-17.02655 Da), protein C-terminal loss of lysine (-128.094963 Da), loss of C-terminal glycine-lysine (-185.116423 Da), loss of C-terminal glycine-lysine with amidation of proline (-186.100443 Da), protein N terminal pyro-glutamate (-17.02655 Da), and lysine glycation (+162.05282 Da). Chymotrypsin cleavage sites were set at low specificity (C terminal to Phe, Trp, Tyr, Met, and Leu residues); Glu-C cleavage sites were set C-terminal to Asp and Glu. Mass tolerance was set at 5 ppm for parent ions and 0.4 Da for fragment ions. A high confidence level of  $|\text{Log Prob}| > 3$  was used for acceptance of peptide identifications obtained through Byonic searches. Glycopeptides were identified using both Byonic and manual inspection of MS/MS spectra.

## Lab 2

### Enzyme Digestion

The NISTmAb was buffer exchanged to 10 mg/mL in 100 mM Tris-HCl, pH 8.2. Prior to digestion, reduction of disulfide bonds was performed by adding 500 mM TCEP and 7 M guanidine-HCl with incubation for 1 hour at 40 °C. Alkylation of the cysteines was then performed with 225 mM sodium iodoacetate at ambient temperature for 1 h in the dark. The NISTmAb was diluted to approximately 3 mg/mL with HPLC-grade water and enzymatically cleaved with Lys-C using an enzyme:substrate mass ratio of 1:20 at 30 °C for 16–20 h. Following incubation, the reaction was quenched with TCEP and TFA and mixed well. The digest was placed at 2 °C to 8 °C for storage prior to analysis by LC-MS.

### Peptide Mapping

The digested sample (5 µg) was injected onto a Waters XBridge BEH C18, 3.5 µm, 2.1 × 150 mm column (Milford, MA) heated to 60 °C and eluted using a 2 h reversed phase gradient at a flow rate of 0.2 mL/min in which mobile phase A was 0.05% TFA in water and mobile phase B was 0.05% TFA in 95% ACN. The peptide mapping experiment was performed on a Waters Alliance HPLC with UV detection at 214 nm and 280 nm, coupled to an ultrahigh-resolution Bruker Daltonics maXis QTOF mass spectrometer (Billerica, MA). MS acquisition parameters included the capillary voltage: 4.2 kV, nebulizer gas: 1.6 bar, dry gas: 8.0 L/min, collision energy: 6 eV, and scan range  $m/z$  200 to  $m/z$  2500 were optimized for peptides on the Bruker maXis QTOF mass spectrometer.

### Peptide Identification

Post-acquisition data processing was performed using Data Analysis software from Bruker Daltonics (Billerica, MA) by matching parent ion masses to a theoretical Lys-C digest of the NISTmAb using Sequence Editor software. Observed masses for peptides were determined from the most abundant multiply charged parent ion in the mass spectrum of each peak and matched to its monoisotopic theoretical mass with an allowed mass error of  $\leq 2$  ppm.

### Non-Reduced Peptide Map

The NISTmAb was diluted to 2 mg/mL with 100 mM Tris, pH 8.2 buffer. A 100 µg aliquot of the mAb was pre-alkylated and denatured with 4 µL of 100 mM IAM (prepared fresh) and 136 µL of denaturation buffer (7 M guanidine-HCl, 100 mM Tris, pH 8.2) for 1 h at 40 °C and protected from light. The protease Lys-C was added at a ratio of 1:10 (enzyme:substrate) and incubated for 6 h.

Prior to loading on the HPLC auto-sampler, which was held at 4 °C, the sample was acidified with a small amount of TFA. The Lys-C digestion mixture (5 µg) was injected onto a column held at 60 °C and peptides eluted with the same gradient, mobile phases, flow rate, and mass spectrometer settings as described above for the reduced and alkylated peptide mapping experiment. In addition, post-acquisition data processing was performed using Data Analysis software from Bruker Daltonics (Billerica, MA) and by matching the disulfide-linked peptide masses to a theoretical, non-reduced Lys-C digest of the NISTmAb using Sequence Editor software.

## Results and Discussion

A representative Lys-C peptide map (Lab 2) analyzed by LC-MS is shown in Figure 13, with corresponding observed mass values for each peak listed in Table 7. This enzyme has been purified and characterized from *Achromobacter lyticus* M497-1, *Lysobacter enzymogenes*, and *Pseudomonas aeruginosa* and cleaves C-terminal to lysine residues, including those followed by proline (77–81). Studies have shown Lys-C to be active within a fairly broad pH range (approximately pH 8 to pH 10), typically exhibiting higher enzymatic activity than bovine trypsin and retaining activity in the presence of chaotropic agents such as urea or guanidine that are often needed to effectively denature mAbs (77–81, 93, 94). In addition to these attributes, Lys-C digestion generates fewer peptides than trypsin, lending itself to simpler UV chromatograms with fewer peaks. However, because there is no cleavage at Arg residues, Lys-C digests may contain a greater number of large peptides than tryptic digests, which may be intractable to complete amino acid sequence coverage by MS/MS.

In the current example, major and minor peaks for the NISTmAb were identified by online LC/MS. The observed masses for each peak in Figure 13 (Table 7) were consistent with the expected Lys-C peptides for this antibody as predicted from a theoretical digest. Several small peptides were not detected in the peptide map, presumably due to insufficient interaction with the chromatographic column (C18) stationary phase. Peptides corresponding to 97% of the L chain and 96% of the H chain amino acid sequences were detected, yielding nearly complete compositional sequence confirmation of the NISTmAb primary structure. High mass accuracy MS is critical for peptide assignments. In the current example, all peptides were identified within a mass error of  $\leq 2$  ppm, thereby providing confident identification. Peptide maps used for product characterization typically employ relatively long chromatographic gradients (>1 h) with the goals of obtaining 100% sequence coverage and identifying any detectable modified peptides. During this characterization stage, low sample throughput and high-efficiency separations are necessary for comprehensive sequence coverage.

Table 7 and Figure 13 are excellent examples of the type of characterization one may see in a typical licensure application to support the validity of a peptide map for identity testing. Such a table may eventually include retention times, relative peak areas, and related performance specifications for peaks deemed to correspond to critical quality attributes. Similar tables were generated for each of the peptide maps collected during preparation of this chapter, each of which will

be summarized below to focus primarily on attributes specific to primary structure confirmation. The reader is directed to the PTMs chapter/Volume 2, Chapter 3 for more detailed information on low-level modifications and methods for their identification.

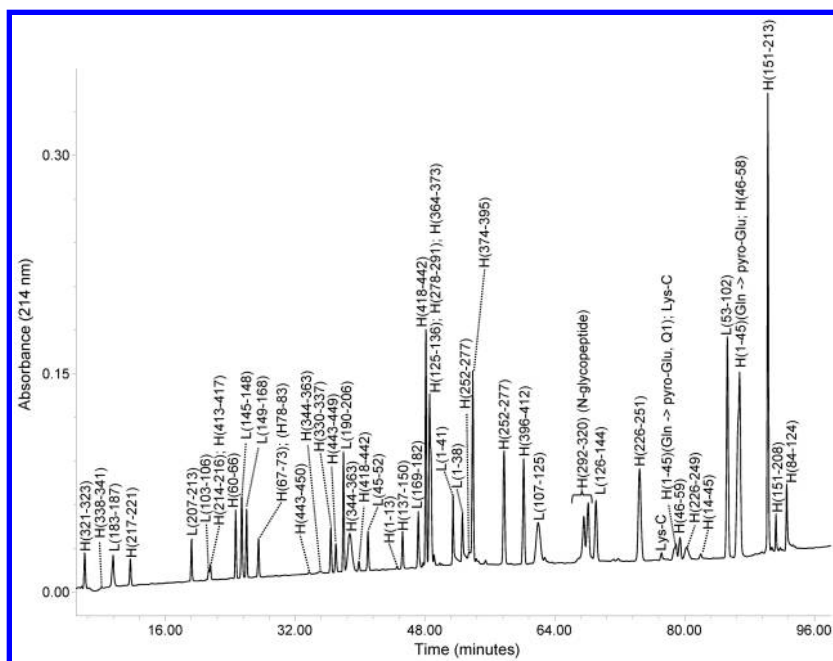


Figure 13. Lys-C peptide mapping of NISTmAb by liquid chromatography-UV-mass spectrometry (LC-UV-MS).

The first column in Table 8 contains a summary of the Lys-C peptide mapping data presented in Table 7 and Figure 13. The table demonstrates that a very high peptide-level coverage was attained and further confirms assertions about modifications determined through intact and middle-down analysis. For example, the peptide map confirms the predominance of glutamine conversion to pyroglutamic acid on the N-terminus of the H chain as well as minor levels of C-terminal lysine present. Localization of these modifications to specific peptides further verifies their terminal location.

The Lys-C peptide map results also indicate the presence of only one N-glycosylation site, further verifying that N-glycosylation is not present on the Fab region of the NISTmAb. The consensus sequence (NST) was identified in the Fc region with a variety of N-glycan compositions (e.g., G0F, G1F) expected for an IgG1 mAb produced in murine-derived cells. The major and minor N-glycoforms identified are in direct agreement with the top- and middle-down results, as well as the released N-glycan results described in the Glycosylation chapter/Volume 2, Chapter 4. Of particular note are the +Hex species listed in Table 8, which suggest that  $\alpha$ -Gal-containing N-glycans may be present in the NISTmAb. This is further supported by data presented in the Glycosylation chapter/Volume 2, Chapter 4.

**Table 7. Theoretical and Observed Peptides for NISTmAb Lys-C Peptide Map**

***Light (L) Chain***

<i>Peptide<sup>a</sup></i>	<i>Theoretical Mass (Da)<sup>b</sup></i>	<i>Observed Mass (Da)<sup>c</sup></i>	<i>Confirmed Sequence</i>
L(1-38)	4275.9987	4275.9927	DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQK
L(1-41)	4558.1680	4558.1644	DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGK
L(39-41)	300.1798	Detected as L(1-41)	PGK
L(42-44)	314.1954	Not Detected	APK
L(45-52)	951.5277	951.5276	LLIYDTSK
L(53-102)	5250.4081	5250.4038	LASGVPSRFSGSGSGTEFTLTISLQPDDEFATYYCFQGSQYPT-FGGGTKV
L(103-106)	487.3006	487.3001	VEIK
L(107-125)	2101.1208	2101.1190	RTVAAPSVFIFPPSDEQLK
L(126-144)	2126.0466	2126.0448	SGTASVVCLLNNFYPREAK
L(145-148)	559.3118	559.3120	VQWK
L(149-168)	2134.9614	2134.9598	VDNALQSGNSQESVTEQDSK
L(169-182)	1501.7512	1501.7504	DSTYLSSTLTLSK
L(183-187)	624.2755	624.2754	ADYEK
L(188-189)	283.1644	Not Detected	HK

**Light (L) Chain**

<i>Peptide<sup>a</sup></i>	<i>Theoretical Mass (Da)<sup>b</sup></i>	<i>Observed Mass (Da)<sup>c</sup></i>	<i>Confirmed Sequence</i>
L(190-206)	1875.9037	1875.9030	VYACEVTHQGLSSPVTK
L(207-213)	869.3338	869.3334	SFNRGEC

**Heavy (H) Chain**

<i>Peptide<sup>a</sup></i>	<i>Theoretical Mass (Da)<sup>b</sup></i>	<i>Observed Mass (Da)<sup>c</sup></i>	<i>Confirmed Sequence</i>
H(1-13)	1379.7770	1379.7764	qVTLRESGPALVK (Gln → pyro-Glu, Q1)
H(1-45)	4788.4620	4788.4581	qVTLRESGPALVKPTQTLTLCTFSGFSLSTAGMSVG-WIRQPPGK (Gln → pyro-Glu, Q1)
H(14-45)	3426.6952	3426.6909	PTQTLTLCTFSGFSLSTAGMSVGWIRQPPGK
H(46-58)	1659.7933	1659.7932	ALEWLADIWDDK
H(59-59)	146.1050	Detected as H(46-59)	K
H(46-59)	1787.8883	1787.8864	ALEWLADIWDDKK
H(60-66)	857.4396	857.4394	HYNPSLK
H(67-73)	831.4814	831.4812	DRLTISK
H(74-77)	449.2122	Not Detected	DTSK
H(78-83)	699.4279	699.4274	NQVVVK
H(84-124)	4631.0607	4631.0577	VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTK
H(125-136)	1185.6394	1185.6390	GPSVFPLAPSSK

*Continued on next page.*



**Table 7. (Continued). Theoretical and Observed Peptides for NISTmAb Lys-C Peptide Map****Heavy (H) Chain**

<i>Peptide<sup>a</sup></i>	<i>Theoretical Mass (Da)<sup>b</sup></i>	<i>Observed Mass (Da)<sup>c</sup></i>	<i>Confirmed Sequence</i>
H(137-150)	1321.6548	1321.6544	STSGGTAALGCLVK
H(151-208)	6186.0205	6186.0128	DYFPEPVTVSWNSGALTSGVHTFPAVLQSS- GLYSLSSVVTVPSSSLGTQTYICNVNHNK
H(151-213)	6713.2910	6713.2844	DYFPEPVTVSWNSGALTSGVHTFPAVLQSS- GLYSLSSVVTVPSSSLGTQTYICNVNHNKPSNTK
H(209-213)	545.2809	Detected as H(151-213)	PSNTK
H(214-216)	360.2009	360.2002	VDK
H(217-221)	627.3704	627.3704	RVEPK
H(222-225)	509.1792	Not Detected	SCDK
H(226-249)	2620.2706	2620.2678	THTCPPCPAPELLGGPSVFLFPPK
H(226-251)	2845.4183	2845.4163	THTCPPCPAPELLGGPSVFLFPPKPK
H(250-251)	243.1583	Detected as H(226-251)	PK
H(252-277)	2955.4190	2955.4191	DTLMISRTPVTCVVVDVSHEDPEVK
H(278-291)	1676.7947	1676.7916	FNWYVDGVEVHNAK
H(292-293)	247.1532	Detected as H(292-320)	TK
H(292-320) <sup>d</sup>	4904.3240	4904.3208	TKPREEQYNSTYRVVSVLTVLHQDWLNGK + [G0F]
H(321-323)	438.2114	438.2113	EYK

**Heavy (H) Chain**

<i>Peptide<sup>a</sup></i>	<i>Theoretical Mass (Da)<sup>b</sup></i>	<i>Observed Mass (Da)<sup>c</sup></i>	<i>Confirmed Sequence</i>
H(324-325)	307.1202	Not Detected	CK
H(326-329)	446.2489	Not Detected	VSNK
H(330-337)	837.4960	837.4958	ALPAPIEK
H(338-341)	447.2693	447.2686	TISK
H(342-343)	217.1426	Not Detected	AK
H(344-363)	2342.1689	2342.1666	GQPREPQVYTLPPSREEMTK
H(364-373)	1161.6064	1161.6048	NQVSLTCLVK
H(374-395)	2543.1241	2543.1216	GFYPSDIAVEWESNGQPENNYK
H(396-412)	1872.9145	1872.9140	TTPPVLDSDGSFFLYSK
H(413-417)	574.3326	574.3324	LTVDK
H(418-442)	3044.3770	3044.3712	SRWQQGNVFSCSVMHEALHNHYTQK
H(443-449) <sup>e</sup>	659.3490	659.3485	SLSLSPG

<sup>a</sup> Samples were reduced, alkylated, and digested with Lys-C. Peptides were separated by reversed-phase-high-performance liquid chromatography (RP-HPLC) with UV detection at 214 nm and mass spectrometry (MS) performed by ultrahigh-resolution-electrospray ionization-quadrupole time-of-flight mass spectrometry (UHR-ESI-QTOF MS). “L” represents light chain, “H” represents heavy chain. Lys-C residues are numbered sequentially starting at the N-terminus of the subunit. <sup>b</sup> Theoretical monoisotopic masses following alkylation with sodium iodoacetate were calculated using the Sequence Editor application in Data Analysis (Bruker Daltonics, version 4.0). <sup>c</sup> Observed monoisotopic masses were determined from the most abundant multiply charged ion in the mass spectrum with Biotoools (Bruker Daltonics). The relative errors between theoretical and observed mass values were less than 2 ppm, which is consistent with the performance specifications of the Bruker Daltonics maXis mass spectrometer for peptide analysis. <sup>d</sup> Additional N-glycoforms are described in Table 8. <sup>e</sup> A peptide containing C-terminal lysine was also detected at a minor level.

Although not discussed in detail in the current chapter, peptide mapping also confirmed the presence of glycosylated and oxidized residues in the NISTmAb as indicated by subunit analysis and presented in more detail in the PTMs chapter/Volume 2, Chapter 3. Collectively, all of the peptide identifications reported are consistent with those predicted earlier by the intact and subunit analysis data. This provides a good example of the utility of orthogonal assays to corroborate results.

During product and method development, multiple peptide mapping strategies may be explored to find the enzyme/column/MS combination that yields the highest level of information for a specific product. For example, the digest efficiency of prospective proteases are one variable to consider. As is evident from Table 7 and is common for any protease, a number of peptides with missed cleavages (e.g., H(151-213)) may be detected in a peptide map as opposed to the fully processed peptides. Missed cleavages can be the result of many factors, including inaccessibility of the protease to cleavage sites and/or reduced enzyme activity due to the presence of certain amino acids in the sequence (95). In addition, individual proteins may contain closely spaced cleavage sites generating very small peptides (2 to 4 amino acids). Such small peptides may not bind to a C18 column or ionize efficiently in the MS source due to limited protonation sites, thereby reducing the likelihood of detection and consequently the level of protein sequence coverage. At times, missed cleavages can improve sequence coverage by providing sequence information in regions that may otherwise produce small peptides. Alternatively, orthogonal proteases whose cleavage specificity would generate larger peptides in those regions may be used to provide complete sequence coverage. An example of this combinatorial approach for bottom-up peptide mapping was performed in Lab 1 using data from separate digestions with different enzymes (Trypsin, Glu-C, and chymotrypsin). Results from the trypsin and Glu-C peptide maps are summarized in Table 8 and Figure 14.

In assessing each method individually, all four peptide maps contained in this chapter yielded significant peptide-level coverage of the NISTmAb. Lys-C, trypsin, and chymotrypsin (data not shown) all resulted in >90% sequence coverage of the H and L chains. Figure 14 presents the summation of peptide-level coverage of the NISTmAb collected in Labs 1 and 2 using the different enzymatic digests. Taken together, these multiple peptide maps provide 100% primary amino acid sequence confirmation of the NISTmAb and again highlight the value of using orthogonal methods.

**Table 8. Summary of NISTmAb Peptide Mapping by Liquid Chromatography-Mass Spectrometry (LC-MS) and by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) Using Multiple Enzymes**

Characteristics	<i>Lys-C (LC-MS)</i>	<i>Trypsin (LC-MS/MS)</i>	<i>Glu-C (LC-MS/MS)</i>
<b>Amino acid sequence confirmation</b>	<p>Light (L) chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 97%: detected peptides represent 208 out of 213 residues</li> <li>• MS/MS not conducted</li> </ul> <p>Heavy (H) chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 96%: detected peptides represent 433 out of 450 residues</li> </ul>	<p>L chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 96%: detected peptides represent 205 out of 213 residues</li> <li>• MS/MS amino acid-level coverage 94%<sup>a</sup></li> </ul> <p>H chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 99%: detected peptides represent 444 out of 450 residues</li> <li>• MS/MS amino acid-level coverage 99%<sup>a</sup></li> </ul>	<p>L chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 100%: detected peptides represent 213 out of 213 residues</li> <li>• MS/MS amino acid-level coverage 95%<sup>a</sup></li> </ul> <p>H chain peptide-level coverage:</p> <ul style="list-style-type: none"> <li>• 62%: detected peptides represent 277 out of 450 residues</li> <li>• MS/MS amino acid-level coverage 56%<sup>a</sup></li> </ul>
<b>Terminal amino acid sequence</b>	<p>N-terminal peptides detected:</p> <p>H(1-13): (Gln → pQ, Q1)</p> <p>L(1-38)</p> <p>C-terminal peptides detected:</p> <p>H(443-449) <i>des</i>-lysine form</p> <p>H(443-450) detected at minor level</p>	<p>N-terminal peptides detected:</p> <p>H(1-5): (Gln → pQ, Q1), H(1-45): (Gln → pQ, Q1)</p> <p>L(1-18), L(1-28)</p> <p>C-terminal peptides detected:</p> <p>H(443-449) <i>des</i>-lysine form</p>	<p>N-terminal peptides detected:</p> <p>H(1-6): (Gln → pQ, Q1)</p> <p>L(1-49)</p> <p>C-terminal peptides detected:</p> <p>H(434-449) <i>des</i>-lysine form</p> <p>H(434-450) detected at minor level</p>

*Continued on next page.*

**Table 8. (Continued). Summary of NISTmAb Peptide Mapping by Liquid Chromatography-Mass Spectrometry (LC-MS) and by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) Using Multiple Enzymes**

<i>Characteristics</i>	<i>Lys-C (LC-MS)</i>	<i>Trypsin (LC-MS/MS)</i>	<i>Glu-C (LC-MS/MS)</i>
	L(207-213)	H(443-450) detected at minor level  L(207-213)	L(195-213)
<b>N-linked glycosylation</b>	<p>N-linked glycosylation site confirmed at Asn300 in the H(292-320) peptide (major) and the H(294-320) peptide (trace)</p> <p>N-glycans detected (identities and relative abundances):</p> <p>Major: G0F, G1F</p> <p>Minor: G2F</p> <p>Trace: G0, G1; Man5; G0F-[GlcNAc], G1F-[GlcNAc], G2F-[GlcNAc]; G1F+[GlcNAc], G2F+[GlcNAc], G3F+[GlcNAc], G2F+[Hex], and G2F+2[Hex]</p>	<p>N-linked glycosylation site confirmed at Asn300 in H(292-304) glycopeptides (major) and H(296-304) glycopeptides (trace)</p> <p>N-glycans detected (identities and relative abundances):</p> <p>Major: G0F, G1F</p> <p>Minor: G2F, G0F-[HexNAc]</p> <p>Trace: Man5, G1F-[HexNAc], G2F+[Hex], G2F-[HexNAc], G2F+[Hex][Hex]</p> <p>Trace levels of aglycosylated H(292-304) and H(296-304) peptides detected</p>	<p>N-linked glycosylation site confirmed at Asn300 in H(298-321) glycopeptides</p> <p>N-linked glycopeptides detected (identities and relative abundances):</p> <p>Major: G0F, G1F</p> <p>Minor: G2F, G0F-[HexNAc]</p> <p>Aglycosylated peptide not detected</p>

<sup>a</sup> Amino acid-level coverage calculated based on observation of an MS/MS b- or y-type ion corresponding to the specific amino acid as identified using software parameters described in Lab 1 materials and methods. MS/MS fragmentation spectra were used to support peptide identification as described in Lab 1 materials and methods.

From a complete peptide sequencing standpoint, the tryptic peptide map provides the highest level of sequence coverage (96% of the L chain and 99% of the H chain) for the NISTmAb using a single enzyme. Trypsin is perhaps the most commonly used protease for bottom-up methods for a number of reasons. This enzyme cleaves C-terminal to Lys and Arg residues and generates peptides well suited for mass spectrometric analysis for several reasons. Size is one such attribute. An average tryptic peptide is composed of only 12 to 26 amino acids, which is typically large enough for suitable chromatographic retention but not too large for informative MS/MS fragmentation (96). Another advantage of tryptic peptides is their charge. Because tryptic peptides terminate in positively charged (basic) Lys and Arg residues, they are readily detected at the MS level, and their chemical properties promote highly informative and easily interpreted y-ion series fragmentation by low-energy CID MS/MS (76).

Glu-C is another enzyme that would typically find use only under unique circumstances. This protease was originally isolated from *Staphylococcus aureus*, and its substrate specificity is buffer-dependent. In buffers containing sodium or potassium phosphate, Glu-C cleaves amide bonds C-terminal to both Glu and Asp residues, and Lab 1 observed cleavage for both residues in urea/Tris digestion buffer in the current experiment. In the presence of ammonium bicarbonate or acetate, however, cleavage after Asp residues is highly preferred (82–84, 97). Although Glu-C digestion resulted in 100% sequence coverage for the L chain of the NISTmAb peptide map, only 62% sequence coverage was obtained for the H chain, which was likely due to the infrequency of cleavage sites in this subunit, creating a prevalence of large, unwieldy peptides.

Chymotrypsin is not used routinely for peptide mapping but may have useful applications in specific cases, such as when complete sequence coverage cannot be achieved using trypsin or Lys-C. This enzyme has lower specificity than trypsin or Lys-C, with C-terminal cleavage occurring predominantly after Phe, Trp, and Tyr and sometimes after Met and Leu residues (85), albeit with lower frequency. This higher number of possible cleavage sites can give rise to smaller peptides than those produced by trypsin or Lys-C. The Lab 1 results obtained for chymotryptic digestion (data not shown) yielded 89% and 91% sequence coverage of the L and H chains, respectively.

The ruling principle for choosing the most appropriate peptide mapping platform is to determine the methodology that will elucidate the broadest range of product attributes. This determination should not be based on protein sequence coverage alone but should also take into consideration which methodologies can be relied upon to resolve and quantify critical quality attributes and identify potential product-related impurities even at low levels. Although a variety of additional factors are involved in such a decision, it seems that in the case of the NISTmAb, a combination of Lys-C and trypsin peptide maps would provide optimum sequence coverage (collective) while allowing very selective and sensitive observation of modifications such as terminal and N-glycosylated peptides. It should be noted that even with both peptide maps, the short H(324-329) peptide is not detected with these enzymes. If this peptide were known for susceptibility to certain critical PTMs or detrimental modifications, it

may then be important to consider using a different enzyme or combination of enzymes.

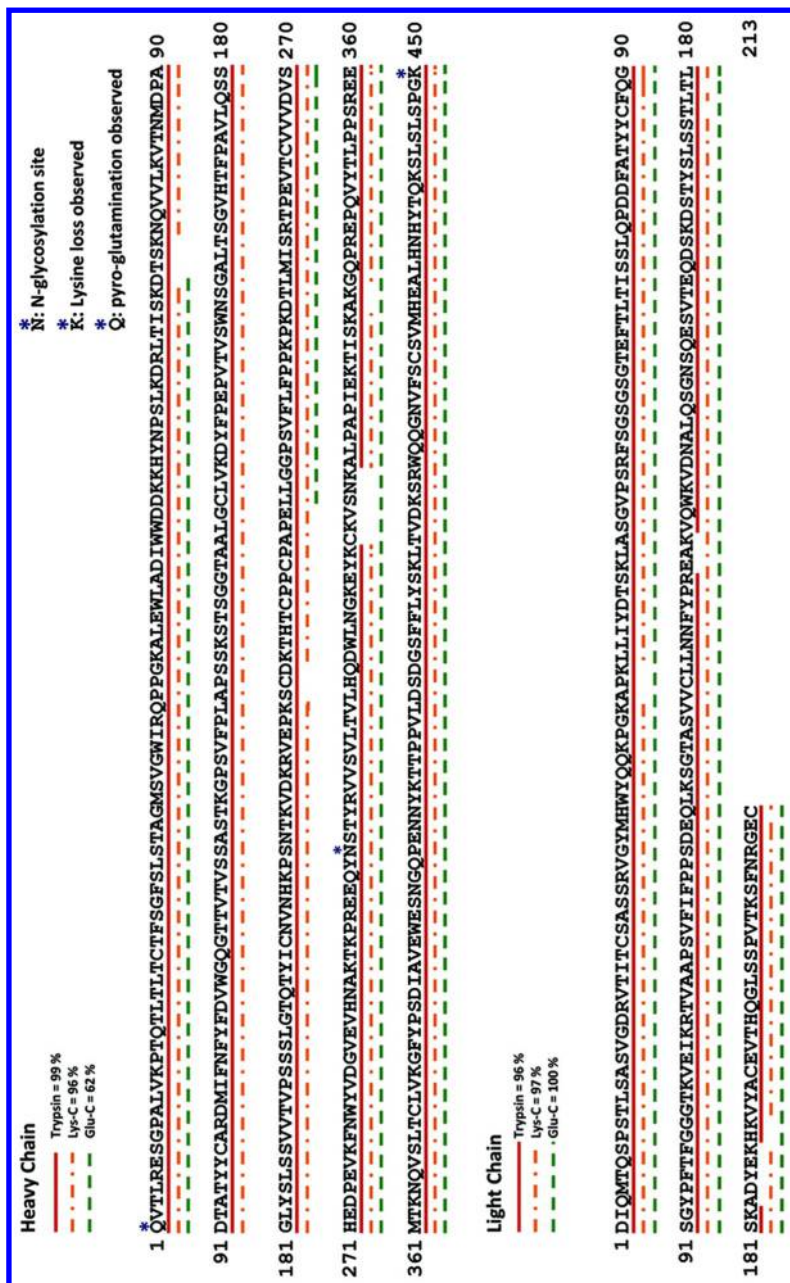


Figure 14. NISTmAb Peptide Coverage. Tryptic (solid underline), Lys-C (dot-dash underline), and Glu-C (dashed underline) peptide-level coverage of the NISTmAb as analyzed by liquid chromatography-mass spectrometry (LC-MS) and/or tandem mass spectrometry (MS/MS). (see color insert)

Although enzyme choice can heavily influence the level of protein sequence coverage obtained via peptide mapping, it is not the only parameter to require optimization. Sequence coverage is also subject to the manner in which the data are processed. The results derived from any LC-MS analysis will undoubtedly change if data processing methods or scoring thresholds are significantly altered, even on the same data set. For example, setting more stringent cutoff values that a data point must meet in order to be considered valid can reduce the number of identified peptides and therefore protein sequence coverage. These parameters include S/N thresholds and mass accuracy tolerance.

In addition to selecting parameters that will ensure accurate results for the sample as a whole, some aspects of an optimized peptide mapping method will include steps or settings designed to focus on a specific attribute. For example, when a particular amino acid residue or its potential modifications are of critical importance to product quality, the MS/MS fragmentation mode, energies, and so forth for its peptide of residence can be specifically optimized to achieve product ions that will provide unambiguous determination of the status of that residue. The same can be said for sample preparation steps, wherein specific steps may be added or modified to target a particular attribute of the mAb such as the identification of glycosylation site(s) (Glycosylation chapter/Volume 2, Chapter 4) or disulfide bonding pattern as discussed below.

### *Non-Reduced Peptide Mapping*

Disulfide bond linkages are essential to the higher order structure of an antibody, and consequently, its biological activity (98). Each different IgG-type (IgG1, IgG2, IgG3, and IgG4) has its own individual set of predicted and unique disulfide bond linkages between the sulfurs of cysteine residues (99–106). IgG2-type mAbs have more complex disulfide bond linkages due to a mixture of isoforms, designated as IgG2-A, -B, and -A/B (49, 50, 107, 108). It is important to have correct disulfide bond linkage formation for proper protein folding, structure, and function. To examine the disulfide bond linkages of an antibody, non-reduced peptide mapping by LC-MS has emerged as a powerful tool (109–111). In the current non-reduced peptide map, the antibody is typically denatured, pre-alkylated, and enzymatically cleaved to generate peptides. Disulfide-containing peptides are then chromatographically separated and identified by LC/MS.

Typical to IgG1 antibodies, the NISTmAb contains 32 cysteine residues in the intact, four-chain protein and is predicted to form sixteen disulfide bonds. Consistent with the structural homology of an IgG1 mAb, these linkages include four inter-chain and twelve intra-chain disulfide linkages (106). The non-reduced peptide map provided confirmation that the cysteine residues of the NISTmAb are correctly paired as predicted. This method uses a pre-alkylation step and digestion at neutral pH with the protease Lys-C to prevent any disulfide scrambling. The predicted disulfide-bonded peptides resulting from a Lys-C digest are shown in the schematic Figure 1 with the paired cysteine residues numbered.



The resulting disulfide-bonded proteolytic peptides were identified by Lab 2 using LC/MS. Only eight unique disulfide-linked peptides are expected, and all eight disulfide-linked peptides were confirmed as shown in Table 9. Use of the non-reduced peptide map demonstrates that the NISTmAb contains the correct disulfide linkages for an IgG1 molecule with no detectable mispaired peptides.

Further site-specific verification using fragmentation techniques can also be performed in this type of analysis. For example, ETD will preferentially cleave cysteine disulfide bonds rather than the backbone of the peptides they connect, resulting in an MS<sup>2</sup> spectrum consisting of the individual peptides originally associated via the disulfide linkage. A second round of fragmentation using HCD or CID may then be performed on the dissociated peptides in an MS<sup>3</sup> experiment to obtain and confirm sequence information of the peptides (112, 113).

### *Peptide Maps for Identity Testing*

Peptide mapping is a robust strategy that provides site-specific information on protein sequence and is therefore a highly useful tool for identity testing. Most often, MS and MS/MS data are used during initial characterization phases to determine peptide identity at each retention time. In subsequent analyses, for example, release testing of drug substance, it is often then inferred that peaks conforming to expected retention times and intensities are of the same composition as those characterized within the same peaks of the initial, fully characterized reference map.

To verify that a peptide mapping method is capable of confirming the identity of the desired drug substance via MS and MS/MS, the method should also be thoroughly tested for its ability to differentiate the product from others because specificity is the most important aspect of an identity assay. To accomplish this, the peptide map may be challenged with stressed product, synthetic peptides, highly similar molecules, and highly unrelated molecules (114). In addition to meeting general criteria such as the absence of new peaks, a visual inspection of retention times and peak area intensities must meet predetermined criteria to confirm the identity of the sample, and this may be done only after a very high-level understanding of method performance and limitations has been established. The use of MS early in the development of the peptide map has become commonplace because such historical data will become useful when significant new or deviant peaks do appear. Cumulative method knowledge improves the likelihood that unexpected peaks have been previously observed during qualification with challenge material and thus can be readily identified. In addition to selectivity, a validated peptide map is often evaluated for precision, as well as robustness, in respect to expected variations in sample preparation and/or analytical operations.

**Table 9. Theoretical and Observed Disulfide-Linked Peptides for NISTmAb**

<i>Disulfide-Linked Peptides<sup>a</sup></i>	<i>Confirmed Disulfide Linkages</i>	<i>Theoretical Mass (Da)</i>	<i>Observed Mass (Da)<sup>b</sup></i>
<b>Intra-chain</b>			
H(1-45)-S-S-H(84-124)	Cys22-Cys97	9301.5025	9301.4958
H(137-150)-S-S-H(151-208)	Cys147-Cys203	7389.6490	7389.6340
H(252-277)-S-S-H(324-325)	Cys264-Cys324	3144.5141	3144.5127
H(364-373)-S-S-H(418-442)	Cys370-Cys428	4087.9567	4087.9524
L(1-41)-S-S-L(53-102)	Cys23-Cys87	9690.5500	9690.5937
L(126-144)-S-S-L(190-206)	Cys133-Cys193	3883.9230	3883.9206
<b>Inter-chain</b>			
L(207-213)-S-S-H(222-225)	Cys213-Cys223	1260.4863	1260.4856
H(226-251)-S=S-H(226-251)	Cys229-Cys229 Cys232-Cys232	5454.7834	5454.7776

<sup>a</sup> “H” denotes the heavy chain, and “L” denotes the light chain; “-” represents one disulfide bond, and “=” represents two disulfide bonds, the putative locations of which are indicated in the second column. <sup>b</sup> Observed monoisotopic masses were calculated from the most abundant multiply charged ion in the mass spectrum. All observed masses agree with theoretical masses to within 5 ppm, which is consistent with the instrument performance specification of the Bruker maXis mass spectrometer for peptide analysis.

## Discussion

The current data set serves as a representative example of how peptide mapping can be used to characterize the primary amino acid sequence of a therapeutic protein. Bottom-up peptide mapping provides a wealth of information on product identity and associated PTMs. By using a peptide mapping approach, a number of amino acids with PTMs were detected in the NISTmAb that could not be conclusively verified by intact and/or subunit analysis methods. The peptide mapping discussion in the current chapter focused primarily on N-terminal, C-terminal, and N-glycosylation PTMs; however, additional discussion of peptide mapping for verifying other PTMs, such as deamidation and methionine oxidation, are discussed in the PTMs chapter/Volume 2, Chapter 3.

As demonstrated above, the practice of peptide mapping is not fail-safe and must have adequate checks and balances (e.g., system suitability, in-house reference standard, analyst training, product knowledge). One must again keep in mind, as discussed in the PTMs chapter/Volume 2, Chapter 3, sample processing may result in the introduction of method-induced modifications (e.g., oxidation, deamidation). The development of an optimized peptide mapping strategy would incorporate stringent evaluation of potential method-induced artifacts and a final robustness evaluation during method qualification. Historical knowledge of the method and product may provide troubleshooting guidelines when results fall outside of known acceptance criteria, and orthogonal methodologies such as intact and subunit analyses can be useful to differentiate product-related artifacts from method-related artifacts. Completely orthogonal techniques, such as separation methods and/or biophysical techniques, may also be utilized when a specific attribute is thought to be affected.

### Analysis Limitations—Reference Material Complexity

This exercise of confirming that the NISTmAb is what it is purported to be must also include a statement of what cannot be ascertained due to subtle aspects of the reference material complexity. That is to say, is each and every chromatographic peak, spectral ion, and collective piece of data fully accounted for and in agreement? When considering the presented data, has every aspect of potential reference material heterogeneity been observed? Although this is the ultimate pursuit, developing technologies and further investigation will undoubtedly reveal new product attributes. It is for this very reason that a widely available reference material characterized to the highest industry standards is necessary. When such questions do arise, a class-specific molecule that may have similar characteristics to the product in question can be widely deployed to evaluate new technology.

A related notion should also be mentioned in regard to product purity. There currently are no absolute means to unequivocally demonstrate the absence of aberrant proteins (i.e., non-intended therapeutic proteoforms) or process-related proteins (host cell proteins [HCPs]). Historically, due to the unavailability of a suitable methodology to certify a material as 100% free of impurities, whether

product-related or otherwise, biopharmaceutical companies have relied on a toolbox approach, using many orthogonal techniques to demonstrate that the drug substance or product does not contain impurities over a specified level. Depending on the nature of other potential process-related impurities, (endotoxins, HCPs, adventitious agents, extractables, and leachables) the level to which a company must demonstrate the removal of these from clinical material is dependent upon the known safety profile of individual impurities. The availability of a reference material of known purity will allow spike/recovery testing for the limits of detection for such impurities, as well as the advancement of technological and methodological means for ensuring product purity.

Each of these discussions becomes relevant to the confirmation of the primary amino acid sequence because LC-MS data sets may contain initially unidentified peaks. This is especially true when new protocols, instrumentation, raw materials, and so forth are implemented into a previously accepted strategy. Changes to the traditional platform may reveal previously undetected species such as the signal peptide, C-terminal amidation, sequence variants, PTMs, or other chemical modifications. Rather than immediately dismissing the new technology change, evaluating such deviations should be initially viewed as alternative outcomes requiring further research using all available components of the analytical toolbox. It is for this reason that the following chapters describe in detail targeted MS-based assays and data assessment for PTMs, sequence variants, and potential impurities. Orthogonal methods available in the toolbox for elucidation of the cause of such deviations are also included.

Finally, it is thought that the reference material described herein will afford an external control when investigations of such alternative outcomes are necessary on manufacturer-specific molecules. To supplement in-house, product-specific standards, a well-characterized material such as the NISTmAb can be run alongside the product of interest. As a molecule that is related in all salient features to other mAb products of its class, the same sample preparation artifacts would likely also be observed. In this sense, the NISTmAb reference material described herein will be useful in differentiating product variability from method artifacts.

## Conclusions

Matching theoretical masses to empirically determined masses of the predicted sequence of the intact NIST reference material and any treated forms of the material (reduced, partial digest, and various enzymatic preparations) can be collectively used to confirm an expected primary sequence. This exercise was conducted by several qualified laboratories and provided a consensus that Figure 1 does indeed correspond to the identity of the NISTmAb reference material, as well as established that intrinsic heterogeneity is present as is expected of all mAb products.

Intact MS analysis was first used to confirm that the intact mass of the protein was consistent with the theoretically predicted composition. Top-down sequencing was also performed on the intact mAb to provide additional

sequence-specific confidence in the assignment. Further mass and sequencing analysis was also performed after treatment of the NISTmAb with the IdeS enzyme and cysteine bond reductions. Analysis of the smaller, ~25 kDa fragments provided higher mass accuracy assignment of intrinsic heterogeneity within the individual polypeptides as well as more complete coverage of the amino acid sequence upon middle-down fragmentation. Finally, results of multiple enzyme digests and peptide mapping analysis were combined to confirm the amino acid sequence of the NISTmAb. Our totality of evidence approach provides orthogonal assurance that the primary structure was elucidated to the extent feasible with current MS technology.

## Acknowledgments

The authors would like to thank Peter Adams (Center for Drug Evaluation and Research) and Michail Alterman (Center for Biologics Evaluation and Research) for their careful review of this manuscript.

## Disclaimers

**FDA Disclaimer:** The findings and conclusions in this article have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any Agency determination or policy.

**NIST Disclaimer:** Commercial equipment, instruments, and materials are identified to adequately specify the experimental procedure and do not imply a recommendation or endorsement by NIST.

## References

1. Smith, L. M.; Kelleher, N. L.; Consortium Top Down, P. *Nat. Methods* **2013**, *10*, 186–187.
2. Sanger, F. *Adv. Protein Chem.* **1952**, *7*, 1–67.
3. Garrett, R.; Grisham, C. *Biochemistry*, 2 ed.; Saunders: Orlando, FL, 1999.
4. Berg, J.; Tymoczko, J.; Stryer, L. *Biochemistry*, 5 ed.; W. H. Freeman: New York, 2002.
5. Brown, J. R.; Hartley, B. S. *Biochem. J.* **1966**, *101*, 214–228.
6. Ryle, A. P.; Sanger, F.; Smith, L. F.; Kitai, R. *Biochem. J.* **1955**, *60*, 541–556.
7. George-Nascimento, C.; Gyenes, A.; Halloran, S. M.; Merryweather, J.; Valenzuela, P.; Steimer, K. S.; Masiarz, F. R.; Randolph, A. *Biochemistry* **1988**, *27*, 797–802.
8. Hancock, W. S.; Bishop, C. A.; Prestidge, R. L.; Hearn, M. T. *Anal. Biochem.* **1978**, *89*, 203–212.
9. Harding, D. R.; Bishop, C. A.; Tarttelin, M. F.; Hancock, W. S. *Int. J. Pept. Protein Res.* **1981**, *18*, 214–220.
10. Veihar, G. A.; Spellman, M. W.; Keyt, B. A.; Ferguson, C. K.; Keck, R. G.; Chloupek, R. C.; Harris, R.; Bennett, W. F.; Builder, S. E.; Hancock, W. S. *Cold Spring Harbor Symp. Quant. Biol.* **1986**, *51* (Pt 1), 551–562.

11. Walsh, K. A.; Sasagawa, T. *Methods Enzymol.* **1984**, *106*, 22–29.
12. Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
13. Chait, B. T. *Science* **2006**, *314*, 65–66.
14. Brady, L. J.; Valliere-Douglass, J.; Martinez, T.; Balland, A. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 502–509.
15. Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28*, 147–176.
16. 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.
17. An, Y.; Zhang, Y.; Mueller, H. M.; Shameem, M.; Chen, X. Y. *mAbs* **2014**, *6*, 879–893.
18. 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.
19. Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86*, 3005–3012.
20. Wang, B.; Gucinski, A. C.; Keire, D. A.; Buhse, L. F.; Boyne, M. T. *Analyst* **2013**, *138*, 3058–3065.
21. Bondarenko, P. V.; Second, T. P.; Zabrouskov, V.; Makarov, A. A.; Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1415–1424.
22. Madsen, J. A.; Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81*, 8677–8686.
23. 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.
24. Garcia, B. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 193–202.
25. Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 4239–4246.
26. Zhang, Z. Q.; Shah, B. *Anal. Chem.* **2007**, *79*, 5723–5729.
27. Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. *Anal. Chem.* **2011**, *83*, 8919–8927.
28. Zhang, Z.; Shah, B. *Anal. Chem.* **2007**, *79*, 5723–5729.
29. Linderstrøm-Lang, K. U. *Proteins and Enzymes*; Lane Medical Lectures; Stanford University Press: Redwood City, CA, 1952; Vol. 6.
30. Edelman, G. M.; Cunningham, B. A.; Gall, W. E.; Gottlieb, P. D.; Rutishauser, U.; Waxdal, M. J. *J. Immunol.* **2004**, *173*, 5335–5342.
31. Strupat, K.; Karas, M.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 89–102.
32. Wei, H.; Tymiak, A. A.; Chen, G. D. *Bioanalysis* **2013**, *5*, 1299–1313.
33. Valeja, S. G.; Kaiser, N. K.; Xian, F.; Hendrickson, C. L.; Rouse, J. C.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 8391–8395.
34. Nicolardi, S.; Deelder, A. M.; Palmblad, M.; van der Burgt, Y. E. M. *Anal. Chem.* **2014**, *86*, 5376–5382.
35. Shaw, J. B.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85*, 8313–8318.
36. Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S. *J. Chromatogr. A* **2006**, *1120*, 112–120.

37. Dillon, T. M.; Bondarenko, P. V.; Speed Ricci, M. *J. Chromatogr. A* **2004**, *1053*, 299–305.
38. Kilpatrick, E. L.; Liao, W. L.; Camara, J. E.; Turko, I. V.; Bunk, D. M. *Protein Expression Purif.* **2012**, *85*, 94–99.
39. Coplen, T.; Bohlke, J.; De Bievre, P.; Ding, T.; Holden, N.; Hopple, J.; Drouse, H.; Lamberty, A.; Peiser, H. *Pure Appl. Chem.* **2009**, *74*, 1987–2017.
40. Dick, L. W., Jr.; Kim, C.; Qiu, D.; Cheng, K. C. *Biotechnol. Bioeng.* **2007**, *97*, 544–553.
41. Dick, L. W., Jr.; Qiu, D.; Mahon, D.; Adamo, M.; Cheng, K. C. *Biotechnol. Bioeng.* **2008**, *100*, 1132–1143.
42. Harris, R. J. *J. Chromatogr. A* **1995**, *705*, 129–134.
43. Liu, Y. D.; Goetze, A. M.; Bass, R. B.; Flynn, G. C. *J. Biol. Chem.* **2011**, *286*, 11211–11217.
44. Luo, J.; Zhang, J.; Ren, D.; Tsai, W. L.; Li, F.; Amanullah, A.; Hudson, T. *Biotechnol. Bioeng.* **2012**, *109*, 2306–2315.
45. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2013**, *85*, 715–736.
46. Tsybin, Y. O. *Chimia (Aarau)* **2014**, *68*, 168–174.
47. Rehder, D. S.; Dillon, T. M.; Pipes, G. D.; Bondarenko, P. V. *J. Chromatogr. A* **2006**, *1102*, 164–175.
48. Wang, L.; Amphlett, G.; Lambert, J. M.; Blattler, W.; Zhang, W. *Pharm. Res.* **2005**, *22*, 1338–1349.
49. Dillon, T. M.; Ricci, M. S.; Vezina, C.; Flynn, G. C.; Liu, Y. D.; Rehder, D. S.; Plant, M.; Henkle, B.; Li, Y.; Deechongkit, S.; Varnum, B.; Wypych, J.; Balland, A.; Bondarenko, P. V. *J. Biol. Chem.* **2008**, *283*, 16206–16215.
50. Wypych, J.; Li, M.; Guo, A.; Zhang, Z.; Martinez, T.; Allen, M. J.; Fodor, S.; Kelner, D. N.; Flynn, G. C.; Liu, Y. D.; Bondarenko, P. V.; Ricci, M. S.; Dillon, T. M.; Balland, A. *J. Biol. Chem.* **2008**, *283*, 16194–16205.
51. Heimer, R.; Schnoll, S. S.; Primack, A. *Biochemistry* **1967**, *6*, 127–134.
52. Utsumi, S.; Karush, F. *Biochemistry* **1967**, *6*, 2313–2315.
53. Turner, M. W.; Bennich, H. H.; Natvig, J. B. *Nature* **1970**, *225*, 853–855.
54. Gergely, J.; Fudenberg, H. H.; van Loghem, E. *Immunochemistry* **1970**, *7*, 1–6.
55. Jefferis, R.; Weston, P. D.; Stanworth, D. R.; Clamp, J. R. *Nature* **1968**, *219*, 646–649.
56. Porter, R. R. *Biochem. J.* **1959**, *73*, 119–126.
57. Edelman, G. M.; Heremans, J. F.; Heremans, M. T.; Kunkel, H. G. *J. Exp. Med.* **1960**, *112*, 203–223.
58. Hsiao, S.; Putnam, F. W. *J. Biol. Chem.* **1961**, *236*, 122–135.
59. Adamczyk, M.; Gebler, J. C.; Wu, J. *J. Immunol. Methods* **2000**, *237*, 95–104.
60. von Pawel-Rammingen, U.; Johansson, B. P.; Bjorck, L. *EMBO J.* **2002**, *21*, 1607–1615.
61. Vlasak, J.; Ionescu, R. *mAbs* **2011**, *3*, 253–263.
62. McLafferty, F. W.; Venkataraghavan, R.; Irving, P. *Biochem. Biophys. Res. Commun.* **1970**, *39*, 274–278.

63. Barber, M.; Jolles, P.; Vilkas, E.; Lederer, E. *Biochem. Biophys. Res. Commun.* **1965**, *18*, 469–473.
64. Barber, M.; Wolstenholme, W. A.; Guinand, M.; Michel, G.; Das, B. C.; Lederer, E. *Tetrahedron Lett.* **1965**, 1331–1336.
65. Biemann, K. *J. Proteomics* **2014**, *107*, 62–70.
66. Biemann, K.; Cone, C.; Webster, B. R.; Arsenault, G. P. *J. Am. Chem. Soc.* **1966**, *88*, 5598–5606.
67. McLafferty, F. W.; Bockhoff, F. M. *Anal. Chem.* **1978**, *50*, 69–76.
68. McLafferty, F. W. *Science* **1981**, *214*, 280–287.
69. McLafferty, F. W.; Todd, P. J.; McGilverey, D. C.; Baldwin, M. A. *J. Am. Chem. Soc.* **1980**, *102*, 3360–3363.
70. Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. *J. Chem. Soc., Chem. Commun.* **1981**, 325–327.
71. Morris, H. R.; Panico, M.; Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. *Biochem. Biophys. Res. Commun.* **1981**, *101*, 623–631.
72. Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233–6237.
73. Tomer, K. B. *Mass Spectrom. Rev.* **1989**, *8*, 445–482.
74. Tomer, K. B. *Mass Spectrom. Rev.* **1989**, *8*, 483–511.
75. Hunt, D. F.; Bone, W. M.; Shabanowitz, J.; Rhodes, J.; Ballard, J. M. *Anal. Chem.* **1981**, *53*, 1704–1706.
76. Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3*, 608–614.
77. Chohnan, S.; Shiraki, K.; Yokota, K.; Ohshima, M.; Kuroiwa, N.; Ahmed, K.; Masaki, T.; Sakiyama, F. *J. Bacteriol.* **2004**, *186*, 5093–5100.
78. Elliott, B. W., Jr.; Cohen, C. *J. Biol. Chem.* **1986**, *261*, 11259–11265.
79. Jekel, P. A.; Weijer, W. J.; Beintema, J. *J. Anal. Biochem.* **1983**, *134*, 347–354.
80. Masaki, T.; Tanabe, M.; Nakamura, K.; Soejima, M. *Biochim. Biophys. Acta* **1981**, *660*, 44–50.
81. Masaki, T.; Fujihashi, T.; Nakamura, K.; Soejima, M. *Biochim. Biophys. Acta* **1981**, *660*, 51–55.
82. Beaudet, R.; Saheb, S. A.; Drapeau, G. R. *J. Biol. Chem.* **1974**, *249*, 6468–6471.
83. Garg, G. K.; Virupaksha, T. K. *Eur. J. Biochem.* **1970**, *17*, 13–18.
84. Houmard, J.; Drapeau, G. R. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 3506–3509.
85. Keil, B., *Specificity of Proteolysis*; Springer-Verlag: Berlin; New York, 1992.
86. Drapeau, G. R. *J. Biol. Chem.* **1980**, *255*, 839–840.
87. Luo, Q.; Joubert, M. K.; Stevenson, R.; Ketchum, R. R.; Narhi, L. O.; Wypych, J. *J. Biol. Chem.* **2011**, *286*, 25134–25144.
88. Yang, Y.; Strahan, A.; Li, C.; Shen, A.; Liu, H.; Ouyang, J.; Katta, V.; Francissen, K.; Zhang, B. *mAbs* **2010**, *2*, 285–298.
89. Valliere-Douglass, J. F.; Kodama, P.; Mujacic, M.; Brady, L. J.; Wang, W.; Wallace, A.; Yan, B.; Reddy, P.; Treuheit, M. J.; Bolland, A. *J. Biol. Chem.* **2009**, *284*, 32493–32506.
90. Goetze, A. M.; Liu, Y. D.; Arroll, T.; Chu, L.; Flynn, G. C. *Glycobiology* **2012**, *22*, 221–234.



91. Boyne, M. T.; Garcia, B. A.; Li, M.; Zamdborg, L.; Wenger, C. D.; Babai, S.; Kelleher, N. L. *J. Proteome Res.* **2009**, *8*, 374–379.
92. Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18132–18138.
93. Sakiyama, F.; Masaki, T. *Methods Enzymol.* **1994**, *244*, 126–137.
94. Tsunasawa, S.; Masaki, T.; Hirose, M.; Soejima, M.; Sakiyama, F. *J. Biol. Chem.* **1989**, *264*, 3832–3839.
95. Gershon, P. D. *J. Proteome Res.* **2014**, *13*, 702–709.
96. Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77*, 5800–5813.
97. Ryan, M. H.; Petrone, D.; Nemeth, J. F.; Barnathan, E.; Bjorck, L.; Jordan, R. E. *Mol. Immunol.* **2008**, *45*, 1837–1846.
98. Liu, H.; May, K. *mAbs* **2012**, *4*, 17–23.
99. Pink, J. R.; Milstein, C. *Nature* **1967**, *214*, 92–94.
100. Pink, J. R.; Milstein, C. *Nature* **1967**, *216*, 941–942.
101. Milstein, C. *Biochem. J.* **1966**, *101*, 338–351.
102. Frangione, B.; Milstein, C. *Nature* **1967**, *216*, 939–941.
103. Frangione, B.; Milstein, C.; Franklin, E. C. *Biochem. J.* **1968**, *106*, 15–21.
104. Frangione, B.; Milstein, C. *J. Mol. Biol.* **1968**, *33*, 893–906.
105. Frangione, B.; Milstein, C.; Pink, J. R. *Nature* **1969**, *221*, 145–148.
106. Edelman, G. M.; Cunningham, B. A.; Gall, W. E.; Gottlieb, P. D.; Rutishauser, U.; Waxdal, M. J. *Proc. Natl. Acad. Sci. U. S. A.* **1969**, *63*, 78–85.
107. Martinez, T.; Guo, A.; Allen, M. J.; Han, M.; Pace, D.; Jones, J.; Gillespie, R.; Ketchum, R. R.; Zhang, Y.; Balland, A. *Biochemistry* **2008**, *47*, 7496–7508.
108. Liu, Y. D.; Chen, X.; Enk, J. Z.; Plant, M.; Dillon, T. M.; Flynn, G. C. *J. Biol. Chem.* **2008**, *283*, 29266–29272.
109. Bean, M. F.; Carr, S. A. *Anal. Biochem.* **1992**, *201*, 216–226.
110. Zhang, W.; Marzilli, L. A.; Rouse, J. C.; Czupryn, M. J. *Anal. Biochem.* **2002**, *311*, 1–9.
111. Mhatre, R.; Woodard, J.; Zeng, C. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2503–2510.
112. Wang, Y.; Lu, Q.; Wu, S. L.; Karger, B. L.; Hancock, W. S. *Anal. Chem.* **2011**, *83*, 3133–3140.
113. Wu, S. L.; Jiang, H.; Lu, Q.; Dai, S.; Hancock, W. S.; Karger, B. L. *Anal. Chem.* **2009**, *81*, 112–122.
114. Wei, Z.; Tous, G.; Yim, A.; Hope, J. N.; Casas-Finet, J. R.; Folena-Wasserman, G.; Schenerman, M. A. *Dev. Biol. (Basel)* **2005**, *122*, 29–47.

## Chapter 2

# Sequence Variants and Sequence Variant Analysis in Biotherapeutic Proteins

Oleg V. Borisov,<sup>\*,1</sup> Melissa Alvarez,<sup>2</sup>  
James A. Carroll,<sup>3</sup> and Paul W. Brown<sup>3</sup>

<sup>1</sup>Novavax, Inc., Gaithersburg, Maryland 20878, United States

<sup>2</sup>Roche Group Member, Genentech, Inc., South San Francisco, California 94080, United States

<sup>3</sup>Pfizer Worldwide Research & Development, Chesterfield, Missouri 63017, United States

\*E-mail: oborisov@novavax.com

A “sequence variant” is a surrogate term covering any unintentional amino acid substitutions, omissions, or insertions during protein biosynthesis. Production of biotherapeutic proteins by living organisms is governed by biological processes responsible for protein production, including DNA replication, RNA transcription, and protein translation steps. Each of the biosynthesis steps has finite fidelity, with error rates ranging from  $10^{-9}$  per base pair for DNA replication to  $10^{-4}$  to  $10^{-5}$  per codon for protein translation.

The occurrence of sequence variants contributes to heterogeneity of recombinant protein therapeutics. Establishing a sequence variant profile of a biotherapeutic product is essential in providing proof of its structure, its manufacturing consistency, and the stability of the producing cell line.

DNA mutations can arise during generation of stable producing cell lines and be amplified by selection pressures used to establish high-producing clones. Once a clone with genetic variant is produced, this mutation will propagate through sub-cloning and can be affected by a cell age. For that reason, the prime goal of cell line (upstream) development is to screen for a possibility of DNA mutations early on during upstream development. In contrast, naturally abundant translational errors can further amplify due to nutrient and amino acid

deprivation in production medium or a non-optimal codon usage in the construct. These errors usually occur at low levels and affect certain amino acids in multiple positions along the protein sequence, and can be both codon-dependent and independent.

Although there is no standardized method for detection of sequence variants, this chapter discusses analytical methodologies for their screening. Peptide mapping with liquid chromatography-mass spectrometry (LC-MS) is currently viewed as a technology of choice for detection of sequence variants in recombinant proteins, with potential for detecting amino acid substitutions at low levels. On the example of sequence variant analysis in the NIST monoclonal antibody (mAb), performed by two independent laboratories, this chapter discusses strategy, capabilities, and limitations of peptide mapping with LC-MS for this application. Comparison of the intra-laboratory data illustrates reasonable agreement between the results, exhibiting a significant overlap in the specific sequence variants detected. Despite the excellent capabilities of peptide mapping method for detecting sequence variants in most cases at low or trace levels, there is no guarantee that such analysis is comprehensive, as discussed in this chapter. The use of multiple and orthogonal analytical methods often assists with the detection of variants missed by one or another method.

## Foreword

In 1949, while studying sickle cell anemia, a blood disorder in which the erythrocytes assume a rigid, sickle shape, Linus Pauling and his collaborators reported significant differences between electrophoretic mobilities of hemoglobin derived from erythrocytes of normal individuals and from those of sickle anemic individuals (*1*). Researchers made a notable prediction that “there is a surface region on the ... hemoglobin molecule (of the sickle cell anemia) which is absent in the normal molecule and which has a configuration complementary to a different region of the surface of the hemoglobin molecule.” They further discovered that under conditions of low oxygen pressure “the sickle cell anemia hemoglobin molecules might be capable of interacting with one another at these sites sufficiently to cause at least partial alignment of the molecules within the cell, resulting in the erythrocyte’s becoming birefringent, and the cell membrane’s being distorted to accommodate the now relatively rigid structures within its confines.” Ahead of its time, this discovery marked the first evidence of the molecular origin of a human disease, a “molecular disease.”

Attributing to Pauling and his collaborators “the realization that sickle cell anemia ... is due to an alternation in the structure of a large protein molecule, an alternation leading to a protein which is by all criteria still a hemoglobin,” Vernon Ingram concluded that “per half-molecule of hemoglobin, this change consists

in a replacement of only one of nearly 300 amino acids, namely, glutamic acid, by another, valine—a very small change indeed” (2). In fact, adult hemoglobin consists of two  $\alpha$ -globin chains and two  $\beta$ -globin chains. The single nucleotide change on the genetic level alters only one amino acid in the  $\beta$ -globin protein chain, but the results are devastating (3).

But yet another notable consequence of Pauling’s work was that it paved the way for the development of improved analytical technologies capable of detecting and resolving complex biological molecules bearing a single amino acid difference.

## Introduction

The paramount concept of modern biotechnology is that it is the art of making a cell proliferate a protein that it normally would not produce. Recombinant DNA technology, based on cloning and expression of the heavy and light chain polypeptide genes in mammalian cell cultures, is currently the principal way for the commercial production of monoclonal antibody (mAb)-based therapies (4–6). Recent advancements in host cell line development, cell culture media and feeding strategies, and bioreactor optimization delivers titers of 1 to 5 g/L and even up to about 10 g/L. Establishment of stable cell line clones with high specific productivities serves as the starting point for the development of a biotherapeutic drug for its commercial production. To achieve high titers, cell line development often targets selection of higher producing clones by cell line sub-cloning and the use of gene amplification (7, 8). Two common Chinese hamster ovary (CHO) cell line expression systems for commercial production of mAbs utilize dihydrofolate reductase (DHFR) and glutamine synthetase (GS) genes for clone selection. They use drug resistance to methotrexate (MTX) and methionine sulfoximine (MSX), respectively, as the tool of choice for gene amplification, targeting improved specific productivity of the cell line for production of a protein of interest. To match the ever increasing market demand along with tuning expression systems, the subsequent downstream purification processes are optimized to produce large quantities of quality biotherapeutic proteins.

Because of the very nature of the biosynthetic processes involved in production of biotherapeutics, the desired product consists of a mixture of related compounds, which defines the heterogeneity profile of that product. Many established co- and post-translational processes accompanying protein expression play essential roles in ensuring proper protein folding and structure, which are known to influence protein function (9). Thus, a wide spectrum of post-translational modifications (PTMs), occurring after a polypeptide chain is assembled, contributes to the heterogeneity of mAbs. PTMs have been the subject of extensive research since the early days of their development as biotherapeutics (10–13).

To date, many commonly occurring PTMs have been characterized, and their effect on product quality is established (12, 14). These are discussed in detail in the PTMs chapter/Volume 2, Chapter 3 of this book. Occurrence of PTMs is governed by the rate of underlying co-translational or post-translational

events. For example, glycosylation is one of the most common and diverse PTMs biosynthesized through metabolic pathways. This non-template-driven process results in a diverse pattern of multiple glycoforms, where the chemical nature of the glycan attached to a specific glycosylation site tends to vary from copy to copy of the same protein. In contrast, biosynthesis of DNA, mRNA, and proteins are template-driven processes. Characteristic of these processes is an overall improved fidelity in maintenance, processing, and transfer of the genetic information. As the consequence, errors occurring prior or during the incorporation of amino acids into polypeptide chain, often referred to as sequence variants (SVs), are much less abundant. These errors and their occurrence and detection are the main scope of this chapter.

Concerns for efficacy and patient safety stimulate the manufacturer of a biotherapeutic to ensure consistent production of a quality product, conduct its extensive characterization throughout the development phases, and establish strict control of its quality. As defined by the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use, the heterogeneity of biotherapeutic products establishes their quality; the degree and profile of this heterogeneity should be characterized to ensure production consistency. When the variants (SVs and PTMs alike) of the desired product have properties comparable to those of the desired product with respect to activity, efficacy, and safety, they are considered product-related substances. The occurrence of a variant is monitored throughout the manufacturing history of that product, and knowledge obtained from preclinical and clinical studies is applied to determine the criticality of that variant. However, when process changes and degradation products result in heterogeneity patterns that differ from those observed in the material used during preclinical and clinical development, the significance of these alterations should be evaluated. Observed differences in heterogeneity prompt manufacturers to conduct additional studies directed toward determining the efficacy and safety parameters of that variant. These points highlight the importance of knowledge about potential variants and their relevant characteristics as well as the importance of the ability to detect these variants.

### **Protein Synthesis by Living Cells**

As formulated by the central dogma of molecular biology, the basis of the biosynthesis of proteins is the transfer of the genetic information (blueprint) to formation of a biopolymer chain of amino acids, which are linked together via amide bonds (15). Copying (transcribing) a particular segment of the genome into the primary transcript messenger RNA (pre-mRNA), creates a copy of that gene, allowing for its subsequent regulated expression. During the process, the RNA polymerase enzyme moves along the DNA template and synthesizes a complementary RNA strand (16). In eukaryotic cells, maturity of mRNA is accomplished by its further processing and splicing to ensure proper decoding of genetic information. Splicing of pre-mRNA molecules is the process during which some domains are excised as removed introns to produce mature mRNA that retains different subsets of the domains of the original RNA. Mature mRNA is then primed for export from the nucleus into the cytoplasm for protein translation.

Each mRNA molecule is further translated into a polypeptide sequence by a complex cellular machinery including ribosomes, transfer RNA (tRNA) and their associated amino acids, and numerous protein translation factors and other enzymes (17). Enzymes, particularly aminoacyl-tRNA synthetases (AARSs), play important roles in the translation process by correctly pairing amino acids with their cognate tRNAs; this process is known as “charging.” Protein translation begins when the ribosome locates the initiation codon, which flags the starting position of the translated region of the mRNA. Delivery of a genetically coded amino acid occurs when corresponding amino-acylated (charged) tRNA diffuses to the peptidyltransferase center of the ribosome where it interacts within the ribosome by matching a three-base codon of the mRNA via anticodon-codon pairing. Subsequently, as the ribosome moves down the mRNA, elongation of the polypeptide chain occurs, and the process (assisted by elongation factors) assembles the polypeptide chain by linking amino acids (provided by correctly amino-acylated tRNAs) sequentially base-paired to the mRNA, three bases at a time. Termination, defined by a position of a stop (nonsense) codon that is not recognized by any tRNA, signals the end of the translation cycle. The polypeptide chain assembly is then terminated by action of proper release factors for further post-translational processing.

### Fidelity of Biosynthesis

Although fidelity of biological information flow is critical, synthesis of a functional protein from genetic information is an error-prone process. Translational errors occur in all biological systems, regardless of whether these take place *in vivo* or *in vitro* during the production of recombinant proteins in a bioreactor. Polypeptide synthesis errors in living organisms can have all sorts of deleterious effects, including disease and cell death, which reduce an organism’s fitness. Evidence shows that disruption of translational fidelity, caused by antibiotics, kills bacteria. Translational errors may also alter protein folding, which is required for its proper function, and the misfolded protein may even be toxic to the cells. Furthermore, protein folding may be affected not only by its primary sequence, but also by the pace of the protein expression affecting the timing of co-translational folding (18, 19). Thus, even seemingly evolutionary neutral synonymous (or silent) mutations, which do not change the amino acid sequence of a protein can have an effect on its function. However, it is hard to unambiguously claim if the effect of these errors is detrimental in all cases. The evolutionary responses to expression errors may also result in adaptations that minimize errors or adaptations that even exploit errors for the organism’s benefit (20).

Erroneous protein synthesis can be a result of numerous factors that cause disruption at any step in protein expression, including errors during DNA replication, mRNA transcription and processing, translational processes, and post-translational events such as protein misfolding, as shown in Figure 1. The error rate of each event depends on a global fidelity of an underlying process, as well as the nature and condition (balanced or stressed) of the studied system. In addition, the technique and type of measurement itself (*in vivo* or *in vitro*) also

contribute to the estimated error rate. The overall fidelity of protein biosynthesis is a composite of the processes leading to that error and the processes of its repair by proofreading and editing. However, the error correction processes are energetically expensive; thus maximum accuracy, due to its high energy cost, is never achieved by living cells (21, 22). When examining fidelity of cellular processes during protein expression it is not surprising that these error rates cover a broad range, with frequencies spreading over several orders of magnitude.

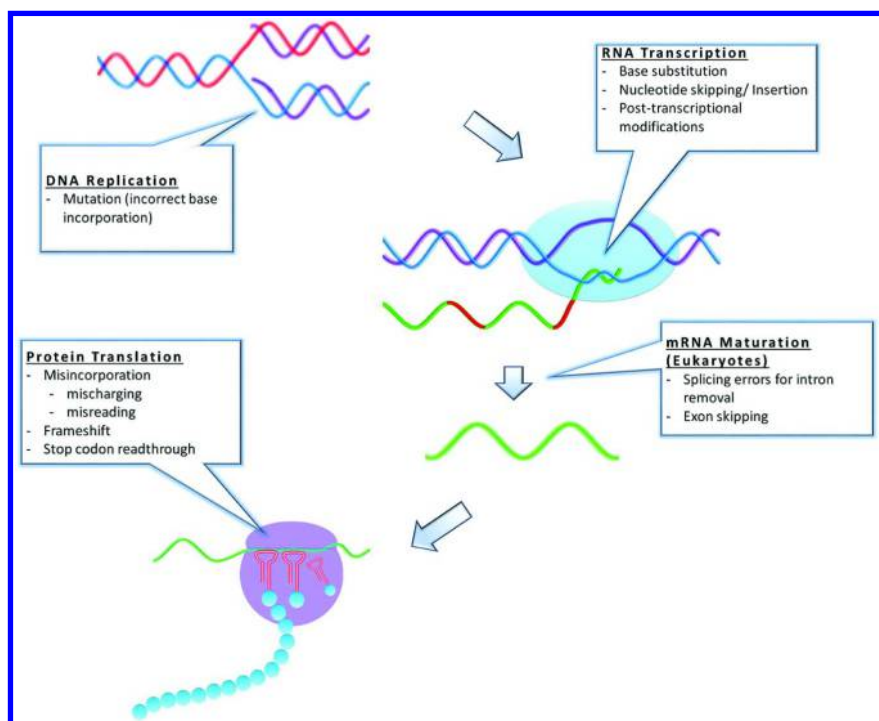


Figure 1. Processes involved in biosynthesis of proteins and potential errors leading to the occurrence of sequence variants.

Accurate DNA replication is important for preserving genetic information, and it is the most critical step of protein synthesis. Owing to evolved extensive proofreading activity of DNA polymerases and associated enzymatic DNA repair processes, a high fidelity of the transfer of genetic information is ensured. Thus, the error rates of copying genetic information are very low and are reported to be in the range of  $10^{-6}$  to  $10^{-11}$  per base pair. In particular, the DNA mutation rate reported for *Escherichia coli* (*E. coli*) is approximately  $10^{-9}$  per base pair (17). Errors during DNA replication can occur when an incorrect base is incorporated into the growing strand of DNA, leading to mismatched base pairs and the introduction of DNA mutation, which can occur in both the coding and non-coding regions of a gene. In the coding region, two types of DNA mutations

are defined as synonymous, causing a non-altered amino acid sequence of a coded polypeptide chain, or non-synonymous, involving substitution of a single amino acid by either a different amino acid (missense mutation) or generation of a premature stop codon (nonsense mutation). Although the underlying processes of DNA replication have high fidelity, the presence of sequence variants in genomes, known as single-nucleotide polymorphism (SNP), as the result of the action of evolutionary processes, is estimated to occur in an individual every 1000 to 2000 base pairs in the human genome (23).

Scientific evidence suggests that overall protein biosynthesis errors are several orders of magnitude more frequent than DNA replication errors. Less is known about the transcription error rates of different organisms, in part because of the challenges of disentangling changes in protein sequences caused by transcription errors from the potentially much more common changes caused by translation errors. Transcription fidelity by RNA polymerases during biosynthesis of pre-mRNA are reported to have errors in the range of  $3 \times 10^{-4}$  to  $4 \times 10^{-5}$  in *E. coli* (24). Recently, a method for identifying transcription errors by sequencing multiple complementary DNAs (cDNAs) originating from the same mRNA molecule was applied to study transcriptome of *Caenorhabditis elegans*. This method revealed transcription errors with rates of  $\sim 4 \times 10^{-6}$  and exhibited C→U base substitution as the most common type of transcription error, occurring due to a G (DNA)/U(mRNA) base pair mismatch (25, 26). Although more frequent, as compared with DNA replication, good fidelity of transcription is checked in part by the enzymes that detect and remove a misincorporated nucleotide during proofreading (27). For eukaryotes, fidelity of a splicing event for intron removal, a process associated with maturation of mRNA, has been reported to have good accuracy with estimated errors of  $10^{-5}$  for higher eukaryotes (28). At the same time, it is also noted that alternative splicing, known to be prevalent in complex eukaryotes with rates reaching about  $10^{-2}$ , is a likely consequence of suboptimal splicing or splicing regulation—programmed events leading to an increase in diversity of expressed proteins for a limited number of genes.

Errors during subsequent mRNA translation on the ribosome can be the result of action of several processes, including perturbation in the mechanism of coded amino acid delivery or errors during mRNA passage through the ribosome (17, 21). The first type of errors results in misincorporation of an erroneous amino acid into the expressed protein sequence, although not affecting its overall length. These errors, often referred to as missense errors by analogy with missense mutations, can be the result of either faulty tRNA charging with a non-cognate amino acid (mischarging or misacylation) or an anticodon-codon mismatch on the ribosome (misreading). Translational errors are reported to occur with frequencies ranging from  $10^{-4}$  to  $10^{-5}$  under typical *in vivo* conditions (21, 22, 29, 30). Misacylation occurs when structurally similar amino acids cannot initially be distinguished with adequate selectivity by AARS and, in turn, charging the corresponding tRNA with a non-cognate amino acid. It was experimentally determined that initial discrimination by AARS between amino acids differing by just one methylene group in their structures cannot be better than a factor of 200 (22). Despite this high degree of uncertainty, tRNA acylation is accompanied by corrective processes of pre-transfer hydrolysis or post-transfer deacylation of mischarged tRNA, by



error editing pathways of cognate AARSs (29–31). Closely related amino acids tend to have closely related codons, although similarity of their codons is not a prerequisite for their recognition by AARS. Leaping ahead of the discussion, the possibility of misincorporation of an erroneous amino acid into the synthesized protein sequence, encoded by a codon, differing by more than a single base pair from that of a correctly coded amino acid, exists and should not be discounted. The detection of this type of misincorporation can challenge algorithms that are based on the SNP principal.

A three-nucleotide unit of tRNA corresponding to the genetic code on the mRNA is called the anticodon. Misreading errors occur when a correctly acylated tRNA mismatches a codon on the mRNA with improper codon-anticodon base pairing. This process leads to the incorporation of an erroneous amino acid in place of the genetically encoded amino acid. It was predicted that most errors of this type should occur at the third position of the codon, which is the most degenerate, although errors due to the wobble at the first and second positions of the codon are also observed in practice (26).

In addition, processivity errors due to non-standard mRNA decoding events can lead to expression of variants of a protein, the sequence of which *a priori* cannot be predicted from the corresponding DNA sequence. These variants can be due to errors during translation initiation, where an alternative to the natural AUG codon initiation sites can produce N-terminus variants of the expected protein. During polypeptide chain elongation, disruption in a sequential three base pair at-a-time reading event can lead to either  $-1$  or  $+1$  frameshift due to, respectively, 2-base (backward) and 4-base (forward) translocations or even bypassing large stretches of mRNA (hopping) (17). A faulty termination event can also lead to the stop codon readthrough event, also known in literature as a leaky stop codon, where tRNA turns the stop codon into a sense codon (21).

It should also be noted that not all differences between the genetic code-derived sequence of a protein and its actual *in vivo* expressed sequence are the result of errors. In particular, not all DNA mutations are due to errors. A so-called beneficial mutation takes advantage of establishing a beneficial phenotype by promoting the frequency of a specific mutation, allowing an organism, for example, to adapt to changing environmental conditions (20). In certain cases, reduced quality control of AARS to charge tRNA with a cognate amino acid, can be used by cells to benefit their survival (29). In other cases, re-coding of the genetic information can be programmed into mRNA by its primary or secondary structure, which alters the triplet-based decoding process and results in alternative nonstandard reading. It allows organisms to take advantage of this flexibility of reading the genetic information and provides novel regulatory options during translation (17, 21, 32). The programmed events occurring at particular re-coding sites are known to occur in viral genes, where they allow incorporation of more genetic information into a relatively short amount of genetic code.

## Biosynthetic Errors during Protein Expression from *Escherichia coli* Production Systems

The synthesis of a functional hormone somatostatin in 1977 was the first published example of expression of mammalian polypeptides in *E. coli* bacteria (33). This *experimentum crucis* paved the way toward the wide use and acceptance of recombinant DNA technology. Historically, *E. coli* systems served as primary hosts for laboratory- and commercial-scale production of biotherapeutics. In fact, the first recombinant protein drugs that received regulatory approval for human use were insulin in 1982 and growth hormone in 1985, both produced in *E. coli* cell lines. In this regard, it seems reasonable to begin the discussion of protein expression errors with production in prokaryotic systems, such as *E. coli* cell lines. Based on the abundant practical experience of mammalian protein production in these cell lines, it is now well known that overexpression of heterologous proteins can lead to unbalanced conditions and increased error rates of the whole protein synthesis machinery. Biosynthesis errors were reported to occur during expression in *E. coli* systems. For example, misincorporation of a single Tyr residue (encoded by UAC codon in the mRNA sequence) by Cys (encoded by UGC or UGU codons) during expression of human  $\alpha$ -synuclein leads to the production of nearly 20% of molecules with Tyr136→Cys misincorporation (34). A highly efficient UGA stop codon readthrough and incorporation of Trp (UGG) in place of the stop codon was also reported during production of human methionyl-neurotrophin 3 (35). This was attributed to a competition of tRNA<sup>Trp</sup>, which is abundant in *E. coli*, with a release factor.

One of the manifestations of the unbalanced conditions is the effect known as the codon bias. Prokaryotes and eukaryotes have subtle differences in frequencies and usage preferences of specific codons, which are balanced by the amounts of cognate tRNAs available for polypeptide synthesis (36). Moreover, the levels of cognate tRNAs were shown to correlate well with the frequency of codon usage. Ideally, in a balanced biological system, the tRNA content available for translation agrees with the codon usage, where the most abundant tRNA isoacceptors are those that have an anticodon that matches the codons that are frequently used. Under sub-optimal conditions, however, translation of the overexpressed heterologous mRNA with a rare (for prokaryotes) codon can lead to a pause at the ribosome, which can cause translational problems (36). Such ribosome stalling is believed to be the primary reason for mid- to low-level codon misreading events, caused by the lack of a proper aminoacyl-tRNA and its substitution by an erroneous aminoacyl-tRNA (17, 36).

One of the best examples of the consequences of codon bias during production of mammalian proteins in *E. coli* is incorporation of Lys in place of Arg residues (or Arg→Lys misincorporation) (37–41). The UCU tRNA<sup>Arg</sup>, encoded in *E. coli* by the argU gene, is one of the rarest tRNAs, constituting only about 3% of the total tRNA<sup>Arg</sup> pool. Its cognate codon, AGA, is similarly rare, particularly in *E. coli*, accounting for about 0.1% and 2% of the total and Arg codons, respectively (42). As a result, misincorporation of Lys at three AGA codons was reported to occur with frequencies of 0.36 to 0.42 during production of a recombinant fusion protein (39), because relative amounts of tRNA<sup>Lys</sup> is nearly 15-fold higher than those of

tRNA<sup>Arg</sup> in the total tRNA pool (41). As a consequence, mistranslation of Arg for Lys produced two variants of human homodimeric triosephosphate isomerase (HsTIM) at almost equal amounts under optimal conditions of overexpression of heterologous protein (41). In another example, expression of bovine placental lactogen, encoded by a structural gene construct containing nine rare Arg codons (AGG and AGA), led to in-frame deletion of two consecutive amino acids from its internal sequence, explained by a translational hop associated with the stall at the AGG codon (43).

Similarly, Arg→Gln misincorporations were associated with the use of the rare Arg CGG codon, which is several times less frequent in *E. coli* than the CAG codon encoding for Gln (44). In another example, a good correlation between Gly→Glu misincorporation and the use of a rare Gly GGA codon led to increased missense errors during expression of a 100-amino acid Protein X with frequencies up to 0.1 (45) and in a fusion domain of a 92-kDa IgG1 Fc-fusion protein (44). A good example of the effect of rare codons and their clusters was also observed during expression of a Human Simplex Virus 2 p27 protease by genetic construct, which contained 11 rare Arg codons CGG, with three of them occurring in a tandem cluster near the C-terminus of the protein (46). Besides the significant levels of Arg→Gln substitutions, researchers found that major forms of the expressed protein were 3 kDa heavier than predicted. The two major species were further found to be due to the +1 frameshift events at both of the second two codons of the CGG triplet, leading to the subsequent translation event and allowing it to proceed in the missense frame until the next termination codon was reached.

In addition to the increased levels of misincorporation, codon bias can manifest itself in the failure of plasmid stability, poor cell viability, and low expression rates of certain recombinant proteins (46, 47).

The effect of physiological demand on expression machinery during synthesis of heterologous proteins can also lead to situations when even relatively abundant codons become limiting if the demand for them is great enough (36). Expression of a recombinant histidine-rich protein II (HRP II)—which contains 36% His residues in its sequence, exceeding the natural occurrence of His residues in a typical *E. coli* protein by 16-fold (36)—exhibited Gln for His misincorporation (48). The extent of the His→Gln substitution was reduced by decreasing the expression rate of the protein, emphasizing that the frequency of codon usage affects the accuracy of protein synthesis, even if these codons are well-represented in the genome of the expressing host.

Another mechanism of misincorporation can be caused by tRNA mischarging due to the relaxed selectivity of aminoacyl-tRNAs synthetases. Of all the AARSs, half of them—including MetRS, IleRS, LeuRS, ValRS, AlaRS, LysRS, GlnRS, ProRS, PheRS, and ThrRS—are known to be less selective in charging tRNA with a cognate amino acid (30). Expression of human thioredoxin protein exhibited multiple misincorporations of Glu→Lys, Ser→Thr, Phe→Ile, Leu→Lys, Leu→Val, and Asp→Glu, with frequencies of  $5 \times 10^{-2}$  to  $10^{-3}$  (49). Although the particular nucleotide sequence of the construct used for expression of thioredoxin was not presented, misincorporation of Lys (encoded by either AAA or AAG) for Leu (encoded by either one of six codons UGA, UUG, CUU,

CUC, CUA, or CUG) cannot be explained based on the principals of SNP. These misincorporations, resulting in incorporation of an amino acid that would otherwise require more than a single base pair change to occur, can be explained by the tRNA charging with an erroneous amino acid by its cognate AARS. It is likely that the mischarging of tRNA<sup>Lys</sup> with Leu by LysRS resulted in the production of Leu-tRNA<sup>Lys</sup> (30), thus causing the observed misincorporation.

The above cases represent misincorporation of one canonical amino acid for another. However, incorporation of non-canonical amino acids is also known to occur during biosynthesis in *E. coli* as the artifact of translational events. Norleucine (Nle) and norvaline (Nva) were reported to incorporate into protein sequences in place of Met and Leu, respectively. Specifically, Nle incorporation in place of Met during expression of recombinant bovine somatotropin (50), interleukin-2 (51, 52), and truncated human macrophage colony stimulating factor (53) was observed to reach levels of 50%, 19%, and 20%, respectively. Apostol et al. reported Leu→Nva misincorporations at multiple locations of recombinant human hemoglobin, ranging from 0.5 to 3% (54). Both Nle and Nva are non-proteinogenic amino acids, implying that their cognate tRNAs do not exist. These amino acids, however, can successfully interact with AARS (Nle with MetRS and Nva with LeuRS), which in turn mischarge corresponding tRNA<sup>Met</sup> and tRNA<sup>Leu</sup>. Limited selectivity of MetRS and LeuRS and relatively high intracellular concentrations of Nle and Nva make these amino acids competitive for incorporation into recombinant proteins.

A good understanding of the root causes of translational errors is essential for eliminating them. Several approaches were suggested to better fit *E. coli* expression systems to produce large quantities of quality mammalian recombinant proteins (55). Codon optimization strategies by site-directed mutagenesis of the target sequence for the generation of codons reflecting the tRNA pool in the host system were used. Replacement of the rare AGA Arg codon with CGU or CGC codons, which are abundant in *E. coli*, allowed Arg→Lys misincorporations to be avoided in several proteins (37, 39, 41, 44). Co-expression of argU or argX genes, encoding the cognate tRNA<sup>Arg</sup> decoding AGA and AGG codons or CGG codon, respectively, largely eliminated translational errors associated with the codon bias (39, 41, 47) and, moreover, improved expression levels of recombinant protein by seven-fold (46). In addition to optimization of gene sequences for production in *E. coli*, careful attention to fermentation conditions was reported to be effective. Varying the bioreactor environment by changing media, temperature, and point of induction was found to play a role in reduction of translational errors to very low levels (44, 49). Likewise, understanding the underlying mechanism of Met→Nle misincorporation allowed researchers to eradicate this error. Nle is a by-product of the leucine biosynthetic pathway, suggested by an increase in Nle intracellular concentration that was measured in response to the increased demand for Leu on the cells (50, 52). Thus, preventing either the formation of Nle by deleting the Leu operon of Nle to eliminate the endogenous synthesis pathway or the overall suppression of Leu biosynthesis by supplying Leu during cell culture or boosting medium supplementation with Met was found to be effective to prevent Met→Nle misincorporation (50, 52).

In summary, under balanced growth conditions in *E. coli*, background errors by the cellular machinery for protein assembly, which are checked by the cellular mechanisms via editing and extensive proofreading processes, can have relatively low frequencies in the range of  $10^{-4}$  to  $10^{-5}$ . In contrast, overexpression of proteins sufficiently upsets the expression system by way of nutritional stress or codon usage mismatch, leading to a significant increase in error rates, often with frequencies of  $10^{-1}$  to  $10^{-2}$ .

### *Biosynthetic Errors during Protein Expression from Mammalian Cell Lines*

*E. coli* expression systems are one of the most economical choices for production of proteins, as they offer simple cultivation requirements, short generation times, fast growth kinetics, and good product titers. One of their shortcomings, however, is the inability to perform post-translational decoration of expressed proteins with glycans, which is often required for a proper protein function (56, 57). The use of mammalian cell lines was established to overcome these issues and is now a well-accepted means to produce properly folded and glycosylated biotherapeutic proteins.

The principles of protein synthesis in prokaryotes and eukaryotes are similar; however, the greater complexity of the eukaryote genome and presence of more steps in the transcription and translation processes require greater variety and complexity of control mechanisms (58). The presence of a more sophisticated proofreading and error editing machinery alone does not automatically assume the greater overall fidelity of protein expression in eukaryotes. In fact, the evidence suggests that eukaryotes are no more or less accurate at protein synthesis than are prokaryotes (20, 59, 60). Reported cases of errors during expression of recombinant proteins from mammalian cell lines are summarized in Table 1 (additional examples can be found in Ref. (26)).

**Table 1. Reported Cases of Sequence Variants in Recombinant Proteins Produced in Mammalian Cell Lines<sup>1</sup>**

Variant	Site(s)	Level per site	Molecule	Expression System	Cause	Effect	Detection		Ref
							Primary	Confirmation	
Tyr→Gln	H(376) <sup>2</sup>	1–27%	IgG1; anti-Her2	CHO; DHFR/MTX	Genetic (polyclonality) TAT→CAA	No effect on binding and Fc effector functions	Tryptic LC-UV map	AAA, Edman, PCR	(61)
V <sub>L</sub> -C <sub>H</sub> crossover	L(Lys107)-H(Ser125) fusion via Pro	N/A	IgG	N/A	Genetic recombination	Unknown	SDS-PAGE, Lys-C LC-UV map	Edman, PCR	(62)
24-AA sequence insertion	Between V <sub>H</sub> and C <sub>H1</sub>	>50%	IgG1; anti-IGF-1R	CHO; DHFR/MTX	Splicing error (intron translation)	No effect on biological activity	Reduced SDS-PAGE, RP-LC with reduced MS analysis of fractions, CEX	Tryptic LC-MS, Edman	(63)
Phe→Leu	11th position from N-terminus	7–10%	Peptide-antibody fusion protein	CHO; GS/MSX	Genetic (somatic mutation) TTC→CTC	Reduced biological activity due to inhibited dimerization	Lys-C LC-UV map	qPCR	(65)
Ser→Asn	Multiple, at AGC codon	0.01–0.2%	IgG1, IgG4	CHO, (also NS0, <i>E. coli</i> )	Mistranslation	Unknown	Tryptic and chymotryptic LC-MS map	Synthetic peptides (spiking)	(60)

*Continued on next page.*

**Table 1. (Continued). Reported Cases of Sequence Variants in Recombinant Proteins Produced in Mammalian Cell Lines<sup>1</sup>**

Variant	Site(s)	Level per site	Molecule	Expression System	Cause	Effect	Detection		Ref
							Primary	Confirmation	
Asn→Ser	Multiple, not codon- or site-specific	1–7%	IgG	CHO	tRNA <sup>Asn</sup> mischarging, due to Asn starvation	Minor decrease in binding for L(Asn35→Ser) mutant	Reduced mass (LC-MS) Lys-C and Asp-N LC-MS map	Synthetic peptides	(67, 68)
Ser→Arg	L(167)	0.8%	IgG1	CHO; DHFR/MTX	Genetic AGT→CGT	Unknown	CEX, <sup>3</sup> tryptic LC-MS map	PCR	(66)
Met→Arg	H(83)	5%	IgG1	CHO; DHFR/ MTX	Not discussed	Unknown	Tryptic LV-UV map		(85)
Pro→Thr	H(274)	42%	IgG1	CHO; DHFR/ MTX	Not discussed	Unknown	Tryptic LC-UV map		(85)
Leu→Gln	H(413)	0.3%	IgG1	CHO; DHFR/ MTX	Not discussed	Unknown	Tryptic and chymotryptic LC-MS map		(85)
Ser→Gly	H(52)	0.2%	IgG1	CHO; DHFR/ MTX	Not discussed	Unknown	Tryptic LC-MS peptide map		(86)

Variant	Site(s)	Level per site	Molecule	Expression System	Cause	Effect	Detection		Ref
							Primary	Confirmation	
Ser→Arg	H(441)	0.3–0.6%	IgG1	CHO	Proposed: genetic AGC→CGC or (AGA or AGG)	Unknown	iCIEF, reduced mass	Tryptic and Lys-C LC-MS peptide map, synthetic peptide (spiking)	(87)
Ala→Ser	L(183) L(152)	7.8–9.9% 0.5–0.6%	IgG4	CHO	L(183): Genetic (GCA→TCA)	unknown	Reduced mass (LC- MS)	Tryptic LC-MS peptide map, qPCR L(183) Synthetic peptide (spiking)	(88)
Thr→Asn	H(Thr24)	2.7–4.0%	IgG1	CHO	Genetic (ACC→AAC)	Not discussed	Tryptic LC-MS map, differential analysis <sup>4</sup>	NGS	(70)
Phe→Tyr (Leu/Ile)	Multiple, not codon- specific	0.3–0.6%	IgG1	CHO	Misincorpora- tion (Phe star- vation)	Not discussed	Tryptic LC-MS map, differential analysis <sup>4</sup>		(70)

*Continued on next page.*



**Table 1. (Continued). Reported Cases of Sequence Variants in Recombinant Proteins Produced in Mammalian Cell Lines<sup>1</sup>**

Variant	Site(s)	Level per site	Molecule	Expression System	Cause	Effect	Detection		Ref
							Primary	Confirmation	
Stop221→Glu	L(C-terminus) 17-aa extension	13.6%	IgG1	CHO	Genetic (Stop codon TAA→GAA)	Not discussed	Tryptic and chymotryptic LC-UV and LC-MS	Reduced LC-MS, Edman, and qPCR	(64)
Tyr→Phe (His)	Multiple	≤3%	IgG1	CHO DHFR/ MTX, GS/MSX	Mischarging tRNA <sup>Tyr</sup> with Phe		Tryptic and thermolysine LC-MS map		(69)

AA, amino acid; AAA, amino acid analysis; CEX, cation exchange chromatography; CHO, Chinese hamster ovary; DHFR, dihydrofolate reductase; GS, glutamine synthetase; *Escherichia coli*, *E. coli*; H, heavy chain; iCIEF, imaged capillary isoelectric focusing; L, light chain; LC-MS, liquid chromatography-mass spectrometry; LC-UV, liquid chromatography with UV detection; MS, mass spectrometry; MTX, methotrexate; N/A, not available; NGS, next generation sequencing; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; RP, reversed phase; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis. <sup>1</sup> Additional examples can be found in Ref (26). <sup>2</sup> Amino acid position indicated in parenthesis. <sup>3</sup> Described in Ref (94). <sup>4</sup> Differential analysis of LC-MS peptide maps using SIEVE software.

In 1993, Harris et al. reported Tyr376→Gln heavy chain variant in CHO-produced anti-Her2 antibody, with levels of the variant molecules reaching up to 27% (61). The origin of the mutation was traced back to the polyclonality of the original transfected pools produced using the DHFR/MTX expression system (61), and the levels of the mutant molecules were found to decrease with cell age, suggesting a difference in the stabilities of normal and mutant cell lines. A crossover event during expression of another IgG molecule occurred between Lys107 of the variable region of the light chain and Ser125 of the constant region of the heavy chain, connected via a Pro residue (62). Splicing between the CGA and the TCC codons encoding for Arg108 of the light and Thr126 of the heavy chains, respectively, produced a new CCC codon encoding for Pro. Genetic rearrangement was identified as the likely mechanism of the unusual recombination event. More than 50% of this anti-IGF-1R IgG1 molecule contained an additional 24-amino acid sequence, inserted between the variable and constant domains of the heavy chain, likely due to intron splicing errors during maturation of the mRNA (63).

Under normal protein expression conditions, a stop codon signals termination of translation at the ribosomal subunits by binding release factors, causing a subsequent release of the polypeptide chain. In one example, a point mutation of a stop codon, TAA, led to the creation of a GAA codon (coding for Glu) in the gene coding for the light chain of an IgG1 molecule. As a consequence, readthrough translation occurred beyond the mutated stop codon until the next alternative in-frame stop codon was reached. This resulted in the insertion of a Glu residue at the C-terminus of the light chain and its extension with 17 additional amino acids, or approximately 14% of the light chain of that molecule (64).

DNA mutations have also been reported to lead to relatively high incidence of errors in protein synthesis. For example, the Phe→Leu variant in the 11th N-terminal position, ranging from 7% to 10% in abundance, in an IgG fusion protein was determined to be due to DNA missense mutation of a TTC codon to a CTC codon (65). Comparison of several cell lines revealed that the probability of spontaneous mutation increased with the number of gene copies transfected into the cells. MTX used in DHFR-based stable cell line selection and amplification systems, with a goal to enhance productivity, was linked to the occurrence of DNA mutations (66). A possibility of increased mutation rates increases during amplification processes and seems to be inevitable during stable cell line development.

Missense mutations that occur on the DNA level are statistically improbable to occur at multiple codons at once. In contrast, mistranslation, often manifested by incorporation of erroneous amino acids at multiple sites of a protein, serve as the signature pattern of translational errors. These differences, in principal, can allow researchers to discriminate between the two main sources of sequence variants in their products. Mistranslations, as discussed above, can be due to mischarging of tRNA with an erroneous amino acid by its cognate AARS, or by a misreading event due to the anticodon of the tRNA mismatch with the codon of mRNA.

Low levels of Ser→Asn misincorporation, ranging from 0.01% to 0.2%, were detected across two IgG1 molecules and one IgG4 molecule to occur at multiple sites encoded by the AGC DNA codon (60). Furthermore, Ser→Asn misincorporations were found to be independent of whether a mAb was

expressed from CHO, NS0, or even *E. coli* cell line hosts, but was affected by cell culture conditions. In the subsequent study, replacement of the AGC codon encoding for Ser63 by either TCC or TCT codons was demonstrated to eliminate the mistranslation (i.e. Ser63→Asn variant) (66). Reverse Asn→Ser misincorporation events, occurring at higher rates of above 1%, were reported to be unrelated of the nature of expressed protein, but were affected by cell culture conditions and, in contrast, exhibited no DNA codon preference (67). A connection between the growth rates and mistranslation in which reduced Asn levels (starvation) in the culture medium was linked with elevated levels of Asn→Ser mistranslation, suggested possible scenarios for the mistranslation to be due either to misreading, where tRNA<sup>Ser</sup> substitutes depleted tRNA<sup>Asn</sup> via codon-anticodon mismatch; or mischarging, where Ser successfully competes for charging onto tRNA<sup>Asn</sup> (68). Supplementing the cell culture medium with adequate amounts of Asn were effective to eliminate the misincorporation (67). A different study demonstrated a strong impact of the producing cell line on the levels of Asn→Ser misincorporation, suggesting that both cell lines and cell culture conditions affect the levels of amino acid substitution (26).

In five separate mAbs produced in CHO, Tyr→Phe sequence variants at multiple locations and at levels  $\leq 3\%$  were detected, independent of the cell line host and selection system (DHFR/MTX or GS/MSX) (69). No link was observed between the Tyr location and the magnitude of these variants and the Tyr codon used (TAC or TAT). However, the misincorporation was found to strongly correlate with the extracellular Tyr concentration, and when it dropped below the threshold concentration of 0.02 mM, the sequence variants were observed. In contrast, excess extracellular Tyr ( $>1$  mM) prevented the formation of the sequence variants at Tyr positions. The data suggested that tRNA<sup>Tyr</sup> mischarging due to the structural similarities between Tyr and Phe is the most likely reason for the occurrence of Tyr→Phe substitution. These studies allowed researchers to develop appropriate feeding strategies aimed at avoiding Tyr starvation conditions and the resulting formation of the sequence variants, while maintaining good productivity of cell lines.

An opposite Phe→Tyr substitution was detected in several production cell clones originating from independently transfected cells at multiple Phe locations (70). PheRS belongs to the class II AARS enzymes, which exhibit both pre- and post-transfer editing (71). However, under Phe starvation conditions, mischarging of its cognate tRNA<sup>Phe</sup> with Tyr can likely be the root cause of this mistranslation event. Further evidence that Tyr misincorporation can be prevented by supplementing cell culture to maintain Phe concentration above 1 mM supports mischarging as the cause of these errors.

The above discussion suggests that the phenomenon of sequence variants in mammalian cell lines (i.e., CHO) is similar to that occurring in bacterial cells (i.e., *E. coli*) and needs to be carefully considered during the development of recombinant biotherapeutics.

## Impact of Expression Errors on Quality of Recombinant Proteins

The impact of translational errors on the properties and function of recombinant proteins is not known *a priori* and is difficult to predict. The effect needs to be examined on a case-by-case basis. A general understanding of the often deleterious effects that the *in vivo* translational errors have on living organisms are documented, but these effects may not necessarily apply equally to biotherapeutic proteins, for which consideration of an indication, bioavailability, clearance mechanisms, dose, and route and frequency of administration are also important factors.

The impact that a particular variant has on the efficacy of a biotherapeutic is an important concern of scientists who deal with occurrence of sequence variants. A great deal of structural and functional knowledge has been accumulated on mAbs, their mechanisms of action (72), and the role of their intrinsic heterogeneity during production in mammalian cell lines (10, 11, 73). This impressive body of knowledge often helps researchers estimate the possible impact of a sequence variant, depending on its location within the sequence of a mAb. For example, occurrence of a sequence variant within the variable domains of a mAb, which contains its complementarity-determining regions (CDRs), raises a concern that antigen binding affinity may be affected. Likewise, sequence variants in the constant domains of a mAb can potentially affect its binding with neonatal Fc receptor (FcRn) and, thus, alter its bioavailability or its effector functions, exhibited through interaction with complement and Fc $\gamma$  receptors. Mutational scanning, based on the site-directed mutagenesis of solvent-exposed amino acids to Ala, was used to map the binding site on human IgG1 for the various Fc receptors, allowing prediction of the potential effect of amino acid substitutions on binding (74, 75). However, evidence exists that mAbs that have identical variable regions, but differ in isotype can manifest variations in their fine specificity and functional affinity as a consequence of their associated constant heavy chain regions (76). This suggests that composition of the constant region affects the secondary structure of the antigen binding sites, thus affecting its specificity. The result implies that the effect of a sequence variant can potentially manifest itself in a domain different from its immediate location, making predictions, based on a primary structure information alone, less reliable. Furthermore, extrapolating from the Walsh and Jefferis (13) discussion on PTMs in biopharmaceuticals, the effect of a modification (PTM or sequence variant) must be evaluated from the point of view of its functional and safety consequences, which are often determined only through clinical evaluation, rather than from the structure of that modification alone. Irrespective of the level and nature, a modification by itself is not as critical because its importance depends solely on the interaction it has with a biological system.

Several studies have explored the effect of sequence variants on properties of proteins. It was shown that His $\rightarrow$ Gln variants of hG-CSF exhibited slightly lower pI values but maintained full *in vitro* biological activity (77). The ability of a Tyr136 $\rightarrow$ Cys  $\alpha$ -synuclein variant to form filaments *in vitro* was not changed; however, dimerization of the mutant protein was shown to significantly increase through disulfide bond formation (34), which potentially can accelerate

protein aggregation and cellular toxicity of  $\alpha$ -synuclein (78). Both Arg- and Lys-containing variants of human HsTIM were reported to be nearly equally catalytically active, but differed in thermostability and susceptibility to urea and proteinase K (41). Functionality of recombinant hemoglobin, assessed by its oxygen affinity and cooperativity, was not disrupted by the presence of 0.73% Leu→Nva variant species when compared with the variant-free control (54).

In multiple cases of mAbs, the effect of a single amino acid change, regardless of its translational or post-translational origin, on the properties of the molecules has been studied. For example, Harris et al. determined that post-translational isomerization of Asp102 in a heavy chain CDR3 region of anti-HER2 IgG1 reduced its potency to 70%, causing serious implications on drug efficacy (79). However, *in vitro* potency (antigen binding) and Fc effector function assays for the same mAb were equivalent for the preparations containing from 1% to 27% of the Tyr376→Gln heavy chain variant (61). In another example, a minor decrease in antigen binding affinity was observed for the light chain containing an Asn35→Ser sequence variant when compared with the correctly expected mAb (67). Dorai et al. determined that a Phe→Leu mutant of the fusion protein (FP) significantly inhibited its propensity to dimerization compared with the expected FP, potentially affecting biological activity of the variant product (65).

In addition to activity, safety to patients is another concern for sequence variants during the development of recombinant protein therapeutics. A potential immunogenic response to a recombinant product and its variants containing missense errors when administered to humans is of an even greater concern, likely exceeding the concern of their possible abnormal bioactivity. A therapeutic protein can be immunogenic because the human immune system categorizes it as non-self (80). The degree of foreignness of the biopharmaceutical protein compared with the natural endogenous proteins affects the immunogenicity of that product and can be influenced by higher-order structure and aggregation state, glycosylation, including the presence of non-human glycan structures, and chemical modifications, including intended modifications and unintended degradation. For example, the impact of isomerization on immunogenicity has been described for tyrosinase-related protein (TRP)-2, where conversion of an aspartic acid to an isoaspartic acid triggered a strong immune response (81). The concern of immunogenicity induced by sequence variants was raised early on by Harris et al. when discussing possible consequences of the Tyr→Gln substitution in the humanized anti-HER2 mAb (61). It was also acknowledged that the potential contribution to an immune response in humans would have been difficult to assign given that roughly 5% of the protein already had a non-human (murine) antibody sequence. Immune responses to therapeutic protein products may pose problems for both patient safety and product efficacy (82). Consequences of immune responses to therapeutic proteins can range from no apparent effect to serious adverse events, including life-threatening complications, neutralization of the effectiveness of therapies, or neutralization of endogenous proteins with non-redundant functions.

Allotypes of IgG proteins are identified by unique epitope(s), recognized by unique serologic reagent(s), and are defined according to the polymorphism in the sequence of the constant region of these molecules (83). Although in most

cases several amino acid substitutions in the constant regions of IgG lead to a change of the allotype, even a single amino acid difference can give rise to an allotypic determinant. This very fact of such polymorphism suggests that allotypic variants can be immunogenic to patients heterozygous for the given allotype. It was proposed that the preexisting antibodies to the allotype that are present on cetuximab, or such antibodies induced in response to its administration, could contribute to resistance to the therapy, thereby manifesting potential importance of allotypes on immunogenicity of proteins (84). Extrapolating the discussion onto the sequence variants of biotherapeutics, it would be difficult to know *a priori* how the presence of a sequence variant could affect immunogenic response in patients. Immune responses to therapeutic proteins and their variants are hard to predict based solely on the knowledge and characterization data, and often require animal testing and clinical data.

Various biological characterization assays are routinely used during development to assess biological activity, antigen specificity, and Fc functionality of recombinant mAbs and their variants, including sequence variants. In practice, however, the effect of low-level sequence variants can be hard to assess due to the natural variability of typical potency assessment bioassays, which, in general, exhibit greater variability than do chemically based methods. In addition, impact and individual properties of such low-level variants are likely to be hindered by an overwhelming background of the major form of that product. Efforts to purify or otherwise manufacture the variant molecule are needed to properly assess its safety and efficacy characteristics by a direct comparison with the correctly expressed protein. In addition to a potential need to introduce control systems for monitoring variant levels during production, these extra efforts would complicate the development of the biotherapeutic product leading to potential delays in timelines, in filing regulatory submission to the government agencies, and in availability and cost of that product to patients in need. Needless to say, it is beneficial to avoid having sequence variants in a biotherapeutic product in the first place; the absence of sequence variants means fewer manufacturing issues during the development of the product. This is likely the reason why most published reports do not detail the effects of variants to a great extent, but agree in suggesting to avoid having sequence variants in biopharmaceutical products (60–70, 85–88).

The previous discussion demonstrates how the genetic sequence of recombinant proteins produced in living cells can undergo mutations that could alter the properties of the protein, presenting potential adverse consequences to patients. To ensure safety and efficacy of a biotherapeutic product and assess its manufacturing consistency, a combination of analytical procedures is used to determine the purity and heterogeneity profile. The ICH guidelines (89) recommend confirmation that the correct coding sequence of the product has been incorporated into the host cell and is maintained during culture to the end of production by means of the proper characterization of the expression construct. In addition, the evaluation of the production stability of the cell line needs to be assessed during the course of a number of population doublings to establish a limit for *in vitro* cell age. The importance of this exercise is evident considering potential occurrence of a phenotypic shift in response to environmental factors

and the cell line age. This implies that the selected cell line should be evaluated for the presence of unintended sequence variants. For that purpose, protein and nucleic acid analytical techniques can be used to assess and verify the amino acid sequence of the expressed protein. Regardless of the significance of a particular variant on the safety and efficacy of the product, one has to be able to detect the sequence variant in the first place.

The fact that underlying biosynthetic processes used for production of recombinant proteins have errors associated with them strongly suggests that a biotherapeutic protein that is free of sequence variants likely does not exist. Thus, detection of biosynthetic errors is a prime step in the process of manufacturing high-quality recombinant protein.

### **Detecting Sequence Variants: Advantage of Using Multiple Orthogonal Methods**

Literature reports on the effect of downstream purification schemes on levels of sequence variants are scarce. It was reported that variants with His→Gln substitution at multiple positions in recombinant human granulocyte colony stimulating factor (hG-CSF), were chromatographically separated during purification to beyond detectable levels in the final purified form of that protein (77). However, the fact that a single amino acid change, constituting, for example, less than 0.1% difference of a typical full-length mAb, can induce only a small effect on the bulk chemical and physical properties of a protein, can challenge the effectiveness of separation by downstream purification of the variant isoforms from the desired product, especially if presence of a sequence variant is not known *a priori*. Hutterer et al. demonstrated that levels of sequence variants in the Fc-fusion protein following a single-step purification by the Protein A affinity column were representative of those in the crude product (44). Furthermore, single- versus three-step purification methods differently affected two types of sequence variants; although Gly→Glu variants decrease two-fold between the two purification steps, almost no change was observed in the levels of Arg→Gln variants. The result demonstrates that one cannot blindly trust that downstream purification completely removes all of the sequence variant isoforms.

A rigorous testing strategy must therefore be implemented to ensure proper clearance and/or consistency of the appearance of sequence variants. For example, various analytical methods were used to analyze translational variants of anti-IGF-1R IgG1 molecule, expressed in CHO and NS0 cell lines (63). The difference between the two molecules was that the CHO-expressed anti-IGF-1R contained an additional 24-amino acid intron sequence, whereas the NS0-expressed mAb lacked this variant. Intact analysis by non-reduced sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and size-exclusion chromatography (SEC) did not reveal any differences between the expected and the variant molecules, having mass difference of 2%. However, the 6% mass difference between the expected and variant heavy chains was hinted at by examining reduced and alkylated SDS-PAGE profiles. Likewise, the heavy chain variant, containing 16 amino acids fewer in its sequence, was successfully

separated from the expected heavy chain and detected by reduced SDS-PAGE (62).

Detection of sequence variants by reversed-phase high-pressure liquid chromatography (RP-HPLC) methods has also been reported. Due to the large size and relatively high hydrophobicity of mAbs, however, separation of variant isoforms on an intact antibody level is hard to achieve. RP-HPLC separation of the anti-IGF-1R splicing variant isoform with an additional 24 amino acids was only possible by employing a rather unique condition of 10% acetic acid in the mobile phase that resolved variant isoforms, enabled their quantitation, and provided fractions for subsequent analysis by mass spectrometry (MS) (63). An even better separation of variant isoforms by RP-HPLC methods was demonstrated for reduced or otherwise fragmented antibodies (63). At the subunit level, Fu et al. also reported improved RP-HPLC resolution of Ala→Ser light chain variant at levels of 8% to 10%, which was detected as a back shoulder peak of the main light chain peak of an IgG4 molecule (88). Although an enhanced analytical RP-HPLC method for analysis of mAbs on intact and fragment levels was developed and successfully applied to monitor the stability and production of mAbs (90), it has not been tested to detect sequence variants.

Among the chromatographic techniques, RP-HPLC is the most compatible with MS detection due to the use of primarily volatile buffers. RP conditions for the best chromatographic separation often optimize at faster flow rates and require the use of mobile phase modifiers such as trifluoroacetic acid (TFA), however, contradicting with the most optimal conditions for MS detection. In a typical scenario, a compromise between the conditions for the best separation and detection by MS needs to be established and utilized. Also, because the separation of closely related variants on an intact and a large polypeptide by RP methods is hard to achieve, simultaneous detection of multiple components in a single spectrum by MS is only feasible when the mass difference of the species exceeds mass resolution and when the relative abundances of these species fall within the dynamic range of a mass spectrometer. For example, detection of Ala→Ser variants with mass difference of 16 Da was not achieved by intact mass analysis using a quadrupole time-of-flight (qTOF) instrument with resolving power of 10,000, even when variants had relative abundance of 8% to 10% (88). In another study, the effect of the size of a polypeptide chain on detection of Asn→Ser variants with mass change of -27 Da was estimated to be at variant levels of  $\geq 2\%$  for the intact heavy chain, of  $\geq 0.5\%$  for the intact light chain, and of  $\geq 0.025\%$  for an average-sized Lys-C peptide fragment (67).

An alternative to intact and/or subunit analysis methods is to use enzymatic digests of the protein. A comparative evaluation of peptide fragments of therapeutic proteins, which is a more sensitive tool to detect subtle differences between samples, is often accomplished by (high-pressure) liquid chromatography with ultraviolet detection (LC-UV) for peptide mapping. Depending on a specific goal and the development phase of the biotherapeutic, a reference point for the comparison can be established by the use of a reference standard, if available, another clone, or a different fraction of the same molecule gathered by an orthogonal separation method. Because most peaks between these samples are the same, even small differences between the samples can be accurately identified.



By conducting spiking studies of one mAb into another, Yang et al. demonstrated that peptide mapping can detect variant peaks with relative abundances of  $\geq 5\%$  with good confidence and of 1% in about half of the cases, using LC-UV detection at a wavelength of 214 nm (85). However, despite good integrity and detectability down to 1% levels attainable by LC-UV peptide mapping method, detection of sequence variants from a single-enzyme peptide map remains a challenge and can be subjected to a lack of resolution of variant peptides or a co-elution with another major peptide. In another study, detection of hypothetical mutations by spiking with two synthetic peptides containing Tyr $\rightarrow$ Gln substitution into the tryptic digest of an antibody was shown to have detection limits of 2% for the peptide eluting in a clear flat region of the LC-UV profile, and only 15% for the peptide which eluted as a shoulder on the descending slope of an existing sequence peak (91). The visual comparison of LC-UV profiles of Lys-C peptide maps did not reveal obvious differences, however, between the variant-free mAb and the sample with less than 7% of Asn $\rightarrow$ Ser variants (67). In another example, a tryptic LC-UV peptide map did not provide the sensitivity to detect the intron-sequence-containing and abundant splicing variant peptide, which co-eluted and was masked by a broad peak due to elution of glycopeptides (63). Finally, the sensitivity of an LC-UV method was not sufficient to reveal the presence of a Ser52 $\rightarrow$ Gly sequence variant at the peptide level of 0.2% in the heavy chain of an IgG1 (86).

It was speculated that the improved detection of variant peptides, in principle, can be achieved by changing separation conditions, such as the gradient and mobile phase composition, of any given LC-UV peptide mapping method, targeting changes to shuffle retention specificities of peptides and causing them to elute in “un-occupied” locations of the chromatogram (92). Analysis of fractions collected from orthogonal methods, such as ion exchange chromatography (IEC) or RP-LC, is a means of sample enrichment with variants of interest. Subsequent interrogation by reduced MS or peptide mapping is an excellent approach to improve detection and, at the same time, provide ground for intra-fraction comparison (87, 88).

Charge-based separation methods, such as IEC and capillary isoelectric focusing (cIEF) are platform methods used to monitor the microheterogeneity of mAb products due to the presence of charge isoforms (93). When compared, the IEC profiles of anti-IGF-1R mAbs with the expected sequence and the product containing the splicing variant with additional 24 amino acids, exhibited non-specific but noticeable differences (63). Separation of charge variants, especially in the basic region of the profiles, is well-established to be very sensitive to a potential sequence variant (63, 79). It should be noted, however, that the sensitivity of these methods is based on the change in charge and not directly due to the change in amino acid composition. Isoforms that do not carry charge differences cannot be effectively resolved by charge-based methods.

Both IEC and cIEF can be extremely effective in detecting low levels of those sequence variants that induce charge differences. For example, IEC successfully separated variant IgG1 molecule with the single Ser $\rightarrow$ Arg mutation in the light chain with abundance of less than 1% from the peak for the major product isoform (94). Similarly, 0.3% to 0.6% Ser $\rightarrow$ Arg sequence variant was detected in the basic

region of the imaged cIEF profile of an IgG1 molecule (87). Moreover, in both cases, IEC and cIEF served as the front-line methods in detecting low levels of these sequence variants. To our best knowledge, these cases are the only published examples where detection of sequence variants at levels less than 1% was achieved for an intact antibody.

In some early characterization work, peptide analyses were often carried out using amino acid analysis (AAA) and N-terminal sequencing by Edman degradation (61). N-terminal sequencing was also used to identify the crossover event between the variable region genes of the antibody, as discussed by Wan et al. (62), and to confirm the intronic translation origin of the variant peptide in the anti-IGF-1R example (63). As a general trend, however, the once-mighty Edman degradation sequencing has been recently challenged by on-line detection with MS because of its excellent capability for high-throughput peptide analysis. Nevertheless, N-terminal sequencing still remains a powerful technology in cases when translational errors lead to proteins' sequence modifications that cannot be easily deduced from examining the amino acid sequence of that protein. For example, Edman was the best-choice sequencing method to identify the variant peptide due to the stop codon mutation event (see Table 1), proving the high value of the sequencing technology (64). It is also the method of choice for the determination of substitutions of Leu and Ile with each other, because they are identical in elemental composition and cannot be distinguished by mass measurement alone.

Popular molecular biology methods, such as polymerase chain reaction (PCR) and qPCR (quantitative polymerase chain reaction), have been successful in confirming occurrence of DNA mutations in biological products (see Table 1 for examples). These methods can provide detection of down to 0.1% of single-nucleotide substitution for a target region of DNA sequence, which requires the design of specific primers, limiting their application to verification of a specific variant (70). Several new DNA sequencing technologies have emerged to show promise for providing fast and cost effective genomic sequencing solutions. The so-called next-generation sequencing (NGS) technologies were reviewed in a number of recent articles (95–97). The NGS technology was demonstrated to detect a 2% sequence variant of a recombinant IgG due to a single nucleotide substitution, with a potential to probe DNA integrity *de novo* down to 0.5% variation levels (70).

## Evolution of Peptide Mapping Methods

Recognition of the molecular origin of disease promoted the development of analytical methods to study proteins. The first successful application of peptide analysis was identification of differences between the normal and the sickle-cell anemia hemoglobins, achieved in 1956 by Ingram, who discovered that the difference between the two proteins was due to a single amino acid on one of the tryptic peptides (98). In its original version, peptide analyses were conducted in a two-dimensional format with a combination of paper electrophoretic and paper chromatographic techniques. Such “fingerprints” consisted of a number of peptide spots arranged (mapped) in a two-dimensional pattern on a filter paper or

a thin-layer cellulose sheet. Introduction of RP-HPLC with ion-pairing reagents in 1970s by Hancock et al. forever changed the way the analysis of proteins and peptides were conducted (99). The term “peptide map,” however, stayed with the method.

Development of soft-ionization techniques, such as fast-atom bombardment, first, and, later, electrospray ionization (ESI), for MS enabled direct coupling of HPLC to the MS detector (100–104). The synergy of the two powerful methods made peptide mapping a highly informative approach to study and identify proteins. Early protein identification approaches were based on detected masses of corresponding peptides and their matching to database entries (105). Further refinements of tandem mass spectrometry (MS/MS) methods and data-dependent acquisition algorithms added an extra level of confidence to peptide assignments by sequencing peptides in the gas-phase during MS/MS experiments (106). Algorithms for matching experimental (high-pressure) LC-MS data with known protein database entries were developed to allow automated identification of proteins and their major PTMs. Among these, SEQUEST (107) and Mascot (108) are two popular programs for database searching used for automated processing of large datasets. Realization of the importance of genetic variation in phenotype prompted development of capabilities to identify peptides containing amino acid substitutions resulting from a SNP (109). Extension of capabilities to search for unknown modifications of peptides, driven by the desire to interpret unmatched peptide spectra, lead to the development of Mascot error-tolerant search (ETS), which is also capable of handling multi-vendor data file formats (110). Benefiting from high mass accuracy attainable by modern high-resolution mass spectrometers (111), ETS is a powerful approach for identification of unexpected PTMs and sequence variants.

### **Role of Peptide Mapping with Liquid Chromatography-Mass Spectrometry**

Peptide mapping is an effective method for evaluating the amino acid sequence and PTMs of proteins and is now used routinely during many stages of pharmaceutical development (65). As analytical technologies become more sensitive, it is likely that products previously believed to be homogeneous will be revealed to contain some microheterogeneity, at least due to the basal error of biosynthesis. For example, by applying high-resolution MS to peptide mapping, Yu et al. reported identification of translational errors with frequencies of  $10^{-3}$  to  $10^{-4}$ , nearing reported errors of cellular processes (60).

Reversed phase separations with 1.5- to 3-hour-long gradients, perfusing 2.1 mm i.d. analytical HPLC or UPLC (ultra performance liquid chromatography) column, packed with 3  $\mu\text{m}$  or  $\leq 2\mu\text{m}$  particles, respectively, are typically employed for evaluation of the integrity of the amino acid sequence of expressed mAbs and are well-suited for detection of sequence variants (85, 88). Such peptide maps are targeted to have enough peak capacity to adequately separate peptides of mAb heavy and light chains with greater than 95% sequence coverage, and to provide enough “peak-free space” in case an unexpected variant is present in the sample. The endopeptidase trypsin is commonly used because it predictably and specifically cleaves both polypeptide chains of a typical mAb into fragment

peptides that have typical molecular masses in the range of 700 and 2500 Da. Tryptic peptides also contain a basic residue, Arg or Lys, at the C-terminus, making them well-suited for MS and MS/MS detection. Typical optimization strategies of digestion conditions target minimization of artifacts induced by sample preparation, such as Asn deamidation, N-terminal Gln cyclization, non-specific cleavages, and secondary products of Cys alkylation (112).

Although trypsin is often the enzyme of choice, the use of other enzymes has been exploited. Employment of a second enzyme peptide map was suggested to improve overall sequence coverage and to enable more complete detection of potential sequence variants at any residue. Such a “two-mapping” method provided researchers with 98.5% sequence coverage for a mAb sample spiked at the level of 1% (85). The published choices of the secondary enzymes include endopeptidases Lys-C (67), chymotrypsin (85), Glu-C (64), Asp-N (88), and thermolysin, the choice of which depends on a particular task or the need to improve coverage of a specific region missing from the primary enzyme peptide map. Reliable detection and identification for peptides is reported for variants with relative abundances of 0.5% (85) and even 0.01% (60).

Quantitation based on MS, in principle, is subject to ionization differences between the normal and variant peptides, the dynamic range of the mass spectrometer, and potential ion signal suppression in cases when peptides of interest co-elute with other peptides. For example, the amounts of Ser variant peptides relative to Asn-containing peptides were reported to be within 15% to 20% higher than values determined by spiking experiments (67), a discrepancy caused partially by better (about 12%) ionization efficiencies of Ser peptide variants and by signal saturation of abundant Asn peptides. Ionization efficiencies of Gly- and Glu-containing variant peptide were shown to be within two-fold of one another, likely due to a change in charge distribution induced by replacement of Gly with the acidic Glu residue (45). However, conservative substitutions of a single amino acid, without causing a significant change in charge or hydrophobicity of a peptide, showed good correlation with true spiked levels of low abundant peptide (85).

Analyses for sequence variants in biotherapeutic proteins, requiring examination of a large number of spectra (often in thousands), can be compared with searching for “a needle in a haystack.” Manual processing of thousands of spectra, produced by modern mass spectrometers, is not practical. Mascot ETS enabled the detection and identification of unknown variants in biopharmaceutical products and became a standard tool in many laboratories in the industry for detection of low-level sequence variants and other modifications in proteins. In fact, the first report on detection of 0.01% of Ser→Asn variant would not have been possible without the use of the software capable of handling large-volume data processing, consisting of tens of thousands of MS/MS spectra in a single data file. Nevertheless, the approach has several limitations. First, the search considers only a single amino acid variation in a peptide (minor change), when compared with its expected sequence in the database. When a modification induces variations involving changes at two or more sites (significant change) in a peptide, the ETS approach will most likely fail to detect these modifications. In the above cases, significant translational errors due to the stop codon readthrough

(64), DNA recombination (62), or intron translation (63) would not be detectable by the ETS search. Detection of the Tyr→Gln missense mutant, caused by a two-base change (TAT→CAA), would challenge data analysis algorithms based on the SNP principals. Second, a frequent complaint is that Mascot ETS contains a large amount of false positives, making data analysis a rather tedious process and requiring extensive manual verification of potential hits.

Mascot software provides a score upon which the “correctness” of a peptide’s assignment to a sequence in the database is based. This score is in part a reflection of the quality of the collision-induced dissociation (CID) data for that peptide. In other words, it reflects the number of sequence information-bearing *b* and *y* fragment ions, originating from cleavages along the peptide’s backbone. However, not all peptides produce good-quality spectra, which often makes it difficult or even impossible to unambiguously identify a peptide. For example, CID fragmentation of peptides containing Pro residues or Pro-Asp (Glu) motifs are known to provide less-informative spectra, making CID spectra for these peptides difficult to interpret. Moreover, assignment of an exact location of a sequence variant in a peptide is often challenged by the lack of sequence-indicative fragment ions or by sequence gaps, because of which the position of that variant cannot be unambiguously assessed. Also, a number of amino acid substitutions are isomeric to each other or to frequent modifications. For example, mass change,  $\Delta M$ , for Ala→Gly = Ser→Thr = Leu→Val = Glu→Asp = Gln→Asn = -14.0157 Da; mass change for Ala→Ser = Phe→Tyr = +15.9949 Da = oxidation; and mass change for Ala→Glu = Gly→Asp = +58.0055 Da = Cys carboxymethylation, which is a frequently used Cys-capping modification. A comprehensive list of chemical modifications and digestion and ionization artifacts, which potentially can be misinterpreted as sequence variants, is provided by Zhang et al. (supporting information in Ref. (26)).

Several novel informatics approaches were recently developed targeting the analysis of biotherapeutics by LC-MS. Zhang et al. developed a computer algorithm for automated analysis of a large-scale LC-MS dataset for identification and quantitation of known and unknown covalent modifications (including sequence variants) by comparing the experimental fragmentation spectrum to the predicted spectrum of each native or modified peptides and on-the-fly determination of their relative peak areas (113). In a second approach, a wildcard search allows the hunt for unanticipated or even unknown modifications alongside known modifications. This algorithm also takes advantage of recalibrating the LC-MS data file on-the-fly for improved mass measurement accuracy (114).

In addition to the automated analysis and interpretation of CID data, a complex work flow for sequence variants analysis by LC-MS takes full advantage of information-rich MS-level data (70). The unified approach aims to overcome limitations of conventional database searching with Mascot ETS and a potential lack or poor quality of MS/MS spectra for a potential variant peptide. For that purpose, a commercially available, label-free quantitation software was successfully employed by Zeck et al. (70) for comparative evaluation of two or more related samples, used as biological references to each other.

The main focus of analytical chemistry during the cell line development is to verify, with a good degree of confidence, that it is free from genetic mutations

and that it is safe to proceed with its further development and establishment of the master cell bank. As a standard practice, several top-performing clones are subjected to the scrutiny of sequence variant analysis at early phases of product development. At that time, evaluation is often conducted on an individual basis by LC-UV and/or LC-MS peptide mapping methods, as well as by comparison of all the tested clones, either by visual inspection of overlaid LC-UV profiles or by comparative analysis of LC-MS data, as described by Zeck et al. (70). Often, LC-UV and LC-MS data are complementary to each other, and sequence variants not detected by LC-MS, for one or another reason, can be detected from the comparison of LC-UV profiles (64). Comparison of multiple clones from different transfections is believed to be a good approach to detect DNA mutations. DNA mutations are reported to cause significant expression errors, with abundance of variant isoforms ranging from under 1% to more than 50%. However, this type of error is random and is a rare event with negligible chance that the same genetic mutation will occur simultaneously in different and unrelated clones. Thus, a comparative approach is a very powerful method for detection of “a needle in a haystack,” which characterizes the occurrence of genetic errors in biotherapeutic proteins well.

In contrast, translational errors are either codon- or amino-acid-specific, suggesting that these errors occur at multiple locations, involving similar sites of the expressed protein. Despite the evidence that these errors produce small amounts of variant isoforms, typically less than 3% in abundance and lower, their successful detection is, undoubtedly, simplified by their multiplexed presence. Even if one site is missed, other chances for identification of sequence variants of a given type still exist. After successful identification and verification of a sequence variant, all other sites can be retrospectively examined to verify the presence or absence of a potential variant at that site, which was presumably missed in the original search. Unlike genetic mutations, translational errors are often successfully fixed by adjusting cell growth conditions or optimizing cell culture and feeding parameters, as illustrated by several examples described above.

### **Detection of Sequence Variants in the Recombinant IgG1 Antibody from NIST (NISTmAb)**

The NISTmAb standard was received by both laboratories at a concentration of 10 mg/mL in 12.5 mM L-His and 12.5 mM L-His HCl, pH 6.0.

The sequence variant analysis of the NISTmAb (IgG1) was conducted by two independent laboratories, Lab 1 and Lab 2. No prior discussions on the use of methods, experimental conditions, and data analysis protocols were made; each lab operated independently, utilizing their existing best practices for execution of this type of analysis. In both cases, peptide mapping was performed using state-of-the-art, high-resolution LC-MS/MS technology.

## Methods Overview

The general strategy used for the detection of sequence variants included the use of two complementary proteolytic digestions, LC-MS/MS analysis with long-gradient separation, and an error-tolerant database searching strategy where unassigned MS/MS (those that do not match the mass of expected proteolytic peptides) is iteratively searched against a comprehensive list of modifications and amino acid substitutions. Here we present an overview of the methodology, with details on sample preparation, experimental conditions, and data analysis given in Tables 2 and 3.

**Table 2. Sample Preparation and Digestion Details**

<i>Lab 1</i>	<i>Lab 2</i>
<i>Denaturation and Reduction</i>	
Approximately 1 mg of the mAb was diluted in denaturation buffer containing 6 M guanidine HCl, 360 mM Tris and 1 mM EDTA at pH 8.6.	An aliquot of 60 $\mu$ L sample (600 $\mu$ g) was diluted into 517 $\mu$ L of denaturing solution, comprised of 7 M guanidine hydrochloride and 0.1 M Tris at pH 8.2.
The sample was reduced by treatment with 10 mM DTT at 37°C for 1 h.	The sample was reduced by the addition of 14 $\mu$ L of 400 mM DTT at 37°C for 1 h.
<i>Alkylation (Cysteine Capping)</i>	
The sample was alkylated with IAA to a final concentration of 35 mM in the dark for 20 min at room temperature. The reaction was stopped by consuming excess IAA with DTT.	The sample was alkylated with the addition of 12 $\mu$ L of 1 M IAA for 60 min at room temperature in the dark.
<i>Desalting</i>	
The sample was buffer exchanged into digestion buffer containing 25 mM Tris and 1mM CaCl <sub>2</sub> at pH 8.1 using NAP-5 desalting columns.	The sample was buffer exchanged into digestion buffer comprised of 50 mM Tris, pH 8.2, using a NAP-5 (Sephadex G-25 gel) filtration column.
80% (800 $\mu$ L) of the eluent from the desalting column was collected to minimize the collection of residual guanidine.	1 mL of the eluent from the desalting column was collected, to give about 0.5 mg/mL of protein in 50 mM Tris buffer, pH 8.2.
The reduced and carboxymethylated sample was split into two fractions.	The purified sample was split into two equal 500 $\mu$ L fractions for proteolytic digestions.
<i>Enzyme 1 Digestion (Trypsin)</i>	
The sample was digested with sequencing-grade modified trypsin added at a 1:25 enzyme to protein ratio at 37°C for 1.5 h.	The sample was digested with sequencing-grade modified trypsin added at a 1:15 enzyme to protein ratio at 37°C for 5 h.

*Continued on next page.*

**Table 2. (Continued). Sample Preparation and Digestion Details**

<i>Lab 1</i>	<i>Lab 2</i>
The trypsin reaction was quenched by acidifying the solution.	10 $\mu$ L of TFA was added to quench the trypsin reaction.
<i>Enzyme 2 Digestion (Thermolysin)</i>	
The sample was digested with thermolysin enzyme using a 1:100 enzyme to protein ratio at 25°C for 30 min.	The sample was digested with thermolysin enzyme using a 1:50 enzyme to protein ratio at 37°C for 3 h.
The thermolysin reaction was quenched by adding a solution of EDTA.	10 $\mu$ L of TFA was added to quench the thermolysin.

CaCl<sub>2</sub>, calcium chloride; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; HCl, hydrochloric acid; IAA, iodoacetic acid; TFA, trifluoroacetic acid.

The strategy used by both labs for peptide mapping of the NISTmAb included a reduction step under denaturing conditions to reduce the protein into its light and heavy polypeptide chains, followed by alkylation with iodoacetic acid (IAA) to prevent the reformation of disulfide bonds. The general strategy used for the detection of sequence variants included the use of two separate proteolytic enzymes to increase sequence coverage and, in many cases, to confirm assignments. Digestion with a highly specific protease, trypsin, was used as a primary enzyme digest for peptide mapping. Thermolysin was used by both laboratories as their platform secondary enzyme for peptide mapping. Thermolysin is a less-specific, thermostable endopeptidase that hydrolyzes protein bonds N-terminal to the hydrophobic amino acid, preferably Leu, Phe, Ile, Met, and Ala residues (115). Thermolysin has proven to be a better choice for the secondary enzyme over chymotrypsin because it is less prone to cause rearrangement at both termini (*i.e.*, transpeptidation). These sample preparation artifacts are known to be abundant in chymotryptic digests and often complicate data interpretation by introducing false positives (116, 117). Peptides generated from the digestion of mAb with thermolysin are typically smaller in size due to the frequency of hydrophobic residues in the mAb and low specificity of the protease.

Despite all efforts to achieve complete digestion, peptide maps are often found to contain some level of species with partial cleavages (118), which are difficult to minimize even with overnight digestion periods. The presence of these species has minimal interference in data interpretation because they are expected and sometimes predictable (119). However, the presence of some level of artifacts or modifications that has similar mass shifts as an amino acid substitution can interfere with the proper assignment of a Mascot hit. If these artifacts are not controlled for or not anticipated, they can potentially impact the quality of the data, thus compromising the ability of the method to detect real sequence variants.



**Table 3. LC-MS Analysis and Database Searching Details**

Lab 1	Lab 2																																										
<b>Liquid Chromatography-Mass Spectrometry Details</b>																																											
<p>100 <math>\mu</math>L of the digests were loaded onto a Phenomenex Jupiter reversed-phase column (300 <math>\text{\AA}</math> pore size, C18, 5 <math>\mu</math>m, 2.1 <math>\times</math> 250 mm, Phenomenex, Torrance, CA) set at 55°C and analyzed using an Agilent 1200 HPLC operating at a flow rate of 0.25 <math>\mu</math>L/min and monitored using a photodiode array UV detector collecting from 190 to 400 nm.</p> <p>Mobile phase A was 0.1% TFA in water, and Mobile phase B was 0.09% TFA and 90% acetonitrile in water.</p> <p>The total gradient program was 190 min:</p> <table border="1"> <thead> <tr> <th>Time (min)</th> <th>% B</th> </tr> </thead> <tbody> <tr><td>0</td><td>0</td></tr> <tr><td>3</td><td>0</td></tr> <tr><td>23</td><td>10</td></tr> <tr><td>140</td><td>35</td></tr> <tr><td>160</td><td>55</td></tr> <tr><td>162</td><td>95</td></tr> <tr><td>165</td><td>95</td></tr> <tr><td>166</td><td>0</td></tr> <tr><td>190</td><td>0</td></tr> </tbody> </table> <p>Peptides were detected online with LTQ-Orbitrap Elite Velos Pro mass spectrometer operated in positive electrospray ionization mode and calibrated externally to provide a mass accuracy of &lt; 5 ppm. Source conditions were optimized for the HPLC conditions used.</p>	Time (min)	% B	0	0	3	0	23	10	140	35	160	55	162	95	165	95	166	0	190	0	<p>40 <math>\mu</math>L of the digests for each enzyme were loaded onto a Waters Xselect Charged Surface Hybrid reversed-phase column (130 <math>\text{\AA}</math> pore size, C18, 2.5 <math>\mu</math>m, 2.1 <math>\times</math> 150 mm, Waters, Beverly, MA) set at 60°C and analyzed using an Agilent 1200 HPLC operating at a flow rate of 0.20 mL/min and monitored using a photodiode array UV detector collecting from 190 to 400 nm.</p> <p>Mobile phase A was 0.1% FA in water and Mobile phase B was of 0.1% FA in acetonitrile.</p> <p>The total gradient program was 180 min:</p> <table border="1"> <thead> <tr> <th>Time (min)</th> <th>% B</th> </tr> </thead> <tbody> <tr><td>0.0</td><td>1</td></tr> <tr><td>150.0</td><td>30</td></tr> <tr><td>160.0</td><td>65</td></tr> <tr><td>161.0</td><td>98</td></tr> <tr><td>163.0</td><td>98</td></tr> <tr><td>163.5</td><td>1</td></tr> <tr><td>164.0</td><td>98</td></tr> <tr><td>166.0</td><td>98</td></tr> <tr><td>166.1</td><td>1</td></tr> <tr><td>180.0</td><td>1</td></tr> </tbody> </table> <p>Peptide were detected online with LTQ-Orbitrap Elite Velos Pro mass spectrometer operated in positive electrospray ionization mode using an internal lock mass ion of the calibrant hexakis(1H,1H,3H-perfluoropropoxy)phosphazene at m/z 922.009798 via dynamic calibration for mass measurement accuracy of <math>\leq</math> 3 ppm.</p>	Time (min)	% B	0.0	1	150.0	30	160.0	65	161.0	98	163.0	98	163.5	1	164.0	98	166.0	98	166.1	1	180.0	1
Time (min)	% B																																										
0	0																																										
3	0																																										
23	10																																										
140	35																																										
160	55																																										
162	95																																										
165	95																																										
166	0																																										
190	0																																										
Time (min)	% B																																										
0.0	1																																										
150.0	30																																										
160.0	65																																										
161.0	98																																										
163.0	98																																										
163.5	1																																										
164.0	98																																										
166.0	98																																										
166.1	1																																										
180.0	1																																										
<b>Mass Spectrometry Conditions</b>																																											
<p><b>MS1 Conditions:</b> scan range 200 to 2000 m/z and 60,000 resolution (at 400 m/z).</p> <p><b>MS2 Conditions:</b> low-resolution CID in LTQ using 2.5 amu isolation window and 35% normalized collision energy.</p> <p><b>Data-Dependent Settings:</b> 1,000 counts for MS2 trigger, top 5 precursors, dynamic exclusion with a repeat count of 4 and exclusion for 45 seconds.</p> <p>Unassigned charge state and singly-charged ions were rejected from triggering MS2.</p>	<p><b>MS1 Conditions:</b> scan range 235 to 2000 m/z and 240,000 resolution (at 400 m/z).</p> <p><b>MS2 Conditions, Run 1:</b> high-resolution CID in LTQ followed by mass analysis in the Orbitrap with resolution of 30,000.</p> <p><b>Data-Dependent Settings, Run 1:</b> 40,000 counts for MS2 trigger, top 5 precursors. Data used to generate exclusion list for MS2 Run 2 condition.</p> <p><b>MS2 Conditions, Run 2:</b> low-resolution CID in LTQ.</p> <p><b>Data Dependent Settings, Run 2:</b> 3,000 counts for MS2 trigger, top 10 precursors, using exclusion list from Run 1.</p> <p>Doubly and triply charged precursor ions were allowed for MS2.</p>																																										
<b>Data Analysis/Database Searching.</b>																																											
<p>Peak list generated with Mascot Distiller and processed with Mascot v. 2.4 as follows:</p> <p><b>Standard search parameters:</b></p> <ul style="list-style-type: none"> <li>MS mass tolerance: 8 ppm</li> <li>MS/MS mass tolerance: 0.8 Da</li> <li>Fixed mods: Carboxymethyl</li> <li>Variable mods: none</li> <li>Tryptic map: enzyme "trypsin," up to 1 missed cleavage</li> <li>Thermolysin map: "no enzyme" specificity</li> </ul> <p><b>ETS parameters:</b></p> <ul style="list-style-type: none"> <li>Same conditions as above, except trypsin specificity</li> <li>Tryptic map: enzyme "trypsin" and "semi-trypsin" with both searches combined for data analysis</li> <li>Thermolysin map: "no enzyme" specificity</li> </ul>	<p>Peak list generated with Excel and processed with Mascot v. 2.4.1 as follows:</p> <p><b>Run 1 Standard search parameters:</b></p> <ul style="list-style-type: none"> <li>MS mass tolerance: 5 ppm</li> <li>MS/MS mass tolerance: 0.05 Da</li> <li>Fixed mods: none</li> <li>Variable mods: none</li> <li>Tryptic map: enzyme "trypsin," up to two missed cleavages</li> <li>Thermolysin map: "no enzyme" specificity</li> </ul> <p>Used to generate exclusion list.</p> <p><b>Run 2 Standard search parameters:</b></p> <ul style="list-style-type: none"> <li>MS mass tolerance: 3 ppm</li> <li>MS/MS mass tolerance: 1.1 Da</li> <li>Fixed mods: none</li> <li>Variable mods: Carboxymethyl and pyro-Glu (N-terminal Q of proteolytic peptides)</li> <li>Tryptic map: "no enzyme" specificity</li> <li>Thermolysin map: "no enzyme" specificity</li> </ul> <p><b>Run 2 ET search parameters:</b></p> <ul style="list-style-type: none"> <li>MS mass tolerance: 3 ppm</li> <li>MS/MS mass tolerance: 1.1 Da</li> <li>Fixed mods: none</li> <li>Variable mods: Carboxymethyl and pyro-Glu (N-terminal Q of proteolytic peptides)</li> <li>Tryptic map: "no enzyme" specificity</li> <li>Thermolysin map: "no enzyme" specificity</li> </ul>																																										

*Continued on next page.*

**Table 3. (Continued). LC-MS Analysis and Database Searching Details**

<i>Lab 1</i>	<i>Lab 2</i>
<i>Post-Search Data Processing</i>	
<p><b>In-house developed software post-search processing:</b></p> <ul style="list-style-type: none"> <li>Automatically integrates data between Mascot ETS results and raw data files.</li> <li>Sorts, organizes, and filters Mascot results (hits due to PTMs, adducts and artifacts separated from the hits due to AA substitution).</li> <li>Generates and integrates reconstructed XICs,</li> <li>Calculates relative levels (%) of sequence variants.</li> </ul> <p>Potential sequence variants are manually validated by comparing their MS/MS spectra with the wild-type, if available.</p> <p>Quantitation is based on tryptic map. Thermolysin data is used for quantitation if variant is not detected in tryptic map even by accurate mass.</p> <p>Variant reporting cutoff: 0.01%</p>	<p><b>Evaluation criteria:</b></p> <ul style="list-style-type: none"> <li>Mass accuracy of 3 ppm (also the search limit for Mascot).</li> <li>Manual evaluation of spectrum quality.</li> <li>Cross-reference hits between the digests. Search considered invalid if not found in trypsin after manual searching of XICs.</li> <li>Elution of variant peptides relative to the wild-type is considered.</li> <li>Possibility of false positives due to artifacts and PTMs is considered.</li> </ul> <p>Quantitation based on XICs to determine relative amounts of variants from tryptic maps</p> <p>If AA substitution involved tryptic cleavage site, thermolysin data is used.</p> <p>Variant reporting cutoff: none</p>
<p>AA, amino acid; CID, collision-induced dissociation; ETS, error tolerant search; FA, formic acid; HPLC, high-pressure liquid chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; PTM, post-translational modification; TFA, trifluoroacetic acid; UV, ultraviolet; XIC, extracted ion chromatogram.</p>	

Due to the improved detection capabilities of modern mass spectrometers, low-level features (often below 1%) of protein therapeutics not previously known are now easily detected. With efficient digestion of the protein as one of the most important goals of peptide mapping, careful optimization of the sample preparation procedure is critical to ensure that the maps do not contain high levels of method-induced artifacts. These artifacts often include N-terminal cyclization, semi- and non-specifically cleaved products, and rearrangement of residues from both termini (116, 117). Furthermore, peptide mapping reactions with trypsin enzyme are normally carried out at high temperatures and pH, thus producing artificial clips and inducing deamidation. One complication of the use of IAA as the alkylation reagent is that it adds a species with the elemental composition of C<sub>2</sub>H<sub>2</sub>O<sub>2</sub>, for a mass of 58.01 Da. This is identical to the change due to a Gly→Asp or an Ala→Glu substitution. Although IAA primarily modifies Cys residues with the addition of a carboxymethyl group, IAA also reacts with Met, Lys, and His residues, with Lys and His residues resulting in a mass addition of 58.01 Da. Overalkylation with double alkylation on Cys residues, resulting in a mass addition of 116.02 Da, has also been observed at approximately 0.2% levels. The need to fine-tune the alkylation conditions should target minimization of these artifacts by balancing alkylation. We note, however, that the complete avoidance of the artifacts is not feasible as both under- and over-alkylated variants are present in the digests at the same time at low levels. Thus, the product ion spectra must be scrutinized carefully to confirm the presence of a potential amino acid substitution and to distinguish it from potential overalkylation. Alternatively, re-analysis of the sample using a different alkylating reagent such as iodoacetamide could be used to rule out sample preparation artifacts. In this dataset, the product ion spectra were evaluated to confirm several Gly→Asp misincorporation events, and only matches confirmed with high confidence were included in the results. Other known artifacts that can interfere with the detection of amino acid substitutions include oxidation; deamidation; transpeptidation; and acetaldehyde reaction on

the N-termini of peptides for addition of C<sub>2</sub>H<sub>2</sub>, or 26.02 Da, which is identical to an Ala→Pro substitution. Note that other potential amino acid substitutions are a nominal 26 Da mass addition but have different elemental compositions and can thus be distinguished by accurate mass measurements. Reaction of acetaldehyde occurs at low levels and appears to be more prevalent when aged reduction and alkylation reagents are used, or when tris(2-carboxyethyl)phosphine is used as the reductant. It should also be noted that interference from host-cell proteins with coincidental peptides, especially from process or incompletely purified samples, or low levels of carryover from previous analyses, can show up as potential sequence variants. In cases of distinguishing true amino acid substitutions from artifacts, all of the available data should be used to confirm that the conclusion is correct.

To ensure sufficient resolution between peaks in the peptide map, both labs utilized long reversed-phase gradients (190 min and 180 min, respectively), similar to that described previously by Yang et al. (85) for the separation of tryptic and thermolysin digests. Although TFA is not the best LC mobile phase modifier for MS detection, it was included in the method used by Lab 1 in order to enhance the resolving power of the column and to improve peptide recovery. In contrast, Lab 2 used formic acid (FA) in the mobile phase in combination with a smaller particle size (2.5 μm), charged-surface hybrid column to provide robust chromatography with minimal peak tailing. These columns are designed to improve separation when non-TFA based mobile phases are utilized.

Both laboratories used LTQ-Orbitrap Elite Velos Pro (Thermo Fisher Scientific, Inc.) for online detection of peptides during LC-MS analysis. Accurate mass detection is highly critical in verifying the fidelity of the amino acid sequence. Not only is it essential for confident peptide assignments by MS, but also for detecting small mass differences between peptides and for resolving nearly isobaric species. For that, Lab 1 used external calibration protocols to provide a mass accuracy of < 5 ppm, whereas Lab 2 employed additional internal lock mass calibration, using hexakis(1H,1H,3H-perfluoropropoxy)phosphazene at m/z 922.009798 for consistent mass measurement accuracy of 3 ppm.

Highly sensitive mass spectrometers in combination with high mass-accuracy detection and rapid low-energy CID acquisition were the methods of choice for sequence variant analysis protocols of both laboratories. Low-resolution CID in the ion trap was used in all Lab 1 experimental protocols. Lab 2 used an iterative approach, which exploited experimentally derived exclusion lists for improved detection of low-level sequence variants. Exclusion lists for each of the digests were first generated from the LC-MS data, where both MS and MS/MS CID data were acquired at high resolution, to provide confident identification of abundant peptides in the sample and to improve the overall sensitivity of the method. A second LC-MS run was then conducted with high-resolution MS and low-resolution MS/MS in the ion trap of the precursor ion selection for MS/MS. The focus of the Lab 2 approach, thus, was on finding all sequence variants at the lowest possible detection, without regard to the throughput. In contrast, the Lab 1 approach used a variant-reporting cutoff level set at 0.01% of relative peak area.

Both laboratories employed Mascot ETS for detection of sequence variants in the NIST IgG1 material. One of the main issues with the use of ETS strategy is an

overwhelming amount of hits that require manual evaluation. Several post-Mascot search criteria can be applied to verify the hits. These criteria can be based on the following (details are given in Table 3):

- Requirement for the presence of a corresponding wild-type peptide.
- Evaluation of the correctness of mono-isotopic peak assignment.
- Requirement for mass measurement error not to exceed the typical error for the majority of other peptides.
- Manual confirmation of spectrum quality and its comparison with that of a corresponding wild-type peptide.
- Requirement for hits to have a certain defined ion-score values.
- Cross-referencing hits between the digests with extra weight on trypsin data.
- Elution of variant peptides relative to a corresponding wild-type peptide.
- Possibility of false positives due to artifacts and common PTMs with mass shifts matching those of potential variants.

Lab 1 used in-house-developed software designed to integrate Mascot ETS results with raw data, automate filtering of results based on a set of pre-defined criteria, generate extracted ion chromatograms (XICs) for peptides, and determine relative amounts of variants. The common practice between the laboratories for determining the relative amounts was to take the ratio of the XIC peak area of the sequence variant divided by the sum of the areas of the sequence variant and wild-type peptide and multiply by 100.

## Results and Discussion

No sequence variants were found in the NISTmAb at a level above 0.1% relative abundance. However, many lower level substitutions were detected at levels ranging from 0.0002% to 0.1%. Amino acid substitutions identified in the NIST IgG1 at or above 0.01% relative abundance are tabulated in Table 4. A 0.01% cutoff was chosen as a reference value, below which amino acid substitutions fall within the realm of the fidelity of biological machinery. As a rule, quality of data (MS/MS spectra) decreases with the relative abundance of species in a sample, making verification of the correctness of assignment difficult for low-level sequence variants. In addition, Mascot output is abundant with false positives and requires careful manual verification, requiring a highly trained analyst.

The results of the comprehensive analysis for sequence variants by the two laboratories are in a reasonable agreement with each other. When combined, a total of 32 sequence variants with relative amounts of 0.01% to 0.1% were reported to be present in the NIST IgG1 molecule, of which nearly half (14 of 32) of the reported variants overlapped between the result of the two laboratories, as shown in Figure 2. There were several instances of the same amino acid substitution occurring at multiple sites throughout the molecule at low levels, specifically Ser→Asn at 10 sites, Gly→Asp at 5 sites, Val→Xle (likely, Ile) at 5 sites, Ala→Thr at 4 sites, and Arg→Lys at 3 sites. These observations are

consistent with a misincorporation event rather than a DNA mutation. Although the codon usage for this specific NIST standard is not known to the authors, all these misincorporations are consistent with a single base difference in the codons. For example, Ser→Asn could be due to AGU to AAU or AGC to AAC, where the difference is in the second position. Similarly, the Gly→Asp substitution could be GGU to GAU or GGC to GAC, again with the difference in the second position. These changes are consistent with a G (mRNA)/ U (tRNA) base pair mismatch during codon recognition, as recently reported by Zhang et al. (26) to occur at any of the three codon positions. Such a change from G to A can lead to these and other misincorporation instances occurring at multiple sites in the sequence (Table 6 in Ref. (26)).

**Table 4. Amino Acid Substitutions at Relative Amounts of  $\geq 0.01\%$  Detected in the NIST IgG1 mAb**

Variant Position	Digest	Peptide Sequence	AA substitution	Lab 1	Lab 2	Comment
L(158)	T	VDNALQSGNSQESVTEQDSK	Ser to Asn (S) [27.01]		0.02	L(158) and L(167) unresolved, both present
H(386)	T	GFYPSDIAVEWESNGQPENNYK	Ser to Asn (S) [27.01]		0.01	
H(30)	T	SGPALVKPTQLTLCTFSGFSL STAGMSVGWIR	Ser to Asn (S) [27.01]		0.02	
L(207)	TL	LSSPVTKS	Ser to Asn (S) [27.01]	0.10		
L(181)	T	DSTYLSSTLLTSK	Ser to Asn (S) [27.01]	0.03	0.04	L(176) and L(181) unresolved, both present, confirmed in TL
L(173)	T	DSTYLSSTLLTSK	Ser to Asn (S) [27.01]	0.07	0.05	Confirmed in TL and high resolution CID
L(170)	T	DSTYLSSTLLTSK	Ser to Asn (S) [27.01]	0.02		
H(411)	T	TTPVLDSDSGFFLYSK	Ser to Asn (S) [27.01]	0.06	0.02	
H(307)	T	VVSVLTVLHQDWLNGK	Ser to Asn (S) [27.01]	0.04		
H(270)	T	TPEVTCVVVDVSHEDPEVK	Ser to Asn (S) [27.01]	0.02	0.10	
L(156)	T	VDNALQSGNSQESVTEQDSK	Gly to Asp(G) [58.01]		0.02	
L(199)	T	VYACEVTHQGLSSPVTK	Gly to Asp (G) [58.01]	0.07	0.02	Confirmed in TL (0.11%)
H(284)	T	FNWYVDVGEVHNAK	Gly to Asp (G) [58.01]	0.03	0.03	
H(141)	T	STSGGTAALGCLVK	Gly to Asp (G) [58.01]	0.04	0.01	Confirmed in TL
H(125)	T	GPSVFPLAPSSK	Gly to Asp (G) [58.01]	0.01	0.01	
L(190)	T	VYACEVTHQGLSSPVTK	Val to Xle (V) [14.02]		0.01	Confirmed in TL
L(57)	T	LASGVPSR	Val to Xle (V) [14.02]	0.02	0.03	Confirmed in TL
H(311)	T	VVSVLTVLHQDWLNGK	Val to Xle (V) [14.02]	0.02	0.01	
H(308)	T	VVSVLTVLHQDWLNGK	Val to Xle (V) [14.02]	0.04	0.03	
H(128)	T	GPSVFPLAPSSK	Val to Xle (V) [14.02]	0.01	(0.002)	Confirmed in TL
H(51)	T	ALEWLAIDWDDKK	Ala to Thr (A) [30.01]	0.01		
H(381)	T	GFYPSDIAVEWESNGQPENNYK	Ala to Thr (A) [30.01]	0.02		
H(144)	T	STSGGTAALGCLVK	Ala to Thr (A) [30.01]	0.02		
H(132)	T	GPSVFPLAPSSK	Ala to Thr (A) [30.01]	0.01	(0.009)	
L(210)	T	SFNR	Arg to Lys (R) [-28.01]		0.09	
L(60)	T	LASGVPSR	Arg to Lys (R) [-28.01]	0.01		
H(318)	T	VVSVLTVLHQDWLN	Asn to Lys (N) [14.05]	0.02		
H(443)	TL	LHNHYTQKS	Ser to Arg (S) [69.07]	0.07		
H(150)	T	STSGGTAALGCLVK	Lys to Arg (K) [28.01]	0.04	0.02	Confirmed in TL (0.05%)
H(68)	T	DRLTISK	Arg to Gly (R) [-99.08]		0.01	Variant and wild-type have different lengths
L(10)	T	DIQMTQSPSTLSASVGDR	Thr to Ser (T) [-14.02]		0.01	
H(138)	T	STSGGTAALGCLVK	Thr to Gln (T) [27.01]		0.01	Two-base substitution

AA, amino acid; CID, collision-induced dissociation; H, heavy chain; L, light chain; T, trypsin; TL, thermolysin.

The use of accurate mass measurements can be extremely helpful in ruling out false positives. Using an internal lock mass in the MS mode allows confident exclusion of potential matches outside of 3 ppm mass error. This can save significant time in the evaluation of potential matches that meet other criteria except for mass accuracy. In addition, collection of the high-resolution product ion spectra in the Orbitrap mass spectrometer can aid in ruling in or out potential sequence variants. Although special software tools used during the analysis automated much of the data processing, a majority of hits that remained were due to false positives. Even after filtering the results, an overwhelming amount of over 600 Mascot hits collected by Lab 1 from the tryptic and thermolysin digests required manual evaluation.

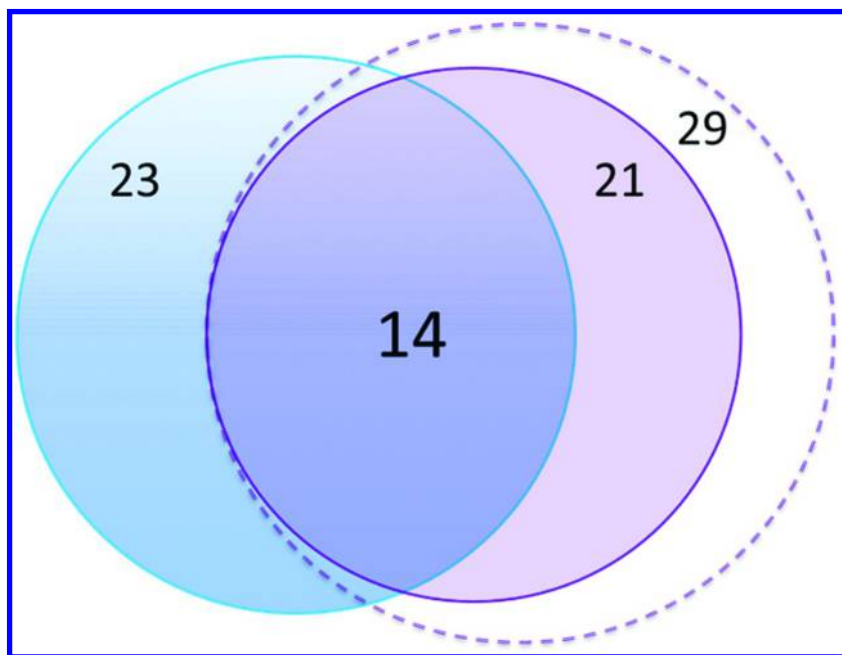


Figure 2. Venn diagram showing the results for sequence variant analysis in the NIST IgG1 material generated by the Lab 1 (blue circle) and Lab 2 (purple circle) with 0.01% reporting cutoff (filled circles) and with no cutoff (open circle) for data from Lab 2.

Several peptides were observed in which more than one amino acid substitution occurred in the same peptide, each generating a peptide with a single misincorporation at a different site. For example, three individual Ser→Asn substitutions were chromatographically separated and detected on the tryptic peptide DSTYLSSTLTLK, containing a total of five Ser residues. XICs of the doubly charged ion for this peptide showed three peaks (Figure 3). The locations of each substitution were manually validated by thorough evaluation

of the MS/MS spectra. In another example, tryptic LC peptide with a sequence of VTITCSASSR exhibited two peaks in the XIC of the Ser→Asn variant form, as shown in Figure 4. There are three Ser residues in this peptide, and therefore three potential sites of substitution for a Ser→Asn misincorporation. In this case, the product ion spectra of the substituted peptide at the two retention times are sufficient to localize the substitutions at the two Ser 24 and 26 or 27 sites. In fact, the Mascot ETS identified both of these Ser→Asn sites correctly and independently, although the two adjacent Ser sites (26 or 27) could not be distinguished from each other. The product ion spectra from the Mascot ETS results output are shown in Figure 5. In addition, these results were confirmed in the thermolysin digest. Based on these data, both Ser24 and Ser(26 or 27) to Asn misincorporations were considered confirmed, and the relative abundances were determined from the trypsin digest. This result illustrates the importance of chromatographic resolution for correct modification site assignments in multiple forms of a variant peptide having isobaric (within the mass measurement accuracy of an instrument) masses; MS/MS spectra alone often do not contain enough information, or the quality of the information is not sufficient, to make an unambiguous assignment.

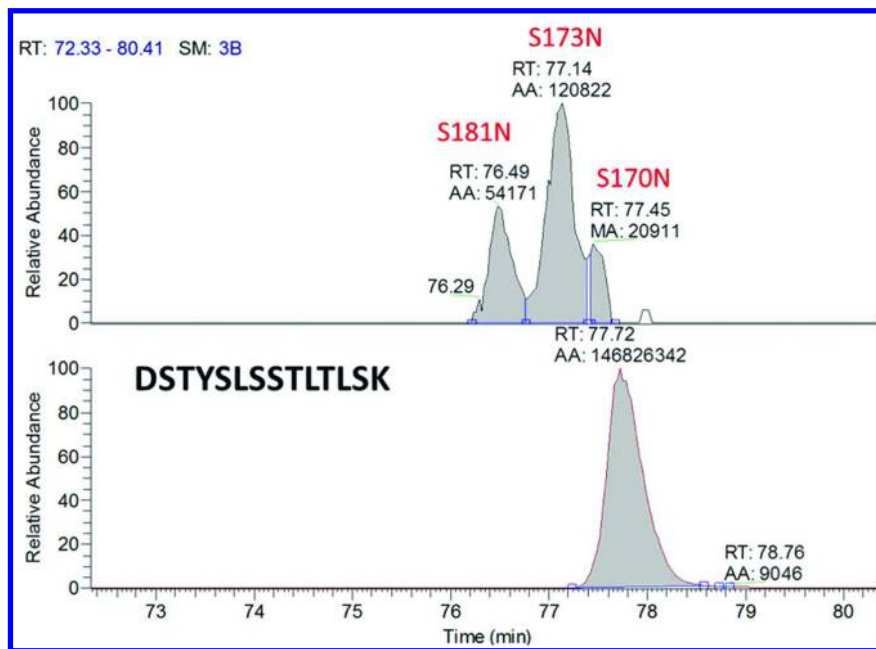


Figure 3. Lab 1 example extracted ion chromatograms for a doubly charged  $m/z$  735.39 Ser→Asn sequence variant (top) and doubly charged  $m/z$  751.88 unmodified (bottom) light chain residues 169 to 182 DSTYLSSTLTLSK peptides.



Another example compares MS/MS spectra for the unmodified STSGGTAALGCLVK tryptic peptide, and its Arg for Lys variant with relative abundance of 0.04% is shown in Figure 6. In this case, the good quality of the CID data allowed unambiguous identification of the location of the substitution. Because the substitution, occurring at the C-terminal end of the peptide, involves tryptic cleavage site and can potentially be due to transpeptidation, the thermolysin map was examined to confirm the correctness of the assignment.

Despite best efforts, some hits are more difficult to assign, particularly ones that are isomeric with PTMs such as alkylation (-58.01 and +58.01) and oxidation (+15.99). In order to assign these hits, careful comparison of MS/MS spectra against the wild-type were performed by manual and very time-consuming examination. A majority of the false positives were found in the thermolysin digest. Due to the no-enzyme-specificity setting in the ETS, the number of potential hits was found to increase exponentially.

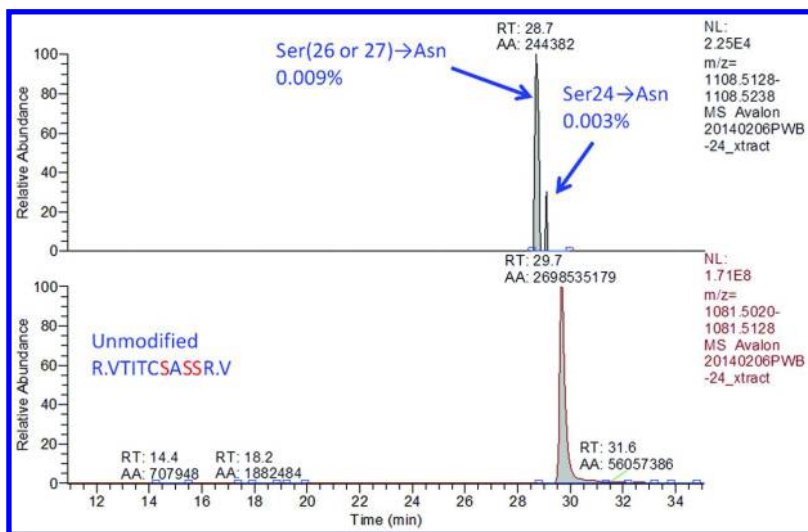


Figure 4. Lab 2 example extracted ion chromatograms, deconvoluted using the Xtract function of XCalibur software (Thermo Scientific, Inc.) for singly charged  $m/z$  1108.52 Ser→Asn sequence variant (top) and singly charged  $m/z$  1081.51 unmodified (bottom) light chain residues 19 to 28 VTITCSASSR peptides, using a mass tolerance of 5 ppm.



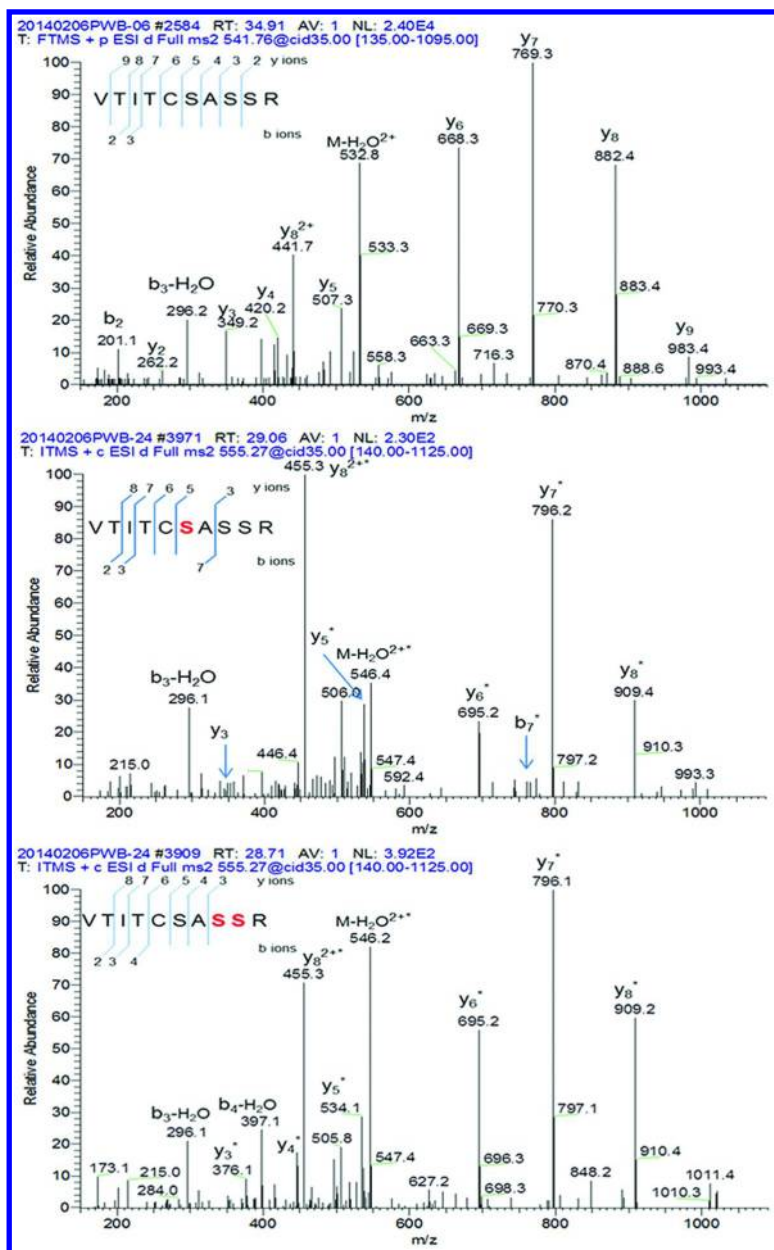


Figure 5. Lab 2 example annotated tandem mass spectrometry (MS/MS) spectra from the Mascot error-tolerant search (ETS) of a doubly charged Ser→Asn (+ 27 Da) misincorporation at light chain position 24 (middle) and 26 or 27 (bottom) compared with the unmodified peptide (top). Ions shifted by 27 Da relative to the unmodified peptide are indicated with an asterisk.

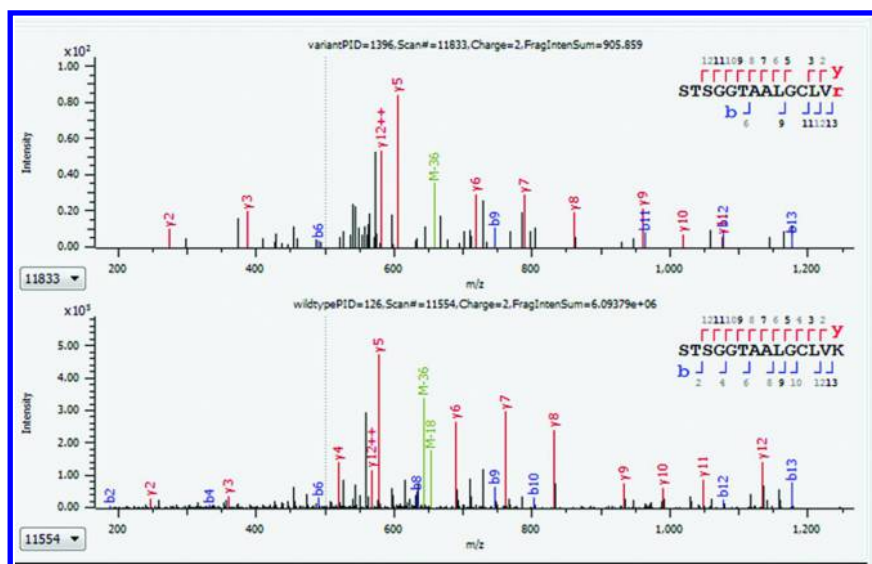


Figure 6. Lab 1 example annotated tandem mass spectrometry (MS/MS) spectra from Mascot error-tolerant search (ETS) of a doubly charged, heavy chain tryptic Lys150→Arg variant (top) and unmodified STSGGTAALGCLVK peptides (bottom).

Most of the peptide variants identified by Mascot were detected from the analysis of the tryptic map. *In silico* analysis of the NISTmAb sequence indicated that more than 20 peptides have molecular masses less than 600 Da. Detection of these peptides is often difficult because many would not be identifiable by Mascot or even retained chromatographically. One peptide generated by trypsin had a length of 63 amino acids and a mass of 6,713 Da. Although this peptide was detectable and identifiable by Mascot, its MS/MS spectra was complex. Localizing the position of the amino acid substitution even with close examination of the MS/MS spectra is often difficult for peptides that large. Thus, tryptic peptide mapping alone cannot provide full coverage of the antibody sequence. It is therefore beneficial to perform a secondary peptide map using an enzyme with orthogonal specificity.

A second peptide map using a complementary enzyme was determined to provide the necessary sequence. Because thermolysin generates relatively smaller peptides, the large, 63-amino acid tryptic peptide was then fractionated into small, more appropriately sized peptides for CID fragmentation. In addition, the difference in cleavage specificity provided overlapping peptide fragments with complementary chromatographic, ionization, and fragmentation properties. Furthermore, two peptides with sequence variants were exclusively identified in the thermolysin map. For example, thermolysin peptide LHNHYTQKS exhibited a +69.01 Da mass shift, corresponding to a Ser→Arg substitution with a level of 0.07%. Comparison of the MS/MS spectra of the sequence variant against the wild-type showed that the mass shift involves the *y*-ion series, with no change in the *b*-ion series. However, this Ser→Arg variant was not detected in the

tryptic map. Because the substitution to Arg introduced an additional tryptic cleavage site adjacent to the original site at Lys, the information on the sequence variant was lost, while a single amino acid Arg was generated. However, Mascot semi-tryptic ETS did positively identify the complementary fragment, LSLSPGK, which resulted from the substitution of Ser→Arg; alone, this result did not allow confirmation of the Ser→Arg substitution.

Chances of detecting sequence variant peptides depend on MS response and can be affected by size, chromatographic separation, ionization and fragmentation efficiency, location, and the effect of the substitution itself. Because of these reasons and the uniqueness of all peptides, establishing a uniform limit of detection and quantitation is not practical and may not be possible. For example, some peptides can be detected at levels below 0.001%, whereas others can be difficult to detect even at levels above 1.0%. Thus, as a more practical approach, it may be suggested to determine levels of detection of potential sequence variants separately for each of the original unmodified peptides in a protein sequence, based on their MS response. Because detection in tandem LC-MS is thresholded to detect peptides reaching a certain minimum number of counts, typically set at about  $10^3$  counts, a variant of an unmodified peptide with  $10^5$  or  $10^7$  counts, will be detected, respectively, at 1.0% or 0.01% levels.

### Modeling Detection of Low-Level Variants from Spiking Experiments

Spiking a known protein of a different identity at different levels into a sample containing the analyte of interest can serve as an excellent test case to evaluate the method performance for the detection of low-level variants. Because it is not possible to know *a priori* which sequence variants are present in a sample and at what level, this model provides the ability to target multiple peptides with a wide range of responses. This sequence variant model can be used to test assay performance, detect deficiencies or limitations of the method, and serve to test the system suitability of the work flow. Specific assay acceptance criteria can be established and integrated in the analysis to determine if the method used for sequence variant screening is performing as expected.

To demonstrate this approach, a second, mouse IgG1 molecule from Waters (Intact mAb Mass Check Standard, Waters Corporation), bearing a 60% sequence homology with the NIST IgG was spiked at the protein level into the NIST antibody sample at 0.2%, 0.5%, and 1.0% (w/w) prior to digestion with trypsin. The spiked samples were prepared and analyzed by Lab 1 using their established work flow. LC-MS/MS data was analyzed with Mascot by searching against a database containing both the heavy and light chain sequences of the Waters IgG1 standard and the NIST IgG antibody listed as two entries.

After tryptic digestion, the spiked sample produced over 40 peptides that differed in length, composition, elution, and ionization. A majority of these peptides at the 1.0% spike level were identified by Mascot, covering 90% of the Waters mAb sample. Undetected Waters mAb peptides were those with sizes <600 Da, and the glycopeptide EEQFNSTFR was not detected because the Mascot search did not include the glycosylation as a modification. At 0.5% and 0.2% spike levels, sequence coverage dropped to 88% and 86%, respectively. The

reduction in sequence coverage upon the decrease of the levels of spiked variant model mAb, suggests that the overall ability of the LC-MS method to detect sequence variants decreases with the relative abundance of a variant peptide in a protein.

As highlighted by this example, the method cannot guarantee detection of sequence variants in 100% of all cases, even at levels of 1.0%. It is important to ensure, however, that the method is tuned and optimized for the best possible detection of low-level variants. Based on the above spiking case study, the system suitability can be introduced with acceptance criteria set to target detection of  $\geq 85\%$  of the spiked variant mAb sequence from the sample spiked with 0.5% of the variant mAb. This model is designed to address potential errors associated with sample handling and preparation as well as with the performance of the entire LC-MS system. This measure is better suited for the detection of low-level variants than the sequence coverage of the major component alone, which is less sensitive to non-optimal experimental conditions or sample and system contamination.

It should be also noted that the use of multiple enzymes with variable specificities should be implemented to improve the chance of detection. As also discussed in this chapter, the arsenal of orthogonal analytical technologies routinely used for product release and characterization serves as a powerful approach to detect sequence variants.

### Analytical Figures of Merit

Modern LC-MS methods are capable of detecting sequence variants at 0.01% (60) and even at levels as low as 0.001% (26), which already fall within the expected basal fidelity of biosynthetic processes ( $10^{-4}$  to  $10^{-5}$ ). This section generalizes characteristics of the LC-MS method for the detection of sequence variants across a protein sequence. Although this discussion applies to the general case of the application of the method, the above spiking studies were used to support the discussion.

Chromatographic separation and speed of MS data acquisition can complicate the detection of variants by crowding the time- $m/z$  domain beyond the ability of the method to produce data for all the species. The issue of undersampling can be substantial in cases where samples are extremely complex and multi-component, such as during MS-based proteomics investigations. In contrast, during the analysis of biotherapeutic proteins by peptide mapping, samples are relatively pure and dominated by a single protein component. Baseline separation for most peptides often is achieved, which is in contrast to MS-based proteomics, where the number of peptides in the sample significantly exceed the ability of the chromatographic method to separate them. A tryptic digest of a typical monoclonal IgG, assuming up to two missed cleavages, contains less than 200 peptides, of which less than 100 are properly cleaved primary sequence peptides. This relatively simple mixture can be effectively separated using a long LC gradient, similar to those described in this chapter, with routine peak capacities of 300.

Ionization efficiencies of peptides are sequence-dependent. This is the major reason why different peptides exhibit variable MS peak intensities in peptide

maps. Broadening of chromatographic peaks or compromised recoveries of particularly hydrophobic species can also play a role. Spiking experiments with two independent mAbs can model the detection characteristics of the LC-MS method (85). The  $\log_{10}$ - $\log_{10}$  plot presented in Figure 7 shows the signal intensities of six peptides from a commercially available Waters mAb spiked into the NISTmAb, where the value of 2 on the  $x$ -axis represents 100%, or pure, Waters mAb. It should be noted that there is at least an order of magnitude span in signal intensities of these peptides detected in the digest of the pure Waters mAb, exhibiting the inherent differences in the ionization efficiencies of these peptides.

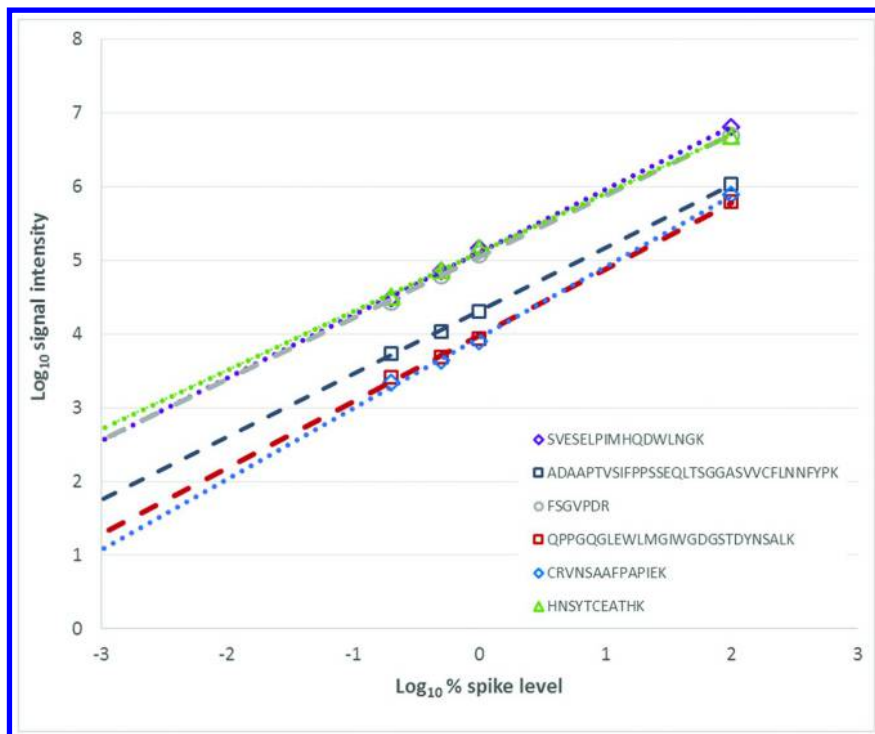


Figure 7.  $\log_{10}$ - $\log_{10}$  dependence of mass spectrometry (MS) signal intensity of six peptides from Waters monoclonal antibody (mAb) on a relative spike level of the Waters mAb into NISTmAb.

### Method Detection Limit

The detection of peptides in a typical data-dependent acquisition method, such as those used in these experiments, is based on the establishment of a certain ion count threshold value, beyond which potential precursor ions are selected for tandem MS analysis. Providing their successful identification during database searching, these precursors gain detected status. A threshold value can be established as the minimum number of ion counts required to determine a

charge state of a potential ion of interest, distinguish it from a random electronic noise, and produce good-quality tandem MS data. In a typical LC-MS method using LTQ-Orbitrap instrumentation, threshold values of 1000 to 3000 counts are often well-suited to providing informative MS/MS spectra for peptides. Thus, this threshold provides an estimate for the lowest signal intensity of an analyte that can potentially be detected by the method. In the Figure 7 example, the *x*-axis values corresponding to intercepts of the extrapolated calibration plots with *y* is 1 (corresponding to signal intensity of 1000 counts) can be used to define the detection limit, based on signal intensity for the model peptides. The detection limits for peptides show strong dependence on their signal intensities in the original pure sample, as plotted in Figure 8. This approach predicts that the minimum spike of Waters mAb that can be detected falls within a wide range from 0.002% to 0.1%, depending on the peptide.

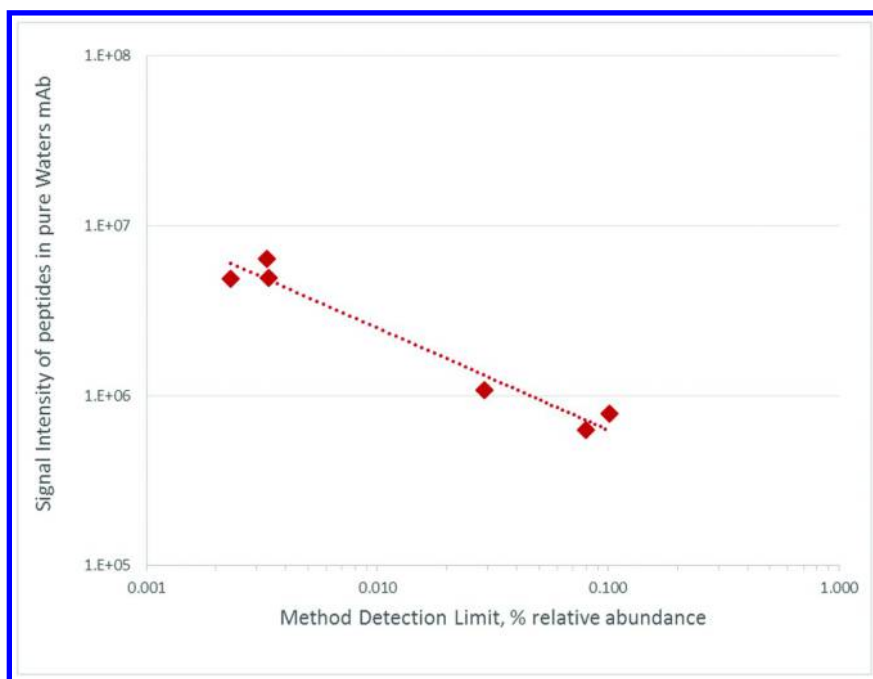


Figure 8. Estimation of method detection limits for variants as a function of signal intensity of primary unmodified peptides based on calibration curves for the six model Waters monoclonal antibody (mAb) peptides shown in Figure 7.

Extrapolating the spiking experiment to the detection of sequence variants in a protein, the detection characteristics of the method can now be estimated. Detection limits for potential sequence variants can be established individually for each of the primary sequence peptides from a 2-point calibration, connecting the plot's origin and signal intensity coordinate of a primary sequence peptide at its 100% abundance, using the following equation:

$$DL, \% = \frac{I_{\text{thrs}}}{I_{100\%}} \times 100\%$$

where  $I_{\text{thrs}}$  is the method's defined threshold for data-dependent acquisition and  $I_{100\%}$  is the signal intensity of a primary sequence peptide.

The assumption of a 100% abundance of a primary sequence peptide is reasonable for detection of low-level variants. Potential differences in ionization efficiencies between original primary sequence peptides and their analogs with sequence variants might be neglected for this estimation, although in some cases may be significant. Furthermore, the pronounced signal saturation exhibited by abundant primary sequence peptides contributes to the underestimation of detection limits estimated by the two-point calibration method, which likely is beyond the variability of ionization efficiencies of the primary sequence peptides and their potential variants. Based on these estimations, method detection limits are predicted to fall within 0.002% to 0.1% for the majority of the peptides in a peptide map of a typical mAb. It should be noted that all the sequence variants reported in this chapter for the NISTmAb were within this range.

### *Precision*

Yu et al. reported that multiple measurements ( $n = 5$ ) on multiple days ( $n = 3$ ) of two mAb molecules exhibited RSDs of 10% to 15% for variants at misincorporation levels of 0.05% or higher (60). In a different study conducted in one of our laboratories (Novavax, Inc., data not shown), measurement repeatability for several peptides in a recombinant protein, with relative abundances from 0.3% to 30%, was below 6% RSD among repeated ( $n = 4$ ) measurements, conducted on two different days, and even using two different 105- and 165-minute-long LC gradients. However, the variability increased to above 20% RSD when the amount of sample loaded onto the column varied at levels of 6  $\mu\text{g}$ , 17.5  $\mu\text{g}$ , and 35  $\mu\text{g}$  of protein. Furthermore, the relative differences in abundances of all of the peptides, regardless of their levels, exceeded 50% between protein loads of 6  $\mu\text{g}$  and 35  $\mu\text{g}$ . This result suggests that to achieve good precision using the LC-MS method for relative quantitation, experimental conditions, such as column loading amount or other parameters that can affect MS response, should be kept consistent between the experiments.

To compare the relative quantitation of sequence variants in the NISTmAb detected by the two laboratories, relative abundances of the 14 overlapping sequence variants were plotted, as shown in Figure 9. The diagonal line in Figure 9 illustrates equivalence of datasets. On average, the results from Lab 1 are higher compared with the data from Lab 2, as evidenced by a larger portion of data points being below the diagonal line. In two cases, results from the two labs differed by almost an order of magnitude (0.02% vs. 0.1% and 0.01% vs. 0.002% reported by Lab 1 and Lab 2, respectively). The majority of the results (71% or 10 out of 14) for sequence variants, however, had relative differences within 100%. The agreement between the datasets is reasonable, given the complete independence

in sample handling, data acquisition, and the way data were generated between the two laboratories.

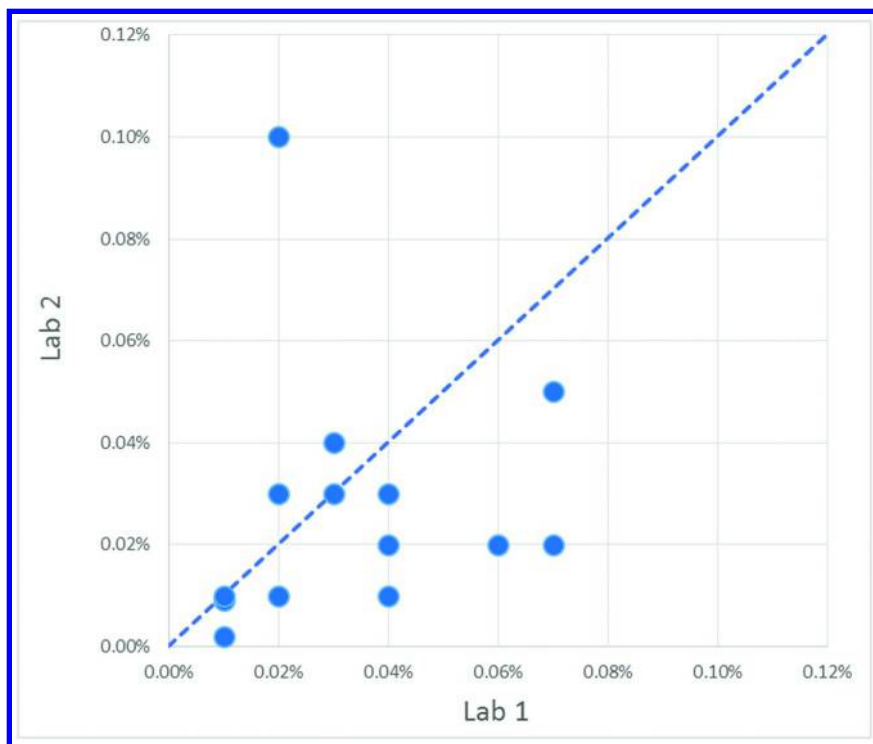


Figure 9. Correlation between relative abundances of sequence variants reported by Lab 1 and Lab 2 in the NIST monoclonal antibody (mAb). The diagonal line indicates one-to-one correlation, shown for visual comparison.

### Accuracy

The ability of an analytical method to accurately report concentrations of analytes is an important metric of that method. Evaluation of accuracy of a method implies the use of standards with known concentration against which signal response of a system can be established during calibration. In contrast, relative quantitation of components in a sample assumes proportionality between their responses and their abundances. For the LC-MS method described in this chapter, however, that is not necessarily true for a wide range of responses and is limited by multiple factors, out of which signal saturation at the upper end of the dynamic range, ionization efficiency, and suppression seem to be the most prevalent.

To get insights on the accuracy of the method, abundance percentages of selected tryptic peptides from Waters mAb, relative to analogous peptides from



NISTmAb, were plotted against 0.2%, 0.5%, and 1.0% protein-level spikes of Waters mAb into NISTmAb. The choice of Waters mAb and NISTmAb tryptic peptide pairs, selected for this evaluation, was such that differences in their sequences do not exceed more than two amino acids. Peptide pairs are presented in Table 5. Regression lines were established for relative abundances of Waters mAb peptides plotted against protein spike levels, allowing evaluation of the accuracy of protein quantitation based on the peptide-level measurement. Slopes of regression lines determine the closeness of peptide quantitation results to actual protein amounts of spiked protein. As presented in Table 5, slopes ranged widely between 0.78 and 2.95, indicating that relative peptide quantitation can under- and overestimate the abundance of variant proteins.

**Table 5. Characteristics of Regression Lines for the Relationship between Abundance Percentages of Several Waters mAb Tryptic Peptides Relative to Analogous NISTmAb Peptides and the Spike Level of Waters mAb on the Protein Level**

<i>Waters mAb (spike)</i>	<i>NIST mAb (major)</i>	<i>Slope</i>	<i>Correlation (<math>R^2</math>)</i>
SQVFLK	NQVVLK	1.72	0.997
LNVQK	LTVDK	1.56	0.999
LEIK	VEIK	2.95	0.996
DSTYSMSSTLTLTK	DSTYLSSTLTLK	0.78	0.995

Thus, the relative quantitation by the LC-MS methodology is not absolute and should be viewed as semi-quantitative. On the other hand, accurate quantitation would require availability of standards for peptides with sequence variants to calibrate the system. This would not be practical in most cases of sequence variant analysis in a biopharmaceutical setting, where fast turnaround and comprehensive sequence coverage are required. Excellent detection characteristics and relatively high throughput of the LC-MS method make it a powerful technology for screening for sequence variants in biotherapeutic proteins.

## Concluding Remarks

In this chapter, we attempted to summarize the current knowledge of the origins, occurrence, significance, and present-day state-of-the-art detection methods for sequence variants in recombinant proteins used as biotherapeutics. We define any unintended amino acid substitutions along the polypeptide chain of a protein as a sequence variant. Because of the nature of production by living organisms, errors associated with the underlying protein biosynthesis govern the occurrence of sequence variants in recombinant proteins. Sequence variants can originate at any step of the complex protein synthesis process, including replication, transcription, and translation events. Sequence variants occurring during DNA replication are known as mutations. Evolutionary DNA replication is the most accurate biosynthesis step, with error rates estimated at  $10^{-9}$  per base pair. Mutations can arise during transfection and gene amplification and

can amplify due to the selection pressures used to establish high-producing clones. Once a clone containing a genetic variant is produced, this mutation will propagate through sub-cloning and can be affected by a cell age. For that reason, ICH guideline (89), emphasizes the need to confirm that the coding sequence is correct and that the sequence is maintained for the course of a number of population doublings to establish a limit for *in vitro* cell age. In contrast, translational errors can be due to either misacylation (or mischarging) of tRNA by its cognate aminoacyl-tRNA synthetases with an erroneous amino acid or a tRNA anticodon mismatching (misreading) with a three-base codon of mRNA, leading to introduction of errors into the growing polypeptide chain. These type of errors are more abundant and occur *in vivo* at the rates of  $10^{-4}$  to  $10^{-5}$  per codon, but they can increase significantly due to nutrient and amino acid deprivation in production medium or an non-optimal codon usage in the construct. In many cases, these errors can be successfully reduced or even eliminated by optimization of cell culture conditions and processes.

In addition, sequence variants are not only limited to a single amino acid substitution. For example, mutations affecting a stop codon, frameshift, or stop codon readthrough, as well as splicing errors during transcription, can produce variants with the addition or omission of multiple amino acids at the same time.

The effect of the sequence variants on potency, efficacy, and safety of biotherapeutics is not known *a priori* and needs to be addressed on a case-by-case basis. A proper risk assessment serves as a gating point in deciding a strategy to evaluate the variants' effect on vital properties of a biotherapeutic protein. Bioassays can be used to assess binding potency or effector functions of IgGs, although the effect of sequence variants at levels of about 1% will likely be too small to be reliably detected by the assays. The effect on immunogenicity of molecules with low-level variants can be difficult to assess because biotherapeutics can already be heterozygous (allele mismatch) for certain patient population groups. In certain cases, animal toxicology or even clinical data also may be needed to address the issue of the presence of a sequence variant.

It is strongly preferable, however, not to have appreciable amounts of sequence variants in biotherapeutics in the first place. Strategies are being developed to improve clone selection to avoid occurrence and potential propagation of genetic mutations. At the same time, cell culture and medium optimization are targeted to reduce commonly occurring misincorporations. However, despite best efforts, sequence variants in recombinant proteins are expected to occur with a certain frequency. Sequence variants normally exist at low levels, however, and are governed by frequencies of biosynthesis errors.

In this chapter, we discussed analytical technologies for screening for sequence variants. Based on published evidence, we conclude that the use of multiple and orthogonal analytical methods often assists with the detection of variants that would be missed by one or another method, especially in cases where sequence modifications involve more than a single amino acid change (Table 1). We note that the majority of the discussed methods deals with analysis on a protein level. Emerging molecular biology methods, such as NGS, are gaining acceptance in performing sensitive analysis of genetic integrity, and these methods deserve consideration.

Peptide mapping with LC-MS is currently the most used first-line method for detection of sequence variants in recombinant proteins. Work flows and tools, originally developed for proteomics applications, have served as the starting point for the development of sequence variant analysis strategies for biotherapeutics. The first such application was demonstrated by Yu et al. (60), who described the synergy of coupling high-capacity peptide mapping with online high-resolution MS followed by an error-tolerant database search for detection of sequence variants. Detection of amino acid substitutions at levels down to 0.001% to 0.1% (26), which is likely to be at the levels of “biological noise,” becomes feasible with new advances in laboratory instrumentation.

In summary, the NIST IgG1 was subjected to sequence variant analysis by peptide mapping using LC-MS at two independent laboratories. The results were found to be in a reasonable agreement with each other, given the relative nature of these measurements. Combined, both laboratories reported the presence of 32 amino acid substitutions at levels of 0.01% to 0.1% of the relative abundance on a peptide level. Out of these 32 variants reported, 14 sequence variants were detected in common in the results of the two laboratories. No sequence variants at levels above 0.1% were detected in the sample by either laboratory. The presented data for the NIST IgG1 molecule can serve as a reference point to scientists who are establishing their methods for sequence variant analysis or developing new strategies for when a high sensitivity for detection is required. The NIST IgG1 sample, accompanied by the sequence variant data, can be used as a system-suitability sample in protocols, such as peptide mapping, to confirm the acceptable performance of a method.

## Acknowledgments

OB thanks Victor Ling for his support and inspiration and Yali Lu of Novavax, Inc., for insightful discussions. JC and PB would like to acknowledge Jason Rouse for his input and guidance and Kathleen Cornelius for her help with sample preparation. MA would like to acknowledge Victor Ling, X. Christopher Yu, Yung-Hsiang Kao, and John Stults for their guidance and support.

## References

1. Pauling, L.; Itano, H. A.; Singer, S. J.; Wells, L. C. *Science* **1949**, *110*, 543–548.
2. Ingram, V. M. *Nature* **1957**, *180*, 326–328.
3. Gabriel, A.; Przybylski, J. *Nature Education* **2010**, *3*, 2–12.
4. Cohen, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 15521–15529.
5. Samaranayake, H.; Wirth, T.; Schenkwein, D.; Rätty, J. K.; Ylä-Herttua, S. *Ann. Med.* **2009**, *41*, 322–331.
6. Agrawal, V.; Bal, M. *BioProcess Int.* **2012**, *10*, 32–48.
7. Kuystermans, D.; Krampe, B.; Swiderek, H.; Al-Rubeai, M. *Cytotechnology* **2007**, *53*, 3–22.

8. Li, F.; Vijayasankaran, N.; Shen, A.; Kiss, R.; Amanullah, A. *mAbs* **2010**, *2*, 466–477.
9. Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*; Roberts and Company Publishers: Greenwood Village, CO, 2005; pp 576.
10. Harris, R. J. *Dev. Biol. (Basel, Switz.)* **2005**, *122*, 117–127.
11. Liu, H.; Gaza-Bulsecu, G.; Faldu, D.; Chumsae, C.; Sun, J. J. *Pharm. Sci.* **2008**, *97*, 2426–2447.
12. Jenkins, N.; Murphy, L.; Tyther, R. *Mol. Biotechnol.* **2008**, *39*, 113–118.
13. Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *10*, 1241–1252.
14. Walsh, G. *Drug Discovery Today* **2010**, *15*, 773–780.
15. Crick, F. *Nature* **1970**, *227*, 561–563.
16. Cramer, P. *Nature* **2007**, *448*, 142–143.
17. Santos, M. A. S.; Tuite, M. F. *Trends Biotechnol.* **1993**, *11*, 500–505.
18. Kimchi-Sarfaty, C.; Oh, J. M.; Kim, I. W.; Sauna, Z. E.; Calcagno, A. M.; Ambudkar, S. V.; Gottesman, M. M. *Science* **2007**, *315*, 525–528.
19. Komar, A. A. *Science* **2007**, *315*, 466–467.
20. Drummond, D. A.; Wilke, C. O. *Nat. Rev. Genet.* **2009**, *10*, 715–724.
21. Parker, J. *Microbiol. Rev.* **1989**, *53*, 273–298.
22. Jakubowski, H.; Goldman, E. *Microbiol. Rev.* **1992**, *56*, 412–429.
23. Boyle, B.; Dallaire, N.; MacKay, J. *BMC Biotechnol.* **2009**, *9*, 75.
24. Imashimizu, M.; Oshima, T.; Lubkowska, L.; Kashlev, M. *Nucleic Acids Res.* **2013**, *41*, 9090–9104.
25. Gout, J.-F.; Thomas, W. K.; Smith, Z.; Okamoto, K.; Lynch, M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 18584–18589.
26. Zhang, Z.; Shah, B.; Bondarenko, P. *Biochem.* **2013**, *52*, 8165–8176.
27. Sydow, J. F.; Cramer, P. *Curr. Opin. Struct. Biol.* **2009**, *19*, 732–739.
28. Fox-Walsh, K. L.; Hertel, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1766–1771.
29. Yadavalli, S. S.; Ibba, M. *Nucleic Acids Res.* **2013**, *41*, 1104–1112.
30. Jakubowski, H. *WIREs RNA* **2012**, *3*, 295–310.
31. Pezo, V.; Metzgar, D.; Hendrickson, T. L.; Waas, W. F.; Hazebrouck, S.; Döring, V.; Marlière, P.; Schimmel, P.; de Crécy-Lagard, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8593–8597.
32. Valle, RP; Morch, MD *FEBS Lett.* **1988**, *235*, 1–15.
33. Itakura, K.; Hirose, T.; Crea, R.; Riggs, A. D.; Heyneker, H. L.; Bolivar, F.; Boyer, H. W. *Science* **1977**, *198*, 1056–1063.
34. Masuda, M.; Dohmae, N.; Nonaka, T.; Oikawa, T.; Hisanage, S.; Goedert, M.; Hasegawa, M. *FEBS Lett.* **2006**, *580*, 1775–1779.
35. Meng, S. Y.; Hui, J. O.; Haniu, M.; Tsi, L. B. *Biochem. Biophys. Res. Commun.* **1995**, *211*, 40–48.
36. Kane, J. F. *Curr. Opin. Biotechnol.* **1995**, *6*, 494–500.
37. Seetharam, R.; Heeren, R. A.; Wong, E. Y.; Bradford, S. R.; Klein, B. K.; Aykent, S.; Kotts, C. E.; Mathis, K. J.; Bishop, B. F.; Jennings, M. J.; Smith, C. E.; Siegel, N. R. *Biochem. Biophys. Res. Commun.* **1988**, *155*, 518–523.
38. Brinkmann, U.; Mattes, R. E.; Buckel, P. *Gene* **1989**, *85*, 109–114.

39. Calderone, T. L.; Stevens, R. D.; Oas, T. G. *J Mol. Biol.* **1996**, *262*, 407–412.
40. Forman, M. D.; Stack, R. F.; Masters, P. S.; Hauer, C. R.; Baxter, S. M. *Protein Sci.* **1998**, *7*, 500–503.
41. Aguirre, B.; Costas, M.; Cabrera, N.; Mendoza-Hernández, G.; Helseth, D. L., Jr.; Fernández, P.; de Gómez-Puyou, M. T.; Pérez-Montfort, R.; Torres-Larios, A.; Gómez Puyou, A. *PLoS One* **2011**, *6*, 1–10.
42. Saxena, P.; Walker, J. R. *J. Bacteriol.* **1992**, *174*, 1956–1964.
43. Kane, J. F.; Violand, B. N.; Curran, D. F.; Staten N.R., N. R.; Duffin, K. L.; Bogosian, G. *Nucleic Acids Res.* **1992**, *20*, 6707–6712.
44. Hutterer, K. M.; Zhang, Z.; Michaels, M. L.; Belouski, E.; Hong, R. W.; Shah, B.; Berge, M.; Barkhordarian, H.; Le, E.; Smith, S.; Winters, D.; Abrosion, F.; Hecht, R.; Liu, J. *Biotechnol. Bioeng.* **2012**, *109*, 2770–2777.
45. Huang, Y.; O'Mara, B.; Conover, M.; Ludwig, R.; Fu, J.; Tao, L.; Li, Z. J.; Rieble, S.; Grace, M. J.; Russell, R. J. *Protein Sci.* **2012**, *21*, 625–632.
46. McNulty, D. E.; Claffee, B. A.; Huddleston, M. J.; Porter, M. L.; Cavnar, K. M.; Kane, J. F. *Protein Expr. Purif.* **2003**, *27*, 365–374.
47. Brinkmann, U.; Mattes, R. E.; Buckel, P. *Gene* **1989**, *85*, 109–114.
48. Schneider, E. L.; King, D. S.; Marletta, M. A. *Biochemistry* **2005**, *44*, 987–995.
49. Harris, R. P.; Mattocks, J.; Green, P. S.; Moffatt, F.; Kilby, P. M. *Biotechnol. Bioeng.* **2012**, *109*, 1987–1995.
50. Bogosian, G.; Violand, B. N.; Dorward-King, E. J.; Workman, W. E.; Jung, P. E.; Kane, J. F. *J. Biol. Chem.* **1989**, *264*, 531–539.
51. Lu, H. S.; Tsai, L. B.; Kenney, W. C.; Lai, P. H. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 807–813.
52. Tsai, L. B.; Lu, H. S.; Kenney, W. C.; Curless, C. C.; Klein, M. L.; Lai, P. H.; Fenton, D. M.; Altrock, B. W.; Mann, M. B. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 733–739.
53. Randhawa, Z. I.; Witkowska, H. E.; Cone, J.; Wilkins, J. A.; Hughes, P.; Yamanishi, K.; Yasuda, S.; Masui, Y.; Arthur, P. *Biochemistry* **1994**, *33*, 4352–4362.
54. Apostol, I.; Levine, J.; Lippincott, J.; Leach, J.; Hess, E.; Glascock, C. B.; Weickert, M. J.; Blackmore, R. *J. Biol. Chem.* **1997**, *272*, 28980–28988.
55. Redwan, E. M. *Arab. J. Biotechnol.* **2006**, *9*, 493–510.
56. Digermann, T. *Biotechnol. J.* **2008**, *3*, 90–97.
57. Jefferis, R. *Nat. Rev. Drug Discovery* **2009**, *3*, 226–234.
58. Raven, P. H.; Johnson, G. B.; Losos, J. B.; Mason, K. A.; Singer, S. S. *Biology*; 8th ed.; McGraw-Hill: Boston, MA, 2007; pp 277–298.
59. Stansfield, I.; Jones, K. M.; Herbert, P.; Lewendon, A.; Shaw, W. V.; Tuite, M. F. *J. Mol. Biol.* **1998**, *282*, 13–24.
60. Yu, X. C.; Borisov, O. V.; Alvarez, M.; Michels, D. A.; Wang, Y. J.; Ling, V. *Anal. Chem.* **2009**, *81*, 9282–9290.
61. Harris, R. J.; Murnane, A. A.; Utter, S. L.; Wagner, K. L.; Cox, E. T.; Polastri, G. D.; Helder, J. C.; Sliwkowski, M. B. *Biotechnology* **1993**, *11*, 1293–1297.
62. Wan, M.; Shiau, F. Y.; Gordon, W.; Wang, G. Y. *Biotechnol. Bioeng.* **1999**, *62*, 485–488.

63. Beck, A.; Bussat, M. C.; Zorn, N.; Robillard, V.; Klinguer-Hamour, C.; Chenu, S.; Goetsch, L.; Corvaia, N.; Van Dorsselaer, A.; Haeuw, J. F. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *819*, 203–218.
64. Zhang, T.; Huang, Y.; Chamberlain, S.; Romeo, T.; Zhu-Shimoni, J.; Hewitt, D.; Zhu, M.; Katta, V.; Mauger, B.; Kao, Y. H. *mAbs* **2012**, *4*, 694–700.
65. Dorai, H.; Sauerwald, T.; Campbell, A.; Kyung, Y. S.; Goldstein, J.; Magill, A.; Lewis, M. J.; Tang, Q.; Jan, D.; Ganguly, S.; Moore, G. *BioProcess Int.* **2007**, *9*, 66–72.
66. 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.
67. Wen, D.; Vecchi, M. M.; Gu, S.; Su, L.; Dolnikova, J.; Huang, Y.; Foley, S. F.; Garber, E.; Pederson, N.; Meier, W. *J. Biol. Chem.* **2009**, *284*, 32686–32694.
68. Khetan, A.; Huang, Y. M.; Dolnikova, J.; Pederson, N. E.; Wen, D.; Yusuf-Makagiansar, H.; Chen, P.; Ryll, T. *Biotechnol. Bioeng.* **2010**, *107*, 116–123.
69. Feeney, L.; Carvalhal, V.; Yu, X. C.; Chan, B.; Michels, D. A.; Wang, Y. J.; Shen, A.; Ressler, J.; Dusel, B.; Laird, M. W. *Biotechnol. Bioeng.* **2013**, *110*, 1087–1097.
70. Zeck, A.; Regula, J. T.; Larraillet, V.; Mautz, B.; Popp, O.; Göpfert, U.; Wiegshoff, F.; Vollertsen, U. E.; Gorr, I. H.; Koll, H.; Papadimitriou, A. *PLoS One* **2012**, *7*, 1–10.
71. Lin, S. X.; Baltzinger, M.; Remy, P. *Biochemistry* **1984**, *23*, 4109–4116.
72. Jefferis, R. *Biotechnol. Genet. Eng. Rev.* **2010**, *26*, 1–42.
73. Zhong, X.; Wright, J. F. *Int. J. Cell Biol.* **2013**, *2013*, 1–19.
74. Shields, R. L.; Namenuk, A. K.; Hong, K.; Meng, Y. G.; Rae, J.; Briggs, J.; Xie, D.; Lai, J.; Stadlen, A.; Li, B.; Foxi, J. A.; Presta, L. G. *J. Biol. Chem.* **2001**, *276*, 6591–6604.
75. Idusogie, E. E.; Presta, L. G.; Gazzano-Santoro, H.; Totpal, K.; Wong, P. Y.; Ultsch, M.; Meng, Y. G.; Mulkerrin, M. G. *J. Immunol.* **2000**, *164*, 4178–4184.
76. Torres, M.; Fernández-Fuentes, N.; Fiser, A.; Casadevall, A. *J. Biol. Chem.* **2007**, *282*, 13917–13927.
77. Lu, H. S.; Fausset, P. R.; Sotos, L. S.; Clogston, C. L.; Rohde, M. F.; Stoney, K. S.; Herman, A. C. *Protein Expression Purif.* **1993**, *4*, 465–472.
78. Zhou, W.; Freed, C. R. *J. Biol. Chem.* **2004**, *279*, 10128–10135.
79. Harris, R. J.; Kabakoff, B.; Macchi, F. D.; Shen, F. J.; Kwong, M.; Andya, J. D.; Shire, S. J.; Bjork, N.; Totpal, K.; Chen, A. B. *J. Chromatogr. B: Biomed. Sci Appl.* **2001**, *752*, 233–245.
80. van Beers, M. M. C.; Bardor, M. *Biotechnol. J.* **2012**, *7*, 1–12.
81. Doyle, H. A.; Zhou, J.; Wolff, M. J.; Harvey, B. P.; Roman, R. M.; Gee, R. J.; Koski, R. A.; Mamula, M. J. *J. Biol. Chem.* **2006**, *281*, 32676–32683.
82. *Guidance for Industry Immunogenicity Assessment for Therapeutic Protein Products*; Draft Guidance, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), February 2013.

83. Jefferis, R.; Lefranc, M.-P. *mAbs* **2009**, *1*, 1–7.
84. Pandey, J. P. *mAbs* **2012**, *4*, 553–554.
85. Yang, Y.; Strahan, A.; Li, C.; Shen, A.; Liu, H.; Ouyang, J.; Katta, V.; Francissen, K.; Zhang, B. *mAbs* **2010**, *2*, 285–298.
86. Que, A. H.; Zhang, B.; Yang, Y.; Zhang, J.; Derfus, G.; Amanullah, A. *BioProcess Int.* **2010**, *8*, 52–60.
87. Ren, D.; Zhang, J.; Pritchett, R.; Liu, H.; Kyauk, J.; Luo, J.; Amanullah, A. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2011**, *879*, 2877–2884.
88. Fu, J.; Bongers, J.; Tao, L.; Huang, D.; Ludwig, R.; Huang, Y.; Qian, Y.; Basch, J.; Goldstein, J.; Krishnan, R.; You, L.; Li, Z. J.; Russell, R. J. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2012**, *908*, 1–8.
89. *ICH Harmonized Tripartite Guideline, Quality of Biotechnological Products, Analysis of the Expression Construct in Cells Used for Production of R-DNA Derived Protein Products*; Q5B, Current Step 4 version, November 30, 1995.
90. Dillon, T. M.; Bondarenko, P. V.; Rehder, D.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S. *J. Chromatogr. A* **2006**, *1120*, 112–120.
91. 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. Anal.* **2000**, *21*, 1099–1128.
92. Battersby, J. E., personal communication.
93. Rea, J. C.; Wang, Y. J.; Moreno, T. G.; Parikh, R.; Lou, Y.; Farnan, D. Monoclonal Antibody Development and Physicochemical Characterization by High Performance Ion Exchange Chromatography. In *Innovations in Biotechnology*; Agbo, E. C., Ed.; InTech: Rijeka, Croatia, 2012; Chapter 19; ISBN: 978-953-51-0096-6, <http://www.intechopen.com/books/innovations-in-biotechnology/monoclonal-antibody-development-and-physicochemical-characterization-by-high-performance-ion-exchang> (accessed January 12, 2015).
94. Alvarez, M.; Tremintin, G.; Wang, J.; Eng, M.; Kao, Y. H.; Jeong, J.; Ling, V. T.; Borisov, O. V. *Anal. Biochem.* **2011**, *419*, 17–25.
95. Ronaghi, M.; Shokralla, S.; Gharizadeh, B. *Pharmacogenomics* **2007**, *8*, 1437–1441.
96. Mardis, E. R. *Annu. Rev. Genomics Hum. Genet* **2008**, *9*, 387–402.
97. Metzker, M. L. *Genome Res* **2005**, *15*, 1767–1776.
98. Ingram, V. M. *Nature* **1956**, *178*, 792–794.
99. Hancock, W. S.; Bishop, C. A.; Prestidge, R. L.; Harding, D. R. K.; Hearn, M. T. W. *Science* **1978**, *200*, 1168–1170.
100. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011–5015.
101. Pappin, D. J.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327–332.
102. Mann, M.; Højrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338–345.
103. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58–64.
104. Yates, J. R., III; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397–408.

105. Henzel, W. J.; Watanabe, C.; Stults, J. T. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 931–942.
106. Arnott, D.; Shabanowitz, J.; Hunt, D. F. *Clin. Chem.* **1993**, *9*, 2005–2010.
107. Eng, J. K.; McCormack, A. L.; Yates, J. R., III *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
108. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.
109. Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R., III *Anal. Chem.* **2000**, *72*, 757–763.
110. Creasy, D. M.; Cottrell, J. S. *Proteomics* **2002**, *2*, 1426–1434.
111. Wei, H.; Tymiak, A. A.; Chen, G. *Bioanalysis* **2013**, *5*, 1299–1313.
112. 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.
113. Zhang, Z. *Anal. Chem* **2009**, *81*, 8354–8364.
114. Bern, M.; Kil, Y. J.; Becker, C. *Curr. Protoc. Bioinf.* **2012**, *40*, 13.20.1–13.20.14.
115. Ambler, R. P.; Meadway, R. J. *Biochem. J.* **1968**, *108*, 893–895.
116. Karty, J. A.; Ireland, M. M. E.; Brun, Y. V.; Reilly, J. P. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2002**, *782*, 363–383.
117. Fodor, S.; Zhang, Z. *Anal. Biochem.* **2006**, *356*, 282–290.
118. Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P. R. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 496–502.
119. Stephen, J. A.; Keevil, E.-J.; Knight, D.; Hubbard, S. J. *J. Proteome Res.* **2007**, *6*, 399–408.



## Chapter 3

# Structural Elucidation of Post-Translational Modifications in Monoclonal Antibodies

Wenzhou Li,<sup>1</sup> James L. Kerwin,<sup>2</sup> John Schiel,<sup>3</sup> Trina Formolo,<sup>3</sup>  
Darryl Davis,<sup>4</sup> Andrew Mahan,<sup>4</sup> and Sabrina A. Benchaar<sup>\*,1</sup>

<sup>1</sup>Amgen Inc., Thousand Oaks, California 91320, United States

<sup>2</sup>Sanovas Inc., Sausalito, California 94965, United States

<sup>3</sup>Analytical Chemistry Division, National Institute of Standards and  
Technology, Gaithersburg, Maryland 20899, United States

<sup>4</sup>Janssen Research and Development, LLC,  
Spring House, Pennsylvania 19477, United States

\*E-mail: [sabrina.benchaar@amgen.com](mailto:sabrina.benchaar@amgen.com)

Therapeutic proteins may undergo a series of modifications throughout their production, processing, and storage. These modifications can include the addition or replacement of functional groups, or structural changes such as folding, cleavage, and racemization. The presence of these modifications can affect therapeutic monoclonal antibody (mAb) biological activity, half-life, and immunogenicity. Post-translational modifications that arise during cellular expression and chemical modifications that may result during the biomanufacturing process and storage must be discovered, tracked, and evaluated for their impact on quality, safety, and efficacy. A review of the most common modifications encountered during therapeutic mAb development and their characterization using modern analytical approaches is presented, with an emphasis on mass spectrometry-based approaches. As a real-world example, a peptide mapping-based characterization study involving three independent laboratories was conducted using the reference mAb provided by the National Institute of Standards and Technology (NIST). Results from this study are discussed to provide a representative sampling of peptide mapping as it is currently applied to characterization of mAb modifications, as well as

considerations to weigh when developing these experiments for qualitative and quantitative analysis.

## Introduction

Most, if not all, proteins are modified following translation from mRNA (1–3). Although it is estimated that the human genome comprises between 20,000 and 25,000 genes (4), the total number of proteins in the human proteome probably exceeds 1 million (5). In addition to mechanisms that generate different mRNA transcripts from a single gene (recombination, alternative splicing, differential termination of transcription, etc.), a major source of this variability is through post-translational modifications (PTMs). PTMs are defined as enzyme-catalyzed processing of the polypeptide chain after translation, involving more than 500 human protein kinases, 150 protein phosphatases, 500 proteases, transferases, and ligases (2). PTMs are vital to regulation of function, signal transduction, cellular regulation, protein targeting and localization, degradation, and structural/conformational rearrangements (1, 2). Some of the most common PTMs found in therapeutic monoclonal antibodies (mAbs) include glycosylation, disulfide bond formation, and proteolytic cleavage of the protein backbone. Downstream processing, formulation, and storage also may result in modifications to the protein, and have been classified as chemical modifications in reference to their non-enzymatic origin (e.g., oxidation, deamidation, glycation, pyroglutamate [pyro-Glu] formation). The introduction of both classes of heterogeneity can affect biological activity, half-life, and immunogenicity and therefore must be comprehensively characterized in therapeutic protein development. For the sake of clarity, “post-translational modification” will be used in the broadest sense for the remainder of this chapter to encompass both types of modifications, a practice that has become relatively commonplace in the biopharmaceutical industry.

Since the development of hybridoma technology in 1975 to produce mAbs in cellular systems (6), mAbs have proven to be valuable for basic immunology research, diagnostic testing, and therapeutic treatment (7–11). It was in 1986 that the first mAb, murinomab-CD3, was approved by the U.S. Food and Drug Administration (FDA) for preventing transplant rejections (12). Since then, many mAbs have been developed to treat a number of diseases, and these biotherapeutics have come to represent a new and rapidly growing class of drug with its own set of regulatory challenges. A unique regimen of analytical characterization is required when filing for review of a new biotherapeutic protein with the FDA. A critical portion of the submission package is a comprehensive characterization of PTMs, which can be particularly challenging when dealing with mAbs due to their complexity and bulky size. The rapid growth of the biosimilar marketplace has accelerated structural characterization into an even more central role. Biosimilars, also known as follow-on biologics, are “generic” versions of marketed therapeutics that may be produced once the protection of the original innovator patent expires. For over 25 years, protocols for accelerated FDA approval of generic small-molecule drugs have been in place. These allow for abbreviated review pathways, often lifting the requirement for clinical trials

if the generic compound is demonstrated to be pharmaceutically equivalent (with respect to active ingredient, purity, dosage, and administration route) and bioequivalent (i.e., similar rate of uptake, metabolism, and localization) (13) to the branded drug. Similarly, the Biologics Price Competition and Innovation Act of 2009 (14) provides for an abbreviated licensure pathway for biosimilars. However, due to the inherent variability of therapeutics produced by biological systems, absolute identity between products is never achieved. Instead, application sponsors are charged with demonstrating a high level of similarity between the innovator and the follow-on to qualify for this shorter approval process. Although small recombinant proteins such as insulin and human growth hormone can be completely characterized to show equivalency using conventional analytical procedures, therapeutic mAbs are much larger and more structurally diverse due to modifications such as N-glycosylation. This heterogeneity necessitates that more stringent criteria for the analytical and biophysical characterization of mAb biosimilars be required to demonstrate the level of structural variability (e.g., PTMs) that can be tolerated without inducing clinically important outcomes, thereby falling under the umbrella of substantial equivalence (15–17). In September 2013, the European Commission approved its first biosimilar mAb, which is a lower-priced biosimilar version of Remicade® (now produced and marketed by Johnson & Johnson and Merck & Co.) named Inflectra™ (Hospira, Inc. and Celltion) (18). More recently, the FDA announced its approval of the first mAb biosimilar in the U.S. market, Zarxio™ (Sandoz, Inc.), which was deemed substantially equivalent to Amgen's product Neupogen® using the abbreviated licensure pathway (19). These follow-on products are likely to pave the way for an increasing number of approved mAb biosimilars within these markets.

In this chapter, we will briefly review analytical methods commonly used for PTM characterization, with an emphasis on high-resolution mass spectrometry (MS)-based techniques. The common PTMs encountered during therapeutic mAb development will be discussed. In addition, the results from an interlaboratory PTM Characterization of the NISTmAb will be presented to demonstrate state-of-the-art peptide mapping while highlighting the subtle nuances of this approach. A representative forced degradation analysis was also conducted to demonstrate its utility in identifying sites prone to modification.

## Methods for PTM Characterization

Global information on the PTMs present in intact mAbs is gained through chromatographic or electrophoretic separation of mAb isoforms based on physicochemical properties, such as size (size exclusion chromatography [SEC] and capillary sodium dodecylsulfate electrophoresis [cSDS]) or charge (cation exchange [CEX] and capillary isoelectric focusing [cIEF]) (Separation chapter/Volume 2, Chapter 5). The resolved peaks can often be associated with certain PTMs and provide information on the global presence and approximate relative abundance of each in a given mAb preparation. For example, CEX and cIEF separate intact mAbs based on charge, and chromatographically resolved

acidic or basic proteoforms (relevant to the dominant species) often contain higher levels of deamidated Asn and/or sialylated glycan variants than the major species (20). Although these separation techniques are well established, they can usually indicate the levels of only a few dominating modification forms, with limited ability to provide detailed and site-specific PTM verification. This drawback has promoted the use of MS-based methods for PTM characterization, which provide an additional dimension of specificity through determination of accurate mass. A variety of MS-based methods may be used to characterize mAb PTMs, which include analysis of intact proteins and/or subunits (21, 22), as demonstrated experimentally in the Primary Structure chapter/Volume 2, Chapter 1, and bottom-up MS analysis of peptides following HPLC (Primary Structure chapter/Volume 2, Chapter 1 and this chapter), capillary electrophoresis, or ancillary separations (23, 24). Each of the methods discussed represent a component of a comprehensive PTM characterization platform and can be used in conjunction with other components for a high level of product understanding. For example, chromatographically resolved intact proteoforms can be fraction collected and further analyzed with orthogonal techniques such as MS. Separation science assays are covered in detail in the Separation chapter/Volume 2, Chapter 5 of this book, therefore this chapter will focus primarily on characterization of unfractionated samples, although the same methods are directly applicable to fractionated species.

### Intact Mass Analysis

Direct mass analysis of the intact mAb can be used to identify structural variants or modifications that result in large mass shifts, such as glycosylation, amino acid truncation, and glycation. The major advantages of this approach are that the sample preparation is minimal and the data analysis is straightforward. Alternatively, some protocols involve mass analysis after reducing the disulfide bonds or performing limited digestion with enzymes such as pepsin, papain, or IdeS (25). Sub-units resulting from these treatments are smaller in size, allowing improved mass accuracy and resolution as well as sub-unit-specific localization of PTMs. Matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) was the initial choice for intact mAb analysis because of its theoretically unlimited mass range in linear mode (26). Insufficient resolution combined the batch-to-batch heterogeneity of mAbs (27, 28), however, demonstrated the limited capacity of MALDI-TOF to form the core approach in intact mass characterization.

Instruments with electrospray ionization (ESI) sources are currently preferred for the majority of intact therapeutic mAb analyses. Electrospray analyses are often coupled with liquid chromatography (LC), providing additional separation before mass measurement. Characteristic charge distributions of proteins/subunits are detected using high-resolution mass analyzers, and ultimately a zero-charge spectrum can be produced (through deconvolution) to display the intact mass. Among various mass analyzers, Fourier transform ion cyclotron resonance (FTICR) MS provides the highest resolving power and mass accuracy for characterization of biomolecules (29–31). Unit mass baseline resolution of the isotopic distribution for an intact 147.7 kDa mAb has been achieved,

albeit requiring preliminary sample preparation, including prior dissociation of non-covalent adducts and extensive manipulation of acquisition parameters to minimize space charge shifts, peak coalescence, and destructive ion cloud Coulombic interactions (31). More routinely, intact protein measurements are collected using ESI coupled with quadrupole time-of-flight (QTOF) or Orbitrap mass spectrometers (32, 33). Intact and middle-down mass analysis of the NISTmAb using Orbitrap and QTOF instruments are discussed in further detail in the Primary Structure chapter/Volume 2, Chapter 1, wherein the NISTmAb glycoforms, C-terminal Lys variants, and glycation heterogeneities were observed within 50 ppm of their theoretical average masses.

Ion mobility mass spectrometry (IM-MS) uniquely provides additional separation on the basis of collisional cross sections, structure, size, and charge, which permits differentiation between isomers, conformers, isobaric species, and native protein complexes (34–36). Notable application examples in the mAb field include the characterization and monitoring of IgG4 Fab arm exchange (34), IgG2 disulfide structure isoforms (37), lot-to-lot heterogeneity in the N-glycosylation profile (38), and the structure of immune complexes (39). A detailed discussion of ion mobility analysis can be found in the Ion Mobility chapter/Volume 3, Chapter 4 in this series.

MS-based approaches for analyzing intact and/or subunit-level species can be very informative regarding the primary structure of a given protein, as well as some of the microheterogeneity present. For example, the NISTmAb was observed to have N- and C-terminal modifications, glycoform heterogeneity, glycation, and even low levels of oxidation through combined intact and middle-down analyses described in the Primary Structure chapter/Volume 2, Chapter 1. However, localization of these PTMs to a specific amino acid residue, a more detailed analysis regarding their relative quantities, and the investigation of their propensity to increase in quantity (e.g., during processing, formulation, and/or storage) are often examined further with bottom-up peptide mapping as discussed in the current chapter.

## Bottom-Up Approaches

The bottom-up approach for mAb characterization is typically referred to as “peptide mapping,” which is essentially an liquid chromatography-tandem mass spectrometry (LC-MS/MS) method coupling ESI with any one of a variety of instrument types (e.g., TOF [time-of-flight], ion trap, Orbitrap). The process involves proteolytic digestion of the mAb (commonly with trypsin), separation of the peptide mixture by chromatography, collection of full MS spectra, fragmentation of selected peptides, and collection of resulting MS/MS spectra. Peptide identification is accomplished by sequence database searches using MS and MS/MS spectra. Peptide mapping is currently the most sensitive and site-specific method used for PTM characterization, but the method is complicated by the overwhelming amount of data analysis required (Bioinformatics chapter/Volume 3, Chapter 7) and the potential for modifications to be induced during sample preparation or by the analytical instruments used.

Peptide mapping experiments used for the purposes of PTM characterization and monitoring are carried out in the same general manner as those described in the Primary Structure chapter/Volume 2, Chapter 1. In addition to providing confirmation of the expected amino acid sequence, PTM peptide mapping is intended to identify all alterations made to that sequence throughout the stages of production, processing, formulation, and storage. PTM peptide mapping methods must be capable of detecting and identifying low-abundance modifications, as well as providing the means to measure their relative quantities and assess their propensity to increase in abundance under various circumstances. Although many PTMs may be easily evaluated using platform methods, some individual PTMs may ultimately require that selective conditions (e.g., gradient, digest conditions, MS settings) be optimized for adequate evaluation due to co-elution with another peptide, poor ionization efficiency, and other analytical problems.

The ability to generate a high-quality, reproducible, and comprehensive peptide map depends on multiple factors, including denaturing conditions, the protease used, LC separation methods, MS instrument type and configuration, fragmentation method, and data analysis software used for processing and interpreting results. Optimizing each factor will best enable the detection of low-abundance species with high confidence. Selecting an appropriate protease can optimize peptide size and ionization efficiency, thereby enhancing sequence coverage. Certain proteases also may be selected to identify specific PTMs. For example, trypsin which cleaves after Arg and Lys residues will not cleave after a glycosylated Lys, and therefore, alternative enzymes such as chymotrypsin, Glu-C, or Asp-N may be needed. Choosing fragmentation methods appropriate for the analytical goal (collision-induced dissociation [CID], higher energy collision-induced dissociation [HCD] or electron transfer dissociation [ETD]) will improve the quality of MS/MS spectra, making peptide identification and PTM localization more reliable. For example, although CID is useful for obtaining glycan composition information from a glycopeptide, ETD may be the method of choice to generate fragments that can confirm the amino acid sequence as well as the site of glycosylation (40, 41).

Peptide mapping is often performed with both UV and MS detectors for visualization and quantification purposes; due to the differences in ionization efficiency or extinction coefficients between modified and unmodified peptide pairs, however, the PTM levels derived from MS and UV signals do not always correlate. It should be noted that in the absence of an internal standard and a calibration curve, these methods only provide relative quantitation (of PTM) rather than absolute levels. Nevertheless, because peptide mapping can provide very detailed sequence information, as well as site-specific PTM information, it is still the top choice for PTM characterizations in therapeutic protein development.

## Most Common PTMs in Monoclonal Antibodies

Although more than 500 kinds of PTMs have been discovered, only a few dozen are commonly observed in mAbs, including—but not limited to—glycosylation, C-terminal Lys clipping, N-terminal pyro-Glu formation,

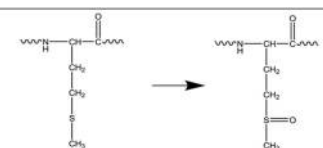
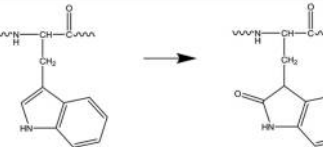
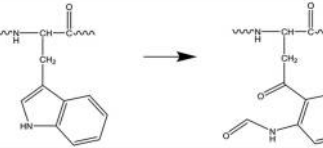
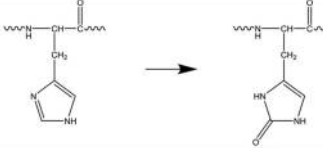
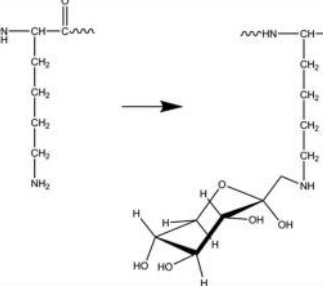

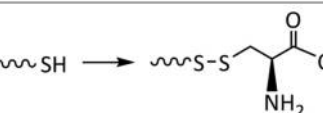

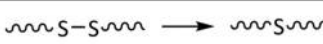
deamidation, oxidation, glycation, cysteinylolation, trisulfide bonding, hydroxylation, non-enzymatic fragmentation (clips), and isomerization. Though a few PTMs exist as dominating forms (e.g., glycosylation, C-terminal Lys clipping, N-terminal pyro-Glu formation), most other modifications present as minor species in very low abundance, and therefore, it is a daunting challenge to structurally characterize them. In this section we provide a brief review of the origins, biological impacts, and analytical challenges for those PTMs most commonly observed in mAbs. Table 1 lists the structures and mass shifts of the PTMs to be discussed in this section.

**Table 1. Post-Translational Modifications (PTMs) in Monoclonal Antibodies (mAbs)**

PTM	Unimod Notation	Native to PTM transition	Delta Mass
Glycosylation (e.g., FA2G1)	dHex(1)Hex(3) HexNAc(4)	<p>... EEQYNSTYR... → ... EEQYNSTYR...</p> <ul style="list-style-type: none"> <li>■ GlcNAc</li> <li>○ Mannose</li> <li>● Galactose</li> <li>▷ Fucose</li> </ul>	+1607.4737
Glycosylation (e.g., Man5)	Hex(5)HexNAc (2)	<p>... EEQYNSTYR... → ... EEQYNSTYR...</p> <ul style="list-style-type: none"> <li>■ GlcNAc</li> <li>○ Mannose</li> </ul>	+1216.4229
C-terminal Lysine Loss	Lys-loss		-128.09496
N-terminal Pyroglutamate	Gln → pyro-Glu		-17.02655
Deamidation	Deamidation		+0.98402
Ammonia-loss (Succinimide intermediate)			-17.02655
	Iso-Asp		+0.98402

*Continued on next page.*

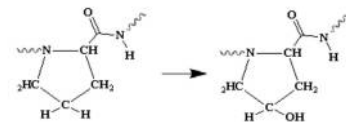
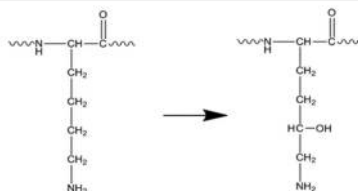
**Table 1. (Continued). Post-Translational Modifications (PTMs) in Monoclonal Antibodies (mAbs)**

<i>PTM</i>	<i>Unimod Notation</i>	<i>Native to PTM transition</i>	<i>Delta Mass</i>
Met Oxidation	Oxidation		+15.99491
Trp Oxidation	Oxidation		+15.99491
Trp Dioxidation	Dioxidation		+31.98982
His Oxidation	Oxidation		+15.99491
Glycation	Hex		+162.05282
Sulhydryl Formation	N/A		+ 2 when intrachain + 1 each chain when interchain
Cysteinylation	Cysteinylyl		+119.0041
Trisulfide Formation	N/A		+31.97207
Thioether Formation	N/A		-31.97207

*Continued on next page.*



**Table 1. (Continued). Post-Translational Modifications (PTMs) in Monoclonal Antibodies (mAbs)**

<i>PTM</i>	<i>Unimod Notation</i>	<i>Native to PTM transition</i>	<i>Delta Mass</i>
Fragmentation	N/A	$\dots\text{TPEVTCVVVDVSHEDPEVK}\dots \rightarrow \dots\text{PEVTCVVVDVSHED-OH}$ $\text{NH}_2\text{-PEVK}\dots$	N/A
Pro Hydroxylation	Hydroxylation		+15.99491
Lys Hydroxylation	Hydroxylation		+15.99491

## Glycosylation

Glycosylation is a major PTM affecting protein folding, conformation, localization and activity, with sugar moieties ranging from simple monosaccharides to complex branched polysaccharides (42, 43). Most therapeutic mAbs have a single N-glycosylation site (e.g., heavy chain N300 in the case of NISTmAb), although some mAbs may have an additional Fab glycosylation site. Glycosylation is known to affect complement activation; antibody-dependent cell cytotoxicity; endocytosis of immune complexes leading to antigen presentation; and inhibition of B lymphocytes, monocytes, and basophils (43–46). Glycans are usually large (a few thousand Daltons) and diverse (several dozen forms coexist), meaning that they are major contributors to mAb heterogeneity. The most straightforward way to characterize glycans is to measure the intact mass of the whole protein or a large part of the whole protein (e.g., heavy chain only, Fc only). These intact and middle-down approaches can yield an idea of the more dominant glycoform compositions, but they provide little information on the inherent structure of this PTM. Analysis of glycosylation during peptide mapping is used to identify the site of glycosylation as well as the glycan compositions present at a given glycosylation site. Glycopeptides of the same apopeptide typically elute in a narrow window during reversed phase (RP) peptide mapping; MS can easily differentiate glycan compositions, however, based on characteristic mass shifts. Table 1 depicts two representative

structures, including a complex biantennary glycan (FA2G1) and a high mannose glycan (Man5) along with the expected shifts in mass versus the aglycosylated apo-peptide. MS/MS fragmentation using CID can be used to gain additional confidence in the glycan assignment, with fragmentation dominated by loss of sequential monosaccharide residues. Higher energy collision dissociation (HCD) in the C-trap of an Orbitrap-type mass spectrometer or HCD in a sector or TOF-TOF instrument is capable of producing greater levels of peptide backbone fragmentation, allowing amino acid sequence information to be determined in the form of b- and y-type ions in conjunction with glycan fragmentation. Electron transfer dissociation also has proven useful in the analysis of glycopeptides, wherein fragmentation occurs primarily through c- and z-type peptide fragments, with the glycan remaining largely intact. Further evaluation of released and derivatized glycans can provide an additional high level of structural detail. A more thorough discussion of glycans and detailed techniques for their analysis and quantification can be found in the Glycosylation chapter/Volume 2, Chapter 4.

### Loss of Lysine on the Carboxyl Tail

C-terminal Lys variants are frequently observed in mammalian cell-derived recombinant antibodies (they may not be observable in other host cell types such as *Escherichia coli* [*E. coli*]). The C-terminal Lys is readily clipped during the manufacturing process, likely due to proteolysis by endogenous carboxypeptidase(s) during cell culture (47, 48). Trace element levels of metals, specifically increased copper or decreased zinc, have been identified as major factors contributing to increased levels of intact C-terminal Lys (48). Studies have shown no effect of C-terminal Lys variation on antibody function, thus it is not considered a critical quality attribute (49, 50). This variation adds to mAb heterogeneity, however, so it is still an important factor to include in the characterization parameters. Because Lys is readily charged, its loss results in a decrease in positive charge that allows for resolution of the modified and unmodified structures by charge-based separation techniques such as ion exchange chromatography (IEX), isoelectric focusing (IEF), and cIEF. In addition, the -128 Da mass shift that occurs upon loss of this terminal residue also can be used to detect and quantify modified versus unmodified forms through the application of MS techniques, including peptide mapping and intact mass analyses.

### Cyclization of Glutamine and Glutamate (Glutamic Acid)

N-terminal pyro-Glu is a cyclic product derived from the rearrangement of Gln or Glu at the N-terminus of a mAb. The reaction can occur spontaneously or be catalyzed enzymatically by glutaminyl cyclase (51–53). When this proceeds spontaneously, the rate of cyclization is much faster with Gln than with Glu. Strategies for cyclization of N-terminal Gln and Glu are often considered when designing peptide therapeutics due to indications that this modification confers increased stability upon peptides, particularly by protecting them from destruction by peptidases (54, 55).

Conversion from N-terminal Gln to pyro-Glu is usually near complete in mAbs (> 95%), and it is known that the conversion occurs primarily inside bioreactors with little contribution from downstream purification and analytical processes. Cyclization that occurs in a bioreactor is more likely to be spontaneous rather than enzyme catalyzed and can be accelerated by higher temperature as well as cell culture media and buffers containing sodium phosphate or ammonium carbonate, although the change in pH results in a minimal effect on the reaction rate (51).

The rate of Glu to pyro-Glu conversion *in vitro* near physiological pH and temperature is comparable to that *in vivo* (52), and pyro-Glu levels are shown to increase with increasing temperature and exhibit pH dependence as well (56). One study has measured faster conversion rates for native conformation of heavy chain compared to the light chain. The rate of this process remained steady for the heavy chain upon denaturation, whereas light chain Glu to pyro-Glu rates increased to levels near that of the heavy chain. This indicates that Glu cyclization can be affected by mAb structure (52).

The Gln conversion to pyro-Glu renders antibodies more acidic, whereas the conversion of Glu to pyro-Glu results in a basic shift. In both cases, the heterogeneity introduced is observable in charge-based assays such as CEX, provided that mobile phase buffers are prepared at a suitable pH. As a result, N-terminal pyro-Glu can be closely monitored during process control. Cyclization of the N-terminus also is readily observable using MS-based peptide mapping because the conversion of Gln or Glu to pyro-Glu results in a mass shift versus the unmodified peptide of -17 or -18 Da, respectively.

## Deamidation and Iso-Aspartate Formation

Deamidation of asparagines is commonly observed and has an important role in regulating the heterogeneity and stability of recombinant mAbs (57–60). Glutamines are also susceptible to deamidation but at a much lower rate unless subjected to particularly harsh conditions such as extreme pH (61, 62). pH, buffer type, and temperature are known factors that can affect the rate of Asn deamidation (62, 63). Primary, secondary, and tertiary structures also are determinants of deamidation, with the most rapid conversion rates occurring when the residue C-terminal to the Asn is a Gly or Ser (64).

Native Asp itself can also undergo isomerization to form isoaspartic acid (iso-Asp), and is another ubiquitous modification that can result in heterogeneity in mAbs (65). The sequences most sensitive to isomerization include Asp-Gly, Asp-Ser, and His-Asp (66, 67). Native Asp first converts to a cyclic imide intermediate, and then either hydrolyzes back to Asp or isomerizes to iso-Asp. In addition, Harris, et al., determined that isomerization of Asp 102 in a heavy chain CDR3 region of IgG1 Herceptin reduced its potency to 70%, causing serious implications on drug efficacy (68). It is likely due to the fact that isomerization results in insertion of an additional methylene group into the backbone, which can influence protein stability and structure. A decrease of antigen binding of several antibodies has also correlated with isomerization of Asp residues (69–72).

Product heterogeneity caused by uncontrolled degradation via deamidation and isomerization can complicate manufacturing consistency. Liu et al. recently reviewed sources of molecular heterogeneities often exhibited by mAbs, including deamidation and isomerization (73). Deamidation and isomerization can occur at different stages of the mAb production process and are one of the major degradation reactions that occur during long-term storage. Physical and chemical stresses generated during manufacturing and storage conditions can compromise the stability of biotherapeutic products. In this regard, formulation strategies aimed at minimizing these effects and assuring appropriate shelf-life of a drug product must be carefully evaluated (74).

Deamidation can be either acid or base catalyzed (63, 75, 76). Under mildly acidic conditions, it involves direct hydrolysis of Asn to produce mainly Asp. At neutral to basic pH, Asn deamidation is believed to proceed through the formation of a metastable cyclic-imide intermediate (succinimide [Asu]) as the result of nucleophilic attack of the carbonyl Asn side-chain by an amide nitrogen C-terminal to Asn. Succinimide formation is followed by its rapid hydrolysis and the formation of  $\alpha$ - (Asp) and  $\beta$ - (iso-Asp) aspartates at an approximately 1:3 molar ratio (63, 77). The rate of succinimide formation is dependent on protein sequence as well as secondary and tertiary structures (78–81). Steric effects of the N+1 residue and conformational restrictions reduce nucleophilic reactivity of the backbone NH centers and, thus, reduce the rate of succinimide formation. Generally, amino acid motifs where the N+1 residue is Gly, Ser, and Asn are the most susceptible to deamidation in mAbs (59, 79, 81–84). Although the majority of iso-Asp formation in a mAb was related to the deamidation of Asn, especially in flexible regions of a molecule (85), the importance of aspartate isomerization on product quality should not be underestimated. Susceptibility of the Asp-Gly sequence to isomerization has been demonstrated as a pH-dependent reaction (77, 86). Further, studies on an IgE have indicated a reduction in binding affinity due to isomerization of an Asp residue in the complementarity-determining region (CDR) of the antibody (69).

Deamidation and isomerization are often a quality-defining attribute of biotherapeutic products, and several analytical techniques may be used to monitor these processes. IEX and IEF have been successfully applied to study the charge microheterogeneity caused by deamidation at the intact level (67, 68). Separation of intact mAbs from their aspartate isomerization products has been achieved using hydrophobic interaction chromatography (HIC) (71). LC-MS/MS peptide maps are routinely used for identification and quantification of deamidation and isomerization products, including the succinimide intermediate.

Edman sequencing can be exploited to distinguish Asp from iso-Asp at the peptide level because the methylene incorporated into the peptide backbone during isomerization prevents cyclization of the phenylthiocarbonyl peptide to form the necessary anilinothiazolone derivative, thus the presence of iso-Asp will terminate the degradation reaction (87). Iso-Asp can also be isotopically labeled, and thus distinguished from Asp, by reaction with protein L-isoaspartyl methyltransferase (PIMT) in the presence of S-adenosyl-L-[methyl-<sup>3</sup>H]methionine (88, 89). The enzymatic addition of a methyl group to an iso-Asp residue not only results in

a differentially labeled peptide, but increases its hydrophobicity and alters its retention time.

It should be noted that deamidation is one of the most challenging PTMs to characterize using MS-based techniques. Deamidation results in conversion of  $-\text{NH}_2$  to  $-\text{OH}$  (+0.984 Da), which has a very similar mass shift as the first  $^{13}\text{C}$  isotope peak of the native peptide (+1.0034 Da). Mass spectral resolution of the deamidated monoisotopic peak from a native peptide cannot be achieved unless the mass spectrometer has ultra-high resolving power ( $> 0.5$  million), far exceeding the range of the widely used TOF or Orbitrap instruments. The isomerization of Asp to iso-Asp demands more complicated analysis due to their isobaric nature and lack of abundant and uniquely identifying MS<sup>2</sup> fragment ions. It was shown that intensity ratios of complimentary b and y ions exhibit characteristic shift due to the presence of iso-Asp (90), however, this method requires use of standards. Recently Cournoyer et al. (91–93) and O'Connor et al. (94) applied respectively electron capture dissociation (ECD) and ETD fragmentation to identification of aspartate isomerization. Presence of iso-aspartate resulted in cleavage along the C $\alpha$ -C $\beta$  backbone of the iso-Asp and the formation of characteristic (c+58) and (z-57) complimentary ions, which allowed differentiation of the two isoforms of aspartate. Nevertheless, interpretation of these spectra is rather complex and thus not amenable to high-throughput analyses. Ultimately, most researchers rely on retention time shifts in RP chromatography to differentiate these modified residues, which highlights the importance of chromatographic separation for resolution of these isomers. In RP chromatography, two deamidation peaks, Asp and iso-Asp, usually elute in a narrow retention time window along with the native peak. The elution order varies depending on peptide sequence, column, and ion-pairing modifier, but typically peptides containing iso-Asp elute earlier, followed by the native Asn peptide, and finally the peptide bearing Asp. When multiple asparagines are present in a single peptide, the peak assignment can be more complicated. This is readily apparent in the extracted ion chromatogram (XIC) of the *m/z* corresponding to native (top) and deamidated (bottom) H(374-395) peptide, GFYPSDIAVEWESNGQPENNYK (a.k.a. “pennyK peptide”), of the NISTmAb heavy chain (Figure 1). Peak assignments are based on literature and experience (95, 96). Five individual deamidation products are observed, two of which co-elute with the native peptide and would therefore result in an overestimate of native peptide in a typical sample.

Deamidation is a spontaneous process which can be promoted at the peptide level by basic pH during enzymatic digestion, masking true degradation of the intact protein. One strategy to differentiate “naturally” occurring degradation from artificially induced modification during sample preparation has been recently developed (59, 81, 97, 98). Here, deamidation products native to the intact protein were discriminated from method-related deamidation by inducing hydrolysis of succinimide moieties on an intact mAb in the presence of H<sub>2</sub><sup>18</sup>O prior to performing tryptic digestion in H<sub>2</sub><sup>16</sup>O (81). The 2 Da shift between native and artificial deamidated residues were subsequently identified using LC-MS/MS techniques.

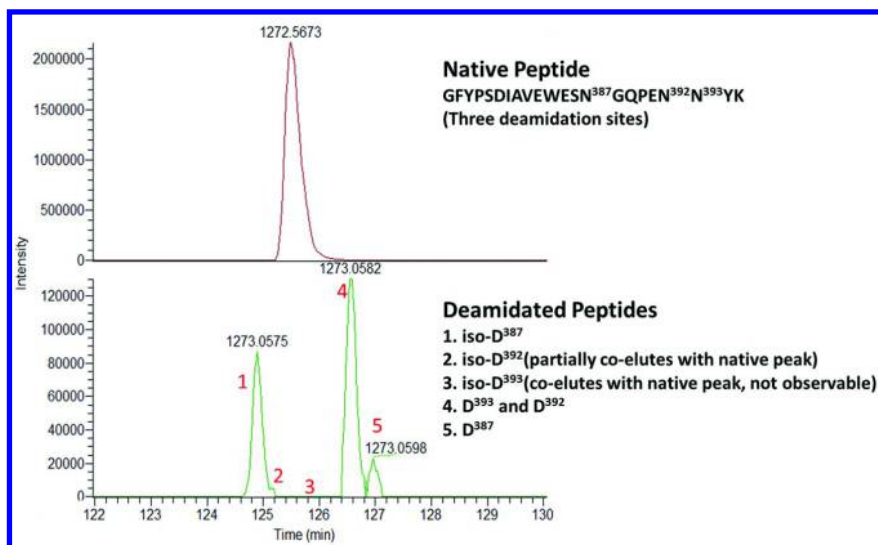


Figure 1. Deamidation of GFYPSDIAVEWESNGQPENNYK peptide in NISTmAb heavy chain Fc region. Top: extracted ion chromatogram (XIC) of peptide GFYPSDIAVEWESNGQPENNYK (native peptide); Bottom: XIC of deamidation products and assignments. All peptides eluted within about a 2.5 minute window.

## Oxidation

Proteins in biological systems are known to be one of the major targets of endogenously generated reactive oxygen species. Radicals and non-radical oxidants can be generated by a wide variety of different cellular processes (99–102) as well as exogenous agents, including radiation (X-ray,  $\gamma$ , UV light, visible light in the presence of a sensitizer), metal ions, and solvents (103–107). Exposure of proteins to attack by free radicals in the presence of oxygen can result in the oxidation of amino acid side-chain groups, peptide backbone fragmentation, protein cross-linking, unfolding, changes in hydrophobicity and/or conformation, and altered susceptibility to proteolytic enzymes (103, 108, 109). The complexity of the chemical pathways involved in amino acid oxidation has been well-described (105, 110). Although all amino acids can undergo oxidation to some degree under certain conditions, those most susceptible to oxidation are the sulfur-containing residues (Cys, Met), the aromatics (Phe, Trp, Tyr), and His. The rate at which this occurs in proteins for any given residue can differ widely depending on the properties of the individual amino acid and neighboring residues, the source of oxidative stress, and environmental parameters such as temperature or pH (103, 105, 106, 111). For example, Met oxidation is favored by peroxides under acidic conditions, whereas cysteines will oxidize more readily via hydrogen peroxide ( $H_2O_2$ ) with increased pH (112). His typically requires metals such as Cu(II) for oxidation to proceed via  $H_2O_2$ , or alternatively, Cu(II) may be used in the presence of an ascorbate sensitizer to initiate His oxidation (113–116).

His oxidation also was recently demonstrated in a mAb upon treatment with a light stress platform (2000 lux (D65 illuminant)/25 °C for 25 days) (117). It was hypothesized that the oxidation pathway was initiated by singlet oxygen ( $^1\text{O}_2$ ) generated from dissolved oxygen upon UV/Vis irradiation (117).

The most commonly oxidized amino acids in biopharmaceuticals are Cys (discussed later in regard to disulfide bonds) and Met, though some Trp and His oxidation products also have been observed. Oxidation of biotherapeutic proteins can alter their physical and biological properties, affecting their potency and stability characteristics (118–120). In the case of human IgG1, two Met residues localized to the CH<sub>2</sub> and CH<sub>3</sub> domains of the Fc region (corresponding to M255 and M431 in the NISTmAb heavy chain) are known to be highly susceptible to oxidation (121–123). Increased aggregation, shifts in melting temperature, increased deamidation rates, and conformational changes have been observed upon oxidation of these residues and highlight the impact of this modification on IgG structure as well as physical and covalent stability (124). Moreover, chemically induced Met oxidation has been demonstrated to decrease the binding affinity of an IgG1 and IgG2 to the neonatal Fc receptor (FcRn) (125, 126) and significantly reduce the serum circulation half-life of IgG1 (127).

The susceptibility of Met residues to oxidation is routinely probed during formulation development by incubation of a mAb in the presence of H<sub>2</sub>O<sub>2</sub> or tert-butyl hydroperoxide (t-BHP) (121, 123, 128). The use of one oxidizing agent versus another is not a trivial choice and is considered with the end goal in mind because their unique specificities will produce differing results. Although both reagents tend to specifically oxidize Met, H<sub>2</sub>O<sub>2</sub> more readily oxidizes less accessible, buried residues, whereas t-BHP is known to target more surface-exposed Met (123, 129, 130). Thus, multiple methods are often recommended to fully evaluate oxidation susceptibilities under various conditions and different exposure types. Following treatment with an oxidizing agent, those Met residues prone to oxidation are determined, the effect of this change on potency and other pharmacological properties of the mAb are evaluated, and strategies are developed to prevent oxidation at these residues.

Oxidized peptides are easily identified using MS since the addition of one oxygen atom to the Met side-chain upon conversion to Met sulfoxide increases the mass of the affected residue by +16 Da (131), and under harsh conditions, the addition of two oxygen atoms will form Met sulfone (+32 Da) (103). At the MS/MS level, Met sulfoxide-containing peptides readily lose methane sulfenic acid (CH<sub>3</sub>SOH) upon CID fragmentation and are thus easily identified by a characteristic loss of 64 Da from the parent mass (132). Although this neutral loss fragment is diagnostic for the presence of Met sulfoxide versus an isobaric Phe residue, it is often the largest ion observed in the MS/MS spectrum. The corresponding lack of peptide backbone fragmentation can impede confirmation of the peptide sequence and also prevent unambiguous assignment of oxidation to a specific residue. This phenomenon can be avoided by fully oxidizing Met sulfoxide to Met sulfone with performic acid (133) because Met sulfone is not subject to neutral loss. However, this additional oxidation step can be avoided by using ETD or ECD to fragment Met sulfoxide peptides along the backbone without the 64 Da neutral loss (134).

Oxidation of Trp residues in biotherapeutics occurs less often than with Met and usually requires different types of oxidizing conditions, such as elevated temperature (135), UV/Vis light (136) or the presence or in situ generation of free radicals (137) to become significant. Even excipients such as polysorbates can become a source of oxidative stress if they undergo autoxidation to form peroxides or free radicals (138–140). One pertinent study showed that alkyl peroxy radicals generated by the autoxidation of polysorbate 20 (PS20) during storage at room temperature or by increased PS20 levels in the formulation were responsible for the Trp oxidation observed in a recombinant humanized Fab fragment and was facilitated by the metal-binding properties of a neighboring His residue (141). This effect was specific to the mechanism of oxidation because oxidation via hydroxyl radicals was not observed when the Fab was exposed to H<sub>2</sub>O<sub>2</sub> or t-BHP.

Characterization of Trp oxidation is analytically challenging because multiple oxidation products usually coexist, such as kynurenine (Kyn) (+4 Da), 5-hydroxyl-tryptophan (5-OH-Trp) (+16 Da), oxindolylalanine diastereomomers (Oia) (+16 Da), 3-hydroxykynurenine (3-OH-Kyn) (+20 Da), *N*-formylkynurenine (NFK) (+32 Da); dioxindolylalanine diastereomers (DiOia) (+32 Da); dihydroxytryptophan (+32 Da) hydroxy-*N*-formylkynurenine (+48 Da) and dihydroxy-*N*-formylkynurenine (+64 Da) (142–144). Alternatively, the indole ring of Trp can undergo oxidation via loss of two hydrogen atoms, generating species with –2 Da (didehydro-tryptophan) (145) or +14 Da (oxolactone) mass shifts (146). Thus, a comprehensive understanding of the Trp reaction pathway is essential for the characterization of peptides with modified Trp residues (144, 146).

As with Met oxidation, it is important to understand the potential for Trp oxidation in mAbs because the effects can be rather dramatic. A key study has shown that oxidation of a single Trp residue in the CDR3 region of a humanized mAb was achieved by UV irradiation or chemical modification via ozone exposure. The resulting ligand binding affinity of the antibody was decreased to 28% of control values, and its potency was measured at 26% of the control after a seven day exposure (147). One strategy commonly used to induce Trp oxidation for subsequent study includes metal-catalyzed oxidation, typically using Fe(II) in the presence of H<sub>2</sub>O<sub>2</sub>, or by using 2,2'-azobis(2-amidinopropane) dihydrochloride (AAPH) (128).

Extreme levels of Trp oxidation can be visualized via changes in solution color. A predominance of NFK and OH-Trp degradation products is indicated by a yellow color, whereas brown reveals the presence of Kyn (145, 148). Of course, more subtle changes in oxidation levels can be detected by analytical methods more sensitive than the eye. Changes in UV/Vis absorption or fluorescence between oxidized and unmodified proteins or peptides are often used as indicators of the presence of Trp oxidation (143, 145, 149), and the separation of Trp oxidized peptides using various chromatographic platforms will result in shifts in retention time as compared to unmodified proteins or peptides (135, 150). However, MS analysis is necessary to identify particular oxidation products or to assign them to a specific Trp residue (145, 149, 151). The mass shifts detailed earlier are easily detected in the parent mass of oxidized peptide ions to identify Trp degradants, whereas isobaric species (e.g., 5-OH-Trp, Oia) can be



distinguished from one another by immonium ions as well as v- and w-series ions resulting from high-energy CID fragmentation (152–154).

Oxidation can happen at nearly all stages of mAb development, such as during cell culture, purification, storage, or even in the process of analytical assays (e.g., in-source oxidation in mass spectrometer ESI sources, on-column during chromatographic separation) (144, 146, 155). Typically, Met and Trp are among the most closely tracked residues. To minimize the extent of oxidation, special attention should be paid to limit exposure to heat, light, and other oxidative stressors during production, storage, and transportation.

## Glycation

Glycation refers to the non-enzymatic addition of a monosaccharide to an amino acid residue via the Maillard reaction (156). This occurs when an unstable Schiff base intermediate is formed via condensation of a reducing sugar carbonyl group with a free amine group (e.g., protein N-terminus,  $\epsilon$ -amino group of Lys or Arg) and then rearranges to form a ketoamine derivative known as an Amadori product (157–172). The most prominent glycation product found in recombinant mAbs tends to be the Amadori product  $N^\epsilon$  fructosyl-Lys, which arises from glycation of a Lys residue and is thus the main focus of PTM studies for biotherapeutic antibodies (173–175).

The main source of this commonly observed PTM in biopharmaceutical production is the reducing sugars included in cell media (e.g. glucose, galactose) (157, 174, 176). The extent of glycation on an antibody can be controlled to some degree by glucose feed strategies as well as through engineering, such as exchanging a Lys for an Arg when possible (157, 165, 174, 177, 178). Glycation appears to have little to no effect on properties such as mAb target binding, in vitro potency or FcRn binding (173, 174, 179), but remains a PTM of interest because it introduces heterogeneity into the sample, particularly in regard to the charge profile (165, 173, 174), such that filing regulations necessitate characterization of the PTM (180).

Unlike the enzymatic addition of N-glycan structures to an Asn residue, a consensus sequence of amino acids is not required for glycation; thus in theory, any otherwise unmodified Lys is susceptible to this PTM. However, the rate of individual modification varies depending on local sequence (e.g., proximity to His or Asp) and higher order structure (e.g., surface exposure) (157–160, 181–184).

Separations of glycated species can be performed using a number of different methods. Because Lys glycation results in loss of a positive charge, the resulting charge difference can be exploited to separate glycated mAb from unglycated mAb using capillary electrophoretic techniques or IEX (185). Boronate affinity chromatography has also been used for separation and analysis of glycated intact mAbs and peptides (157, 160, 174, 186, 187).

The conversion of Lys to  $N^\epsilon$  fructosyl-Lys is readily detected using MS techniques due to the characteristic +162 Da shift in mass. Although this can be measured using intact or subunit analysis if the modified species is sufficiently abundant, this mass shift also can arise from heterogeneity in the number of hexose residues comprising the population of N-glycans conjugated to the mAb

(e.g., F1A1G2 has a +162 Da mass shift compared to F1A1G1). However, Lys glycation can be confirmed at the intact or subunit analysis level if a +162 Da species remains after treatment of the sample with PNGase F to remove N-linked glycans (165, 175). Of course, to assign the glycation moieties to specific Lys residues, bottom-up methods must be used. In the case of a +324 Da shift, the increased mass could indicate two glycations on the same IgG molecule or could be attributed to retention of a Val-His-Ser sequence (+323.15935 Da) at the N-terminus of either antibody chain if non-specific cleavage of the signal peptide occurs during cellular processing (157, 188).

Typically, individual glycated peptides have low abundance, which can make them difficult to detect as well as quantify. In addition, CID fragmentation of a glycated peptide often produces only minimal peptide backbone fragmentation (Figure 2). Instead, a number of water losses and characteristic ions formed by intramolecular rearrangement of the glucose moiety (-54.031 Da and -84.042 Da) tend to dominate the spectra (189). The corresponding lack of y- and b-ions also can hinder detection by software algorithms if these fragments are not intense enough or do not provide enough coverage of the predicted ion series to pass statistical filters. Some studies have addressed this issue by using neutral losses of H<sub>2</sub>O to trigger MS<sup>3</sup> or multistage activation events to obtain more complete ion series information upon compilation of the resulting spectra (190). Alternatively, derivatization with sodium borohydride and phenylboronic acid has been used to stabilize the sugar moiety, which not only improves y- and b-ion fragmentation but also causes a shift in RP-HPLC retention time to aid in the detection of glycated peptides in the peptide map (176, 179, 191). The challenge of generating informative ion series fragmentation in MS/MS also can be overcome using alternate fragmentation methods such as ETD, which produces sufficiently informative c- and z-ions for peptide identification (186, 187, 192, 193), or HCD, which generates spectra including both the neutral loss ions characteristic of the sugar adduct as well as a portion of predicted y- and b-ions.

The presence of glycation prevents trypsin and Lys-C cleavage at the C-terminus of the modified residue (184, 185, 194). Thus, if peptide maps of glycated and unglycated samples are compared, a new peak will typically appear to indicate the presence of the glycated missed cleavage peptide. Although digestion with these enzymes may be useful for easily visualizing glycated peptides, they are not the enzymes of choice for quantitation of this PTM. Because the modified peptide comprises a missed cleavage and the unmodified species represents only one-half of that peptide, a comparison is essentially being made between two different peptides. To attempt quantitation of glycated and unglycated peptides, the first decision to make would be which of the two unmodified peptide halves that comprise the modified peptide should be used for the calculation. Some methods suggest that the XIC should be generated using the longer of the two peptides (195), although it also could be argued that the cleaved peptide half containing the residue that is modified in its counterpart should be included. One study quantified relative glycation abundance by dividing the XIC of the glycated peptide by the sum of XICs of the two halves of the unmodified, fully tryptic peptides (173). Regardless, it is almost certain that the longer, modified peptide and the shorter, unmodified peptide components have different

ionization efficiencies, and it is difficult to predict how reliable the comparison between the two abundances would be. Fortunately, with the availability of alternate enzymes such as chymotrypsin or Glu-C, modified and unmodified peptides of the same sequence can be generated and used for better quantitative outcomes (168, 179, 196). Methods involving forced glycation with isotopically labeled reducing sugars have also been developed for quantitation of this PTM (175, 197).

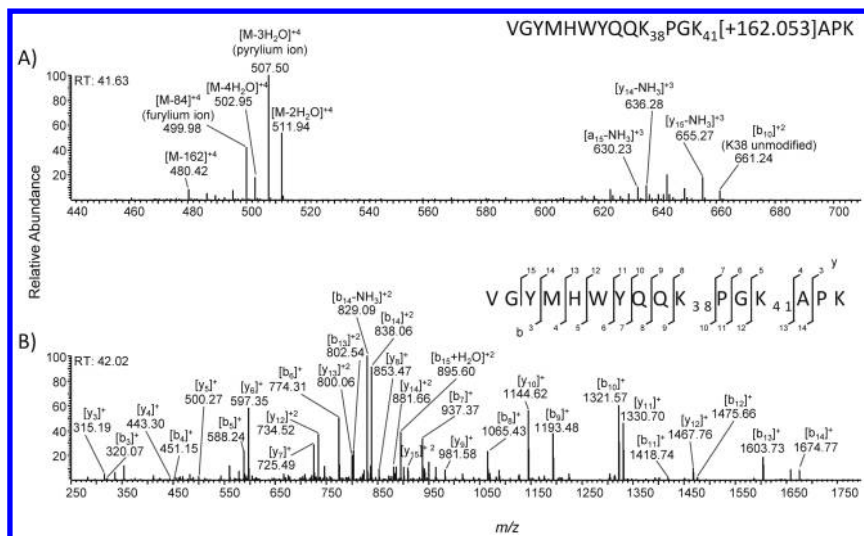


Figure 2. Collision-induced dissociation (CID) spectra of glycosylated and unglycosylated peptides. (A) Fragmentation of the NISTmAb glycosylated peptide L (29–44) (VGYMHWYQQKPGKAPK) results in preferential cleavage of the hexose moiety over the peptide backbone, with characteristic product ions representing sequential losses of water from the adduct as well as a  $-84$  Da ( $-3\text{H}_2\text{O}-\text{HCHO}$ ) neutral loss resulting from intramolecular rearrangement of the hexose moiety. Low levels of peptide backbone fragmentation include the  $b_{10}^{+2}$  ion at  $m/z$  661.24, indicating that K38 is unmodified, thus suggesting that the modification is localized to K41. (B) Nearly complete  $y$ - and  $b$ -ion series are generated by CID fragmentation of the unglycosylated peptide. The parent ion selected for fragmentation in panel A is the quadruply charged, first monoisotopic peak at  $m/z$  521.0104; the parent ion fragmented in panel B is the doubly charged peak at  $m/z$  959.4888. Subscripts in the peptide sequence indicate the position of Lys residues within the light chain sequence.

## Variants Involving Cysteines

Knowledge of disulfide bond structures characterized decades ago for immunoglobulins has been augmented by the more recent identification of free sulfhydryls, cysteinylations, thioether formation, and trisulfide bonds. A disulfide bond can undergo reversible  $\beta$ -elimination to initially form one dehydroalanine

and one persulfide on constituent chains. Continued degradation of the persulfide converts into a Cys residue (free sulfhydryl) representing a point of no return for the native disulfide. At this stage, the original dehydroalanine may react with the newly formed free sulfhydryl to form a non-reducible thioether bond (198). Free sulfhydryls may alternatively become covalently modified by a free Cys in the solution, which is referred as “cysteinylation” (199, 200), or undergo a second round of  $\beta$ -elimination to form a second dehydroalanine. Subsequent hydrolysis of dehydroalanine residues may contribute substantially to fragmentation of the antibody hinge region to produce Fab and Fab-Fc components and is accelerated by heat and increasing alkaline pH (37, 201–203). The effects of fragmentation induced by these Cys modifications on the biological activity of mAbs have yet to be elucidated; it stands to reason that the Fab fragment alone will not have Fc-mediated effector properties, however, whereas the Fab-Fc component may display at least some decrease in potency (204). Although these detrimental Cys modifications were not detected in this study, typical disulfide bonding of the NISTmAb was addressed in the Primary Structure chapter/Volume 2, Chapter 1.

Trisulfide bonding is a modification formed by insertion of a sulfur atom into an existing disulfide bond. Though seldom affecting mAb functions, trisulfides are commonly observed in all IgG subtypes and predominantly found in the linkages between heavy-heavy chains and heavy-light chains (205). The level can reach as high as 39%. The generation of trisulfides is correlated with Cys levels in cell culture media because Cys can contribute to the production of hydrogen sulfide, which then reacts with mAbs. Optimizing Cys feed strategies can minimize these trisulfide variants, eventually leading to lower heterogeneity for the target molecule (206). In the current study, trisulfides were not observed in the NISTmAb.

## Fragmentation of Intact mAb

Clipping, also known as protein fragmentation, is a non-enzymatic degradation pathway through which peptide bonds are cleaved via specific mechanisms (204). Although essentially all of the peptide bonds in the product can be degraded, rates vary over several orders of magnitude. It was reported that specific residue pairs in mAbs are especially prone to peptide bond cleavage, including Asp followed by Gly, Ser, Val, Tyr, Phe, Pro, Asn, Leu, Asp, or Lys; Pro or Gly followed by Asp; Gly followed by Gly, Ala, Leu, or Val; Ala, Leu, or Ser followed by Gly; Gly followed by Tyr; Gly, Ala, or Leu followed by Ser; Cys followed by Cys; and Asn followed by Ser, Val, Leu, or Pro. (204). The cleavage rates are sequence-, pH-, and temperature-dependent (207). A frequently observed and monitored attribute is Asp-Pro cleavage, which tends to accelerate at acidic pH. cSDS is a frequently used technique to detect and monitor protein clips. Intact/reduced mass analysis (by LC/MS as well as MALDI-TOF) also can provide very useful information on the presence of large fragments resulting from Asp-Pro clipping.

## Hydroxylation

Hydroxylation products initially detected in collagen, including hydroxyprolines and hydroxylysines, also can be detected in mAbs but at more reduced levels. Although the +16 Da mass shift is the same as for oxidation, all hydroxylation reactions are enzyme-catalyzed by prolyl or lysyl hydroxylases in the lumen of the endoplasmic reticulum following protein synthesis. These modification levels therefore can only be manipulated in clone selection and at the cell culture stage. Downstream processes such as purification or storage may not affect hydroxylation levels (208, 209). Although little is known about the effect of hydroxylation on mAb immunogenicity and potency, unusually high hydroxylation levels should be avoided, especially for those residues in the CDR regions. It should be noted that hydroxylation was not observed in the NISTmAb in this study, and is therefore not included in the interlaboratory discussion below.

## Interlaboratory PTM Characterization Using Peptide Mapping

The summary above clearly demonstrates that mAbs typically contain various PTMs that can contribute to heterogeneity and potentially affect drug efficacy, safety, and clearance rates. It is of paramount importance to characterize, and inevitably ensure consistent PTM levels. This is especially true when these modifications occur in regions that are involved in mechanism of action (e.g., CDRs that recognize and bind antigens, Fc gamma receptor binding sites involved in antibody-dependent cellular cytotoxicity [ADCC], C1q binding sites involved in complement-dependent cytotoxicity [CDC]). To ensure product quality, detrimental modifications must be closely tracked and monitored during process development, manufacturing, transportation, and storage (182).

### Purpose of This Study

To demonstrate the use of peptide mapping for detection and determination of PTM levels present in a mAb, several laboratories collaborated to perform trypsin peptide mapping on the NISTmAb. Peptide mapping is a routine assay that every biopharmaceutical laboratory is capable of performing; however, a variety of steps in sample processing, instrumental analysis, and data interpretation exist. In an effort to gain a broad representation of peptide mapping practices currently utilized in the biopharmaceutical industry, participating laboratories were asked to analyze the samples according to their routine analytical work flows without specification of a uniform method. The interlaboratory comparison was therefore intended to:

1. Evaluate interlaboratory PTM identifications and quantifications in the absence of a predefined protocol to gain insight into the capabilities of current analytical techniques.
2. Highlight perturbations and variances in peptide mapping techniques that may affect results, such as differences in sample handling, instruments and software platforms.

3. Demonstrate the power of peptide mapping for PTM characterization and the factors to consider when interpreting data.
4. Use the results to identify areas in which additional, targeted interlaboratory studies may be used to further elucidate best practices and selectively evaluate the aspects leading to variability.

## Samples and Experimental Conditions

Our interlaboratory PTM characterization study was based on the NIST reference mAb (control). Each laboratory performed peptide mapping and PTM identification/quantification studies using their individual platform methods and returned their results for comparison with the other laboratories. Because the main focus was to evaluate the comparability of results between, rather than within, laboratories, replicates were not required. However, Lab 3 generated triplicate analyses from separate tryptic digests performed with the same protocol as an intra-laboratory metric against which we could evaluate the interlaboratory results.

It should be highlighted that no uniform peptide mapping protocol was provided for this study. As a result, although all of the distributed samples were subjected to reduction, alkylation, trypsin digestion, LC-MS/MS analysis, and database search processes, significant differences in method details were used by each laboratory. It is well known that sample preparation can contribute to oxidation, deamidation, and other PTM artifacts. Therefore, the differences in methods utilized in this study are important factors to consider in order to achieve a meaningful interpretation of the results. Summaries of the peptide mapping protocols and instrumental parameters employed by each laboratory are given in Table 2 and Table 3. Lab # designations were randomly assigned.

**Table 2. Sample Preparation and Digestion Details**

<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>
<i>Denaturation and Reduction</i>		
Denaturation and Reduction		
Approximately 0.1 mg of the mAb was buffer-exchanged to 6 mol/L guanidine HCl, 250 mmol/L Tris and 1 mmol/L EDTA at pH 7.5 (pH adjusted with NaOH) using Bio-spin P6 columns (Bio-Rad) according to manufacturer's specifications.	500 µg of the mAb was denatured with 7.5 mol/L guanidine HCl, 250 mmol/L Tris and 1 mmol/L EDTA at pH 7.5. Total volume 500 µL.	Approximately 0.1 mg of mAb sample was mixed with 300 µL of 8 mol/L guanidine HCl with 4 mmol/L EDTA prepared in pH 8.0, 1X DPBS.

*Continued on next page.*

**Table 2. (Continued). Sample Preparation and Digestion Details**

<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>
2.0 $\mu\text{L}$ of 500 mmol/L DTT was added for a final DTT concentration of 9.8 mmol/L. The sample was reduced at 37 °C for 30 min.	3 $\mu\text{L}$ of 500 mmol/L DTT was added. The sample was reduced at room temperature for 30 min.	Sample was reduced at 37 °C for 60 min after adding 10 $\mu\text{L}$ of 1.0 mol/L DTT for a final DTT concentration of 24 mmol/L.
<i>Alkylation (Cysteine Capping)</i>		
4.8 $\mu\text{L}$ of 500 mmol/L IAM was added for a final IAM concentration of 22 mmol/L. Alkylation proceeded in the dark for 30 min at room temperature. The reaction was quenched with 2 $\mu\text{L}$ of 500 mmol/L DTT.	The sample was alkylated with the addition of 7 $\mu\text{L}$ of 500 mmol/L IAM for 15 min at room temperature in the dark.	Sample was alkylated by adding 24 $\mu\text{L}$ of 1 mol/L IAM (final concentration of 55 mmol/L) and incubating for 60 min at room temperature in the dark. The reaction was quenched with 15 $\mu\text{L}$ of 1 mol/L DTT.
<i>Buffer Exchange</i>		
The sample was buffer exchanged at room temperature into freshly prepared digestion buffer containing 2 mol/L urea in 100 mmol/L Tris, pH 7.8 (pH adjusted with NaOH) using Bio-spin P6 columns (Bio-Rad) according to manufacturer's specifications. Digestion buffer was added for a final IgG concentration of 1.42 $\mu\text{g}/\mu\text{L}$ .	The sample was buffer exchanged at room temperature into freshly prepared digestion buffer containing 50 mmol/L Tris, pH 7.5 using NAP-5 columns according to manufacturer's specifications. Final mAb concentration $\sim$ 0.8 $\mu\text{g}/\mu\text{L}$ .	The sample was buffer exchanged into 50 mmol/L Tris, 1 mmol/L $\text{CaCl}_2$ , pH 8.0 using a Zeba spin desalting column.
<i>Enzyme Digestion (Trypsin)</i>		
Sequencing grade modified trypsin (Promega Gold) was prepared in 0.05 mol/L acetic acid at a concentration of 1.0 $\mu\text{g}/\mu\text{L}$ . 2 $\mu\text{L}$ (2ug) trypsin was added to a 50 $\mu\text{L}$ (70 $\mu\text{g}$ ) aliquot of the reduced and alkylated IgG (enzyme to protein ratio = 1:35 [w/w]). Digestion proceeded using the CEM Discover (model 908005) microwave hydrolysis system for 30 min at 50 °C, 50 W power, with $\Delta$ temp of 5 °C, high-speed stirring, cooling	The sample was digested with Roche sequencing grade trypsin (dissolved in water) added at a 1:20 enzyme to protein ratio at 37 °C for 30 min. 10 $\mu\text{L}$ of TFA (30% v/v) was added to quench the trypsin reaction.	To each sample, 1 $\mu\text{L}$ of 1mg/mL Promega sequencing grade modified trypsin in 50 mmol/L acetic acid was added. Digestion was allowed to proceed for 2 hours at 37 °C. 0.5 $\mu\text{L}$ of TFA per 50 $\mu\text{L}$ of sample was used to quench the trypsin reaction.

*Continued on next page.*

**Table 2. (Continued). Sample Preparation and Digestion Details**

<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>
on, and in SPS mode. The tryptic digest was dried to completion using vacuum centrifugation, then stored overnight at $-20\text{ }^{\circ}\text{C}$ . Dried samples were resuspended in $142.8\text{ }\mu\text{L}$ of $0.1\%$ TFA for a final concentration of $0.5\text{ }\mu\text{g/mL}$ .		

DTT = dithiothreitol, DPBS = Dubelco's Phosphate Buffered Saline obtained from Life Technologies, EDTA = ethylenediaminetetraacetic acid, iodoacetamide = IAM, mAb = monoclonal antibody, TFA = trifluoroacetic acid, Tris = tris(hydroxymethyl)aminomethane.

**Table 3. Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis and Database Searching Details**

<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>																																																		
<i>LC-MS Analysis</i>																																																				
LC instrument: Dionex UltiMate 3000 UHPLC MS: Thermo LTQ-Orbitrap Elite Column: Supelco Bio Wide Pore C18, $2.1\text{ mm ID} \times 150\text{ mm}$ Column temp: $40\text{ }^{\circ}\text{C}$ Autosampler temp: $7\text{ }^{\circ}\text{C}$ UV = $280\text{ nm}$ FR = $0.200\text{ mL/min}$ $10\text{ }\mu\text{g}$ injection	LC instrument: Waters Acquity UPLC MS: Thermo LTQ-Orbitrap Velos Column: Agilent Zorbax C18, $4.6\text{ mm ID} \times 150\text{ mm}$ Column temp: $50\text{ }^{\circ}\text{C}$ Autosampler temp: $8\text{ }^{\circ}\text{C}$ UV = $214\text{ nm}$ FR = $0.100\text{ mL/min}$ $20\text{ }\mu\text{g}$ injection	LC instrument: Acquity UPLC M-Class MS: Waters Xevo G2 XS QTOF Column: Waters M-class HSS T3, $0.3\text{ mm} \times 150\text{ mm}$ , $1.8\text{ }\mu\text{m}$ Column temp: $60\text{ }^{\circ}\text{C}$ Autosampler temp: $4\text{ }^{\circ}\text{C}$ to $6\text{ }^{\circ}\text{C}$ FR = $0.005\text{ mL/min}$ $0.12\text{ }\mu\text{g}$ injection																																																		
MS/MS: top 10 method, Orbitrap for parent (resolution = $60,000$ ), ion trap for fragments	MS/MS: top 4 method, Orbitrap for parent (resolution = $60,000$ ), ion trap for fragments	MS/MS: top 10 method, (resolution = $30,000$ to $60,000$ )																																																		
Mobile phase A was $0.1\%$ FA in water and Mobile phase B was $0.1\%$ FA in acetonitrile.	Mobile phase A was $0.02\%$ TFA in water and Mobile phase B was of $0.02\%$ TFA in acetonitrile.	Mobile phase A was $0.1\%$ FA in water and Mobile phase B was $0.1\%$ FA in acetonitrile																																																		
The total gradient program was $160\text{ min}$ : <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Time (min)</th> <th>% B</th> </tr> </thead> <tbody> <tr><td>0</td><td>3</td></tr> <tr><td>10</td><td>3</td></tr> <tr><td>120</td><td>45</td></tr> <tr><td>125</td><td>97</td></tr> <tr><td>135</td><td>97</td></tr> <tr><td>136</td><td>3</td></tr> <tr><td>160</td><td>3</td></tr> </tbody> </table>	Time (min)	% B	0	3	10	3	120	45	125	97	135	97	136	3	160	3	The total gradient program was $210\text{ min}$ : <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Time (min)</th> <th>% B</th> </tr> </thead> <tbody> <tr><td>0</td><td>0.5</td></tr> <tr><td>5</td><td>0.5</td></tr> <tr><td>185</td><td>45</td></tr> <tr><td>186</td><td>100</td></tr> <tr><td>190</td><td>100</td></tr> <tr><td>191</td><td>0.5</td></tr> <tr><td>210</td><td>0.5</td></tr> </tbody> </table>	Time (min)	% B	0	0.5	5	0.5	185	45	186	100	190	100	191	0.5	210	0.5	The total gradient program was $100\text{ min}$ : <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Time (min)</th> <th>% B</th> </tr> </thead> <tbody> <tr><td>0</td><td>2</td></tr> <tr><td>3</td><td>2</td></tr> <tr><td>70</td><td>30</td></tr> <tr><td>75</td><td>40</td></tr> <tr><td>80</td><td>70</td></tr> <tr><td>82</td><td>85</td></tr> <tr><td>82.1</td><td>2</td></tr> <tr><td>100</td><td>2</td></tr> </tbody> </table>	Time (min)	% B	0	2	3	2	70	30	75	40	80	70	82	85	82.1	2	100	2
Time (min)	% B																																																			
0	3																																																			
10	3																																																			
120	45																																																			
125	97																																																			
135	97																																																			
136	3																																																			
160	3																																																			
Time (min)	% B																																																			
0	0.5																																																			
5	0.5																																																			
185	45																																																			
186	100																																																			
190	100																																																			
191	0.5																																																			
210	0.5																																																			
Time (min)	% B																																																			
0	2																																																			
3	2																																																			
70	30																																																			
75	40																																																			
80	70																																																			
82	85																																																			
82.1	2																																																			
100	2																																																			

*Continued on next page.*



**Table 3. (Continued). Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis and Database Searching Details**

<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>
<i>LC-MS Analysis</i>		
Peptides were detected online with LTQ-Orbitrap Elite mass spectrometer operated in positive electrospray ionization mode and calibrated externally. Source temperature was set at 250 °C, spray voltage at 3.8 kV and capillary temperature at 350 °C.	Peptides were detected online with LTQ Orbitrap Velos mass spectrometer operated in positive electrospray ionization mode using lock mass ion at <i>m/z</i> 391.2841. Source temperature was set at 200 °C, spray voltage at 3.8 kV, capillary temperature was set to 250 °C.	Peptides were detected online with the Waters Xevo G2 XS QTOF mass spectrometer operated in positive electrospray ionization mode using Glu-Fib lock mass ion at <i>z</i> = +2, <i>m/z</i> = 785.8427 Da. Capillary voltage was set at 3.25 kV, with the sampling cone at 40 V, source offset at 80 V. Cone gas flow was set at 50 L/hr and source temperature at 80 °C. Desolvation temperature was 220 °C and gas flow rate 600 L/hr.
<i>MS Conditions</i>		
MS1 Conditions: scan range 110 to 2000 <i>m/z</i> and 60,000 resolution (at 400 <i>m/z</i> ).	MS1 Conditions: scan range 200 to 2000 <i>m/z</i> and 60,000 resolution (at 400 <i>m/z</i> ).	MS1 Conditions: scan range 300 to 2000 <i>m/z</i> and 22,000 resolution (at 400 <i>m/z</i> ); operated in W optics mode with normal dynamic range. The data acquisition rate was set to 0.500 s, with an 0.01 s interscan delay. Data were collected in continuum mode.
MS2 Conditions: low-resolution CID in LTQ using 2 <i>m/z</i> isolation window and 35% normalized collision energy.	MS2 Conditions: low-resolution CID in LTQ using 2 <i>m/z</i> isolation window and 35% normalized collision energy.	MS2 Conditions: peptides were isolated ± 1.500 Da and fragmented using a linear collision energy ramp from 60 eV to 80 eV. Scan time was set at 0.500 s with interscan time of 0.014 s.
Data Dependent Settings: 500 counts for MS2 trigger, top 10 precursors, dynamic exclusion with a repeat count of 2, repeat duration of 30 s and exclusion for 45 seconds. Unassigned charge states were rejected from triggering MS2.	Data Dependent Settings: 1000 counts for MS2 trigger, top 4 precursors, dynamic exclusion for 18 seconds. Unassigned charge states were rejected from triggering MS2.	Data Dependent Settings: MS2 was triggered for a Top 5 experiment by ions passing an intensity threshold of 5000. Real-time dynamic exclusion was set for 15.0 sec. Each cycle of experiments was followed by analysis of the lock-mass reference.
<i>Data Analysis, Database Searching</i>		
Data analyzed with Byonic (Protein Metrics, San Carlos, CA): <u>Standard search parameters:</u> MS mass tolerance: 10 ppm MS/MS mass tolerance: 0.4 Da Fixed mods: Cys carbamidomethylation (+57.021464 Da) Variable mods: Met oxidation (+15.994915 Da) Asn deamidation (+0.984016) succinimide intermediate (-17.02655 Da) protein C-terminal loss of Lys (-128.094963 Da) protein N-terminal pyroglutamination (-17.02655 Da) Lys glycation (+162.05282 Da) Tryptic map: enzyme "trypsin," up to 4 missed cleavages	Data analyzed with MassAnalyzer (195): <u>Standard search parameters:</u> MS mass tolerance: 5 ppm MS/MS mass tolerance: 0.5 Da Fixed mods: Cys carbamidomethylation (+57.021464 Da) Variable mods: Met, Trp +15.9949 Met, Trp, Tyr +31.9898 Asn +0.9840 Asn, Gln -17.0265 Lys +162.0528 protein N-terminal pyroglutamination (-17.02655 Da) Asp, Ser, Thr, Glu -18.01 Lys, Pro +15.9950 Tryptic map: enzyme "trypsin".	Data analyzed with Proteome Discoverer (Thermo Scientific, San Jose, CA): <u>Standard search parameters:</u> MS mass tolerance: 10 ppm MS/MS mass tolerance: 0.4 Da Fixed mods: Cys carbamidomethylation (+57.021464 Da) Variable mods: Met, Trp oxidation (+15.994915 Da) Asn deamidation (+0.984016) protein C-terminal loss of Lys (-128.094963 Da) protein N-terminal pyroglutamination (-17.02655 Da) Tryptic map: enzyme "trypsin", up to 2 missed cleavages

*Continued on next page.*

**Table 3. (Continued). Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis and Database Searching Details**

Lab 1	Lab 2	Lab 3
<i>Quantitative Analysis</i>		
For peptides containing modified residues and their unmodified counterparts, XICs were created and AUC calculated using the Xcalibur 2.2 (Thermo Scientific, San Jose, CA) Qual Browser. Peptide mass tolerance was set at 10 ppm and the monoisotopic peak for each detected charge state was used. Relative abundances were determined by calculating the percent AUC of peptides containing a certain modified residue relative to all peptides (modified and unmodified) containing that residue.	MassAnalyzer automatically calculates peak areas, which include all the isotope peaks from all major (> 1/6 of the base peak) peptide forms and charge states. MassAnalyzer also reports the percentage of each modification, so generally there is no manual quantitation needed.  To manually verify the quantitation by MassAnalyzer, only the monoisotopic peak for each detected charge state was used. Relative abundances were determined by calculating the percent AUC of peptides containing a certain modified residue relative to all peptides (modified and unmodified) containing that residue.	Quantitative analysis was performed manually. XICs of the monoisotopic peaks for +2, +3, or +4 charge states of the peptides of interest were generated using theoretical molecular weight to two decimal places and a mass tolerance of $\pm 0.1$ Da. The charge state with highest intensity for the peptide was used for quantitation. Mass accuracy and correct charge state was confirmed for each peptide. Reconstructed ion chromatograms of the unmodified and modified peptides were integrated using peak areas to calculate relative modified percent. Note: In some cases the primary ion was +1 charge state.
<small>AUC = area under the curve, CID = collision-induced dissociation, FR = flow rate, FA = formic acid, ID = interior diameter, LC chromatography, LTQ = linear trap quadrupole, MS = mass spectrometry, MS/MS = tandem mass spectrometry, QTOF = quadrupole-of-flight, TFA = trifluoroacetic acid, UHPLC = ultra-high-pressure liquid chromatography, XIC = extracted ion chromatogram</small>		

PTMs identified in the NISTmAb (control sample) are presented in the “PTMs identified in NIST reference mAb” section, with direct comparison of quantified levels from the three laboratories. Peptides were identified based on precursor mass and MS/MS, and quantified through XIC peak area measurement, as described in Table 3.

### Quantification Strategies

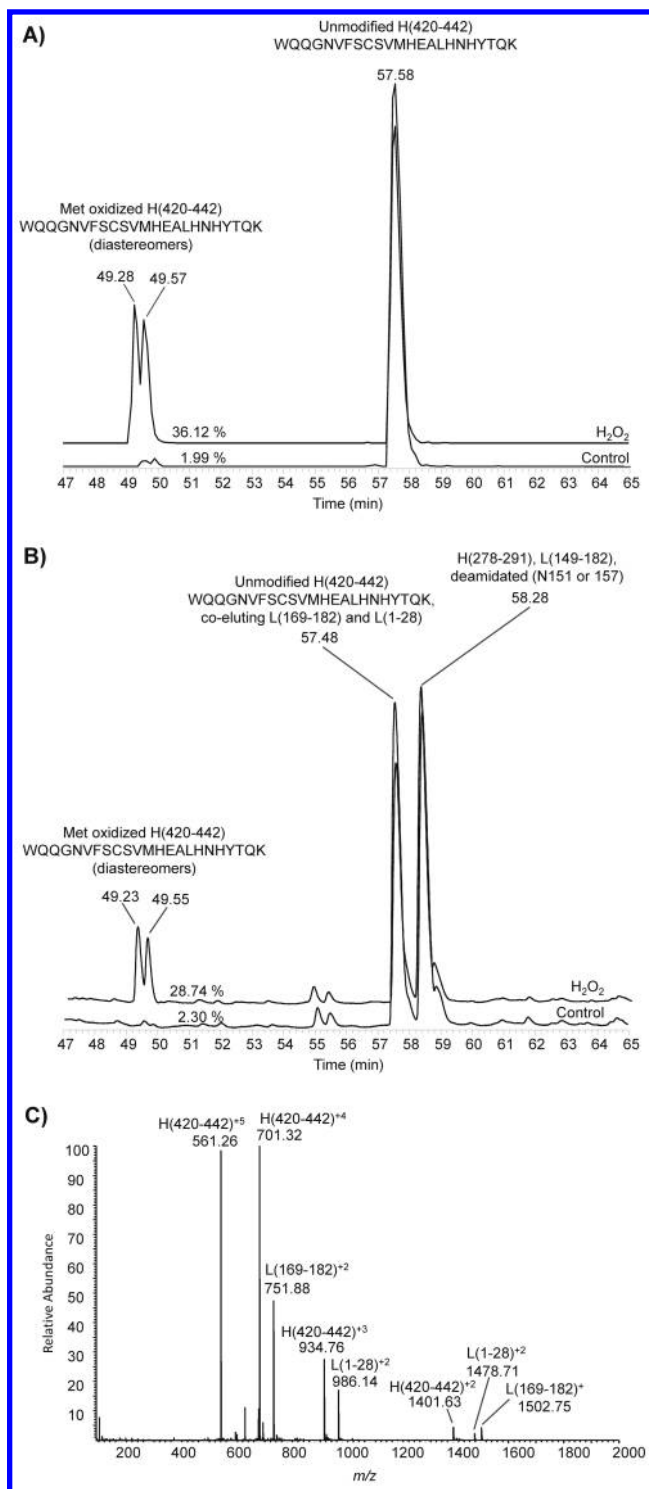
Because this study reports all quantitative PTM levels as percentages, it is important to understand the basis for percentage calculation and clarify how these quantifications are performed in each laboratory. As a typical practice, peptide mapping in the biopharmaceutical industry during characterization stages is performed with both UV and MS-based detection. Quantification of individual peptides by UV analysis can be complicated by the presence of co-eluting peptides. Figure 3 illustrates the differences in quantitative outcomes that are often observed between UV and XIC quantitation methods. Upon treatment with H<sub>2</sub>O<sub>2</sub> for 4 hr at 37 °C, M431, located in NISTmAb peptide H(420-442), was found to have high levels of oxidation as compared to the control (see section below). As shown in Figure 3, the relative levels of oxidation were quantified by calculating the area under the curve (AUC) for peaks generated from summed charge state monoisotopic parent ion XICs (Figure 3A) as well as their UV signal (Figure 3B). Relative abundance of the modification of interest is calculated as:

$$\% \text{ Relative abundance} = \left[ \frac{\text{AUC of modified peptide}}{\text{AUC of total peptides}} \right] \times 100$$

where AUC of total peptides = all modified forms + unmodified form.

Quantitation by XIC indicated that approximately 36% of H(420-442) underwent Met oxidation when subjected to oxidative stress, whereas the control sample only contained low levels of oxidation (approximately 1.99%), which were near the limit of detection. The relative abundance of Met oxidation in the H<sub>2</sub>O<sub>2</sub>-treated peptide when UV peaks were quantified was considerably lower (approximately 29%) than the value obtained from XIC quantitation. Examination of the MS1 spectra collected throughout the elution of the unmodified peak (57 to 58 min) shows co-elution of additional peptides (Figure 3C). The co-eluting peptides contribute to UV absorption and thereby gave rise to an overestimation of unmodified H(420-442) abundance and, consequently, an underestimation of the oxidized form. This example highlights the need to be aware of confounding factors when making such measurements and also demonstrates the additional selectivity conferred by MS versus LC alone.

Although all three laboratories only utilized XICs from MS signals to report relative levels of PTMs in our interlaboratory study, there are still many factors to consider when interpreting these results. For example, because the XIC is reconstructed from MS1 data, duty cycle or scan speed of the instrument can significantly affect the quantification. Proper integration of a peak requires a sufficient number of data points to fit a truly representative shape. Most modern instrumentation allows data-dependent acquisition of MS/MS spectra for increased protein sequence coverage (e.g., dynamic exclusion of abundant ions after fragmentation to sample peptides of lower abundance). However, although the use of these settings may increase peptide identification, they typically provide fewer MS1 scans for quantification or even result in under-sampling of the peak. To circumvent this issue, methods have been used that include only MS1 scans when quantitation is anticipated, with MS/MS performed for identification purposes in a separate run (210). Table 2 shows that the individual laboratories in this study chose a variety of data-dependent settings (i.e., Top-N method) for analysis of the NISTmAb sample. This parameter, as well as additional differences in LC methods (e.g., different gradient, mobile phase, column, temperature) and MS instrument types (Orbitrap vs. QTOF), may account for substantial differences in the overall PTMs identified and measured.



*Figure 3. Quantitation of Met oxidized peptide H(420-442) (WQQGNVFSVSMHEALHNHYTQK). Tryptic digests of a NISTmAb sample exposed to 0.04% H<sub>2</sub>O<sub>2</sub> for 4 hr at 37 °C and an untreated sample were analyzed by liquid chromatography-UV-tandem mass spectrometry (LC-UV-MS/MS).*

*(A) Quantitation of unmodified and Met oxidized peptides by extracted ion chromatogram (XIC); (B) Quantitation of unmodified and Met oxidized peptides by UV; (C) MS spectra summed over elution time (57 to 58 min) of unmodified peptide peak indicating co-eluting peptides. Note in (A) and (B) that oxidation of Met forms a chiral center at the sulfur atom and therefore elutes as a “double peak,” whereas the parent peptide elutes as a single peak. Data provided by Lab 1.*

The method by which the XIC is prepared is also important to consider. One must keep in mind that:

1. Most peptide intensity signals are split because they tend to ionize as a number of charge states.
2. The charge state distribution of an analyte is dependent on a number of factors, including its concentration (211).
3. Ion suppression by a co-eluting peptide may affect one member of the modified/unmodified peptide pair of interest but not the other.
4. A co-eluting peptide of similar *m/z* may be included in the XIC, depending on the width of the mass tolerance window.
5. The amino acid residue subject to the modification of interest may be present in more than one peptide form due to missed cleavages, partial cleavage, and other modifications.
6. A peptide may have additional types of modifications besides the modification of interest.
7. The isotopic peaks, which can be easily resolved in QTOF or Orbitrap instruments, may or may not be considered in the final XIC, depending on software settings.

Consequently, there are multiple properties that demand attention when generating an XIC for quantitation, including—but not limited to:

1. Charge state (use of all detected charge states vs. most abundant charge state).
2. Isotope distribution (use of monoisotopic peak vs. most abundant isotopic peak or entire isotopic envelope).
3. Incomplete digestion (number of missed cleavages/partial cleavages to consider for quantification).
4. Additional modifications (derive total AUC from the peptide with the modification of interest and unmodified peptide vs. including additional amino acid modifications of the same peptide in the value for total peptide abundance).
5. Intensity above a given threshold (use of the limit of detection for the threshold or greater value to increase confidence).

6. Mass tolerance (too large a tolerance may include co-eluting peptides with similar  $m/z$ , whereas too narrow a tolerance may exclude some information from the peptide of interest; this setting will vary between instruments, depending on mass accuracy).

Many of these parameters are fixed within the software coding and thus not freely manipulated by the user; it is important, however, to understand the methods used by the algorithm to calculate XICs in order to have a full understanding the resulting data.

One example illustrating the need to carefully consider the methods to be used for quantitation is presented by Lab 3 in Figure 4. The heavy chain of the NISTmAb terminates with the tryptic peptide SLSLSPGK, although the data from all three laboratories show that the majority of the heavy chain has undergone Lys loss at the C-terminus. Thus, the *des*-Lys peptide SLSLSPG is highly abundant. The presence of the Lys on the unmodified peptide renders it more readily ionized to the higher charge state ( $z = +2$ ) than the clipped peptide, which is instead more abundant at  $z = +1$ . Table 4 presents relative abundances of the *des*-Lys peptide as calculated by using several XIC methods that differ only in the charge states included in the peak integration. The values vary from a 99.54% abundance of SLSLSPG relative to SLSLSPGK when only the +1 charge state is considered, to as low as 61.63% relative abundance of the *des*-Lys form when only the +2 charge state is included. Differences in charge state distributions between corresponding modified and unmodified peptides are only one factor that can have a great impact on data analysis results. This example highlights the importance of understanding the data processing methods one uses, however, and the extent to which the methods themselves influence the results.

In the current study, Lab 1 used Bionic software (Protein Metrics Inc., San Carlos, CA) to first identify the PTMs present on tryptic peptides of the NISTmAb with up to four missed cleavages. Manual quantification was then performed via XICs in Xcalibur ( $\pm 10$  ppm of theoretical monoisotopic value) that included all detected charge states of modified peptides and their unmodified counterparts. Lab 2 utilized MassAnalyzer (195) for the peptide identification, and the automated quantification includes all the isotope peaks from all major ( $> 1/6$  of the base peak intensity) peptide forms and charge states, and further verified using XICs. Lab 3's quantitation method relied upon manual XIC extraction of monoisotopic peaks for the +2, +3, or +4 charge states of the peptides of interest, using theoretical molecular weight to two decimal places and a mass tolerance of  $\pm 0.1$  Da. The charge state with highest intensity for the peptide was used for quantitation. Mass accuracy and correct charge state was confirmed for each peptide. Reconstructed ion chromatograms of the unmodified and modified peptides were integrated using peak areas to calculate relative modified percent.

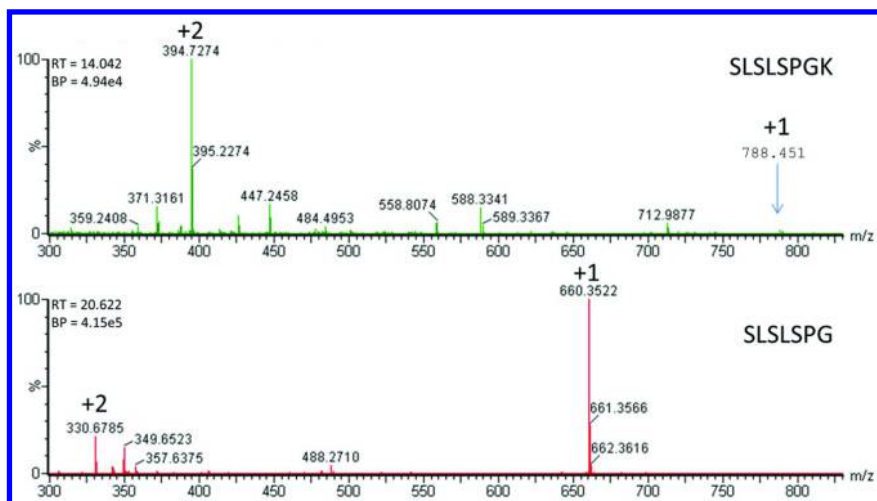


Figure 4. Presence of C-terminal Lys changes the charge state distribution of the terminal heavy chain peptide. Mass spectra collected by Lab 3 during peptide mapping shows ions corresponding to the heavy chain C-terminal peptide SLSLSPGK (top) and the clipped des-Lys form (bottom). The presence of the readily ionizable Lys residue on the unmodified peptide results in a highly abundant  $z = +2$  charge state, whereas the  $z = +1$  ion is more abundant for the modified peptide.

**Table 4. Relative Abundance Calculations of des-Lysine Peptide<sup>a</sup>**

Charge State Included	Relative Abundance of des-Lysine (%)
$z = +1$	99.54
$z = +2$	61.63
$z = \text{most abundant per species}$	88.12
$z = +1, +2$	89.72

<sup>a</sup> Lab 3 generated area under the curve (AUC) values for the extracted ion chromatograms (XICs) of the C-terminal peptide of the heavy chain with and without Lys. Relative abundance values were calculated for the des-Lys peptide using methods that differ according to which charge state was considered.

Values reported during quantification exercises using LC-MS will inherently fluctuate due to the variety of experimental and data analysis conditions that may affect the final value. This fact is clearly demonstrated in the NISTmAb study discussed below. When conducted under consistent and optimized operating conditions, however, these fluctuations can be kept to a minimum and very reproducible results can be attained (212, 213). Therefore, although peptide mapping may have its limitations, as discussed above, it remains the staple

technique for observing the widest range of PTMs while simultaneously providing quantitative results. The main items to keep in mind when interpreting these data are that:

1. Quantification with peptide mapping is in fact relative.
2. There exists the potential for additional modifications to be introduced during the sample preparation as the sample is exposed to the various buffers, temperatures, and other conditions used throughout the process.
3. Sample analysis itself may introduce artificial modifications because peptide integrity may be affected by source settings, such as temperature or voltage.
4. Data acquisition and/or processing protocols may introduce variability in identifications.

### **PTMs Identified in the NIST Reference mAb**

In this section, the PTMs identified in the NIST reference mAb (control sample) are presented, along with the relative abundances calculated by each laboratory. The PTM types covered in this section include: glycosylation, C-terminal Lys loss, N-terminal pyro-Glu formation, deamidation, oxidation, and glycation. Some PTM types discussed in the Most Common PTMs in mAbs Section are not reported here either because they are not routinely tracked by each laboratory or simply because they were not detected or identified in the NISTmAb. This section gives an overview of the PTM distributions throughout the NISTmAb molecule along with specific discussions pertaining to items to consider that are specific to the modification type. It should be noted, as discussed above, that these are relative quantification values rather than absolute measures of PTM content because the reported number may be influenced by digestion conditions and/or the quantification methods used to obtain the values, as highlighted below.

#### *Glycosylation*

Table 5 is a list of glycans identified in the NISTmAb control sample by the three laboratories. Each laboratory utilized compositional matching (theoretical  $m/z$  to experimental  $m/z$ ) as a means to identify the glycan species occupying the N300 N-glycosylation site. Although each laboratory readily identified this residue as being predominantly glycosylated (> 98%), the identities of the glycans reported at this site varied between laboratories, particularly for the low abundance species.

Quantification of glycosylation is challenging with peptide mapping because of the existence of numerous glycan forms and the difficulty in achieving good chromatographic separation of each species. In a typical RP method used for peptide mapping, the elution behavior of glycopeptides is dominated by the hydrophobic peptide rather than the unique hydrophilic properties of each conjugated N-glycan of differing composition and/or structure (214, 215). As



a result, glycopeptides comprising the same amino acid sequence elute within a narrow window, with a great deal of co-elution. The number of glycan species that can be identified is typically affected by ion suppression, such that low-abundant species may go undetected or may not generate enough signal for confident quantification. Due to the variations in chromatographic methods used, the population of low-abundant glycopeptides identified and quantified by each laboratory will manifest greater differences than the high-abundant species.

**Table 5. Putative Glycans Identified in the NIST Reference mAb (Control) Using Trypsin Digest and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) on Heavy Chain N300**

<i>Glycan</i>	<i>Lab 1 (%)</i>	<i>Lab 2 (%)</i>	<i>Lab 3<sup>b</sup> (%)</i>
FA2G1 (H4N4F1)	30.04	39.22	33.00 ( $\pm 0.70$ )
FA2 (H3N4F1)	35.96	33.71	39.04 ( $\pm 1.55$ )
FA2G2 (H5N4F1)	7.35	12.36	6.03 ( $\pm 1.15$ )
FA1 (H3N3F1)	14.96	3.40	10.40 ( $\pm 0.53$ )
FA2G2Ga1 (H6N4F1)	1.01	3.16	1.29 ( $\pm 0.48$ )
FA1G1 (H4N3F1)	6.08	2.42	4.81 ( $\pm 0.19$ )
FA2G2Ga2 (H7N4F1)	0.92	1.58	0.63 ( $\pm 0.30$ )
FA1G1Gc1 (H4N3F1Sg1)	--	1.11	--
FA3G1 (H4N5F1)	--	1.09	0.97 ( $\pm 0.06$ )
M5 (H5N2)	0.80	1.02	0.87 ( $\pm 0.31$ )
FA1G1Ga1 (H5N3F1)	0.98	0.96	1.25 ( $\pm 0.18$ )
A1 (H3N3)	--	0.75	0.56 ( $\pm 0.32$ )
Unglycosylated	1.90	0.72	1.17 ( $\pm 0.12$ )

<sup>a</sup> Oxford nomenclature followed by monosaccharide composition (in parenthesis), where H = Hexose, N = HexNAc, F = Fucose, Sg = N-glycolylneuraminic acid. <sup>b</sup> Value in parenthesis represents  $\pm 1$  standard deviation for three separate digestions and analyses.

Although glycosylation is not chemically affected by digestion methods, as many other PTMs can be, glycan moieties are relatively labile and are prone to monosaccharide fragmentation during ionization (216, 217). These in-source degradation products will have  $m/z$  values that correspond to smaller glycopeptide compositions. Many of the chemical changes induced by PTMs will cause a shift in retention time compared to the unmodified peptide. Thus, degradations that are artificially derived from in-source events can be distinguished from pre-existing PTMs because the former will have the same retention time as the unmodified peptide. Because glycopeptides with differing glycan compositions do not easily resolve chromatographically, however, glycan compositions resulting from in-source loss of saccharide residues cannot generally be distinguished

from their native forms using comparative retention times. This will certainly lead to an overestimation of some glycoforms, particularly those of smaller composition (and a corresponding underestimation of their original glycoform) (218). For example, the FA1G1 composition will ostensibly result from in-source degradation of FA2G2, and consequently in the underestimation of FA2G2. After reviewing each laboratory's tune method, we found that Lab 1 operated the capillary temperature at 350 °C according to the HESI setting recommendations, whereas Lab 2 customized the capillary temperature to 250 °C to avoid in-source fragmentation. Lab 3, in contrast, was using a capillary column with 220 °C but a much lower flow rate, which may increase the chance of in-source fragmentation. This may be the dominating factor contributing to the differences between the three laboratories (Table 5): larger glycans (e.g., FA2G2, FA2G1, FA2G2Ga1) are in higher abundance in Lab 2 than in Labs 1 and 3, whereas Labs 1 and 3 detected higher levels of smaller glycans (FA2, FA1, FA1G1). Peptide mapping can be a useful tool to ensure appropriate localization of collective glycan site occupancy and the relative level to which that site is unoccupied; however, detailed glycoanalysis (both for qualitative identification and quantitative glycan analysis) has traditionally been performed on released glycans, as discussed in detail in the Glycosylation chapter/Volume 2, Chapter 4. It should be noted, however, that all glycans identified in Table 5 also were identified using the targeted glycan methods described in the Glycosylation chapter/Volume 2, Chapter 4.

### *C-Terminal K and N-Terminal Pyro-Glu*

The NISTmAb DNA sequence encodes a Lys residue at the carboxy terminus of the heavy chain. Due to the prevalence of C-terminal Lys clipping that occurs during the mAb manufacturing process (47), it is routine in mAb characterization platforms to assess the level to which this has occurred. Peptide mapping of the NISTmAb using trypsin identified the presence of both a fully intact peptide (designated as +K) as well as its counterpart with C-terminal Lys truncation (*des*-Lys). The data reported from all three laboratories were in agreement that the *des*-Lys species dominated in relative abundance (> 86%) over the unmodified C-terminal heavy chain peptide as observed through peak integrations in XICs (Table 6).

The C-terminal Lys loss results were almost the same between Lab 1 and Lab 3 (89.85% vs. 89.73%). Lab 2, however, detected *des*-Lys at a relative abundance of approximately 87%, which represents a minor variation (interlaboratory coefficient of variation [CV] = 1.9%). In this case, the interlaboratory variation was only slightly higher than the intra-laboratory variation seen by Lab 3 when all digestion factors were well controlled (intra-laboratory CV = 0.75%), indicating little to no method-related variation. Although the authors considered this to be well within assay reproducibility, analyses of the various data sets for this peptide were placed under additional scrutiny. Because Lab 1 and Lab 2 both included all charge states in their quantitation calculations, this minor discrepancy is not likely due to the differences in charge state distributions as discussed

earlier. Another possible explanation explored was the potential for differential ion suppression in each of the laboratories. As depicted in Figure 4, Lab 3 achieved chromatographic resolution of both the +K and *des*-Lys forms from any interfering peptides, whereas data from Labs 1 and 2 show the +K form co-eluting with other major peaks (data not shown). Considering that Labs 1 and 3 produced similar quantitative values, ion suppression is not likely to be responsible for this minor variation. This minor variation may instead have been due to differences in background noise levels or slightly different elution profiles created by different chromatographic methods; ultimately, however, all three laboratories' results were deemed to be comparable within expected reproducibility.

**Table 6. C-Terminal Lysine Loss and N-Terminal Pyro-Glu Identified in Control NISTmAb Using Trypsin Digest and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)**

<i>Post-Translational Modification</i>	<i>Position</i>	<i>Lab 1 (%)</i>	<i>Lab2 (%)</i>	<i>Lab 3<sup>a</sup> (%)</i>
C-terminal Lys Loss ( <i>des</i> -Lys)	HC K450	89.85	86.89	89.73 (±0.67)
N-terminal Pyroglutamate	HC Q1	99.65	99.64	> 99

<sup>a</sup> Value in parenthesis represents ± 1 standard deviation for three separate digestions and analyses.

The sequence of the NISTmAb contains an N-terminal Gln residue on the heavy chain. Conversion to the cyclized pyro-Glu was found in high abundance (> 99%) by all three laboratories in the NISTmAb N-terminal peptide, with only trace levels of the unmodified (QVTLR) Gln residue detected (Table 6). Lab 1 and Lab 2 measured the same level of N-terminal pyro-Glu, whereas Lab 3 was unable to quantify the levels because the intensity of unmodified peptide signal was too low. Gln cyclization has been reported to increase with longer digestion times, and therefore should be optimized to minimize artifactual conversion. Digestion time did not seem to influence the results in the current experiment as all laboratories obtained similar results. In addition, the N-terminal pyro-Glu levels measured by the three laboratories is consistent with both the intact and middle-down analyses as discussed in the Primary Structure chapter/Volume 2, Chapter 1, indicating that the peptide mapping methods used provided minimal method-induced conversion of the N-terminus to pyro-Glu.

### *Deamidation*

As expected for any therapeutic protein, deamidation of a number of Asn residues were detected in the digested NISTmAb sample. Several Asn residues within the conserved regions of human IgGs have been demonstrated in the literature as being more prone to deamidation than others (79, 219). In particular, the Asn-Gly motif deamidates most readily, which is thought

to be due in part to minimal steric hindrance during formation of the initial cyclized succinimide intermediate. Commonly, susceptible asparagines are found within tryptic peptides corresponding to NISTmAb H(305-320) and H(374-395) (VVSVLTVLHQDWLNGK and GFYPSDIAVEWESNGQPENNYK). Indeed, these peptides were identified as unmodified, deamidated (Asp and/or iso-Asp), and succinimide (Asu) forms in tryptic digests of the NISTmAb. As an example, Lab 1 data observed between 3% and 5% modification of one or more Asn residues in these peptides, whereas modification of other peptides containing Asn residues remained below 1% (Table 7).

A comparison between Labs 1 and 2 shows that Lab 1 reported, on average, higher levels of deamidation. Evaluation of the digestion conditions shows Lab 1 used a slightly elevated temperature (37 °C vs. room temperature) during denaturation and reduction, double the time for alkylation, and higher temperature for digestion (50 °C vs. 37 °C). These conditions likely contributed, to some extent, to the slightly higher levels of deamidation. However, including Lab 3 (which performed digestion at an elevated pH = 8.0 that may cause increased deamidation) in the comparison increases the variability observed between laboratories, with no clear trend found to be associated with deamidation site, sequence, or specific method conditions. Despite this, Lab 3 performed triplicate analyses of the NISTmAb and was able to obtain very reproducible results. Collectively, this highlights the difficulty in monitoring deamidation using reconstructed ion chromatograms due to issues of co-elution, susceptibility to digest conditions, differences in software integration protocols, and isotopic overlaps.

Deamidation is a PTM that has long been known to be susceptible to digestion conditions. Elevated pH, elevated temperature, and increased duration of reduction/alkylation and digestion times have all been shown to increase levels of deamidation (220). The various factors also may be interrelated; for example the pH of ammonium bicarbonate has been reported to drift to more basic pH during extended digestions, resulting in additional deamidation (213). Alkaline pH also is detrimental to identifying inherent succinimide intermediates, as they are readily converted to Asp under these conditions. Various approaches for minimizing digest-induced deamidation have been developed. The use of low pH (6.0) and low temperature (4 °C) was implemented with long digestion times to reduce deamidation; however these digestions may require multiple days to yield optimal coverage (221). The speed of digestion has been directly targeted by using elevated pressures and microwave irradiation, also with some success in minimizing artifactual degradation (222–224). An alternative route to speed digestion is through increasing enzyme activity through removal of potential inhibiting agents (e.g., guanidine) prior to digestions (213). Combination strategies with various parameters optimized are the ultimate pursuit for “high fidelity” digestion protocols resulting in peptides more reflective of the native mAb composition (212). New enzymes for peptide mapping also are being explored, including a secreted aspartic protease 9, which was shown to produce approximately 3.4 kDa fragments with reduced deamidation due to its amenability to low pH digest conditions (225).

**Table 7. Asparagine Deamidation Identified after Tryptic Digestion of NISTmAb Control Sample<sup>a</sup>**

			<i>Approximate Relative Abundance of Deamidation (%)</i>		
			<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3<sup>b</sup></i>
H	N78	Q79	0.40	0.13	3.30 (±0.10)
H	N86	M87	0.11	---	0.27 (±0.06)
H	N162	S163	---	0.42	---
H	N279 or N289	W280 or A290	0.31	---	2.50 (±0.10)
H	N318	G319	5.04	0.18	4.90 (±0.17)
H	N328	K329	0.79	---	---
H	N387	G388	2.62 <sup>c</sup>	2.25	0.67 (±0.06) <sup>c</sup>
H	N392	N393		2.99	
L	N136 (or N137, Lab 1)	N137 (or F138, Lab 1)	0.32	0.23	0.93 (±0.06)
L	N151	A152	0.73 <sup>c</sup>	---	---
L	N157	S158		0.15	

<sup>a</sup> It should be noted that Lab 1 and Lab 3 added Asp, isoaspartic acid (iso-Asp), and succinimide (Asu) together, whereas Lab 2's report does not include Asu in the calculation. Residues neighboring the deamidated Asn at the C-terminal side are shown because they are known to influence the susceptibility of Asn to modification. <sup>b</sup> Value in parenthesis represents ± 1 standard deviation for three separate digestions and analysis. <sup>c</sup> Abundances distinguished between residues of same peptide for Lab 2 based on retention time as described in Figure 1, but calculated together for Labs 1 and 3.

In Lab 1 and Lab 3 results, peptides with Asn residues were identified in the NISTmAb, and the modification of Asn to Asp/iso-Asp or Asu was manually quantified by integrating XICs derived from the monoisotopic peaks of all detected charge states (Lab 1) or the most intense charge state (Lab 3). The relative abundance of peptides containing Asp/iso-Asp or Asu were added together as products of deamidation. Lab 2 quantified deamidation in much the same manner, but the percentages were derived from the software directly followed by manual verifications. The software employed by Lab 2 does not include Asu in the calculation, but rather as a separate category of ammonium loss as this is a common loss observed for many peptides regardless of sequence. In most cases, this did not result in significant differences in the relative quantitation calculated. In one example, however, Lab 2 manually incorporated Asu relative abundance into the calculation for heavy chain residue N318 modification and reported an increase in the percent deamidation at this site from 0.18% to 2%. Although this variation in calculation does not completely account for the relatively large disparity between laboratories at this site, it does exemplify the need to understand the underlying principles of software calculations when comparing datasets.

The use of XICs to monitor deamidation are further complicated by the fact that deamidation results in a mass shift of +0.98402. The monoisotopic mass of the modified peptide is therefore isobaric with the first  $^{13}\text{C}$ -containing isotope of the unmodified peptide. The use of XICs will therefore inherently overestimate the quantity of deamidation when the modified peptide co-elutes with the native peptide. Software algorithms have recently been developed to combat this issue using isotopic envelope modeling and sequence-specific predictive chromatography (226, 227).

When considering data such as those presented in Table 7, it is important to keep in mind the complications that may accompany the analysis of deamidated peptides. As mentioned earlier, chromatographic conditions may greatly affect which peaks are ultimately resolved and should be a critical aspect considered when comparing methods for deamidation assessment. Variations in digest conditions and quantification methodology also may affect the relative value of deamidation observed as already discussed. These factors together explain the apparent inconsistency between laboratories, and the reported values therefore represent the range of values expected in the absence of a tightly controlled, pre-defined protocol for digestion, instrumental analysis, and data analysis.

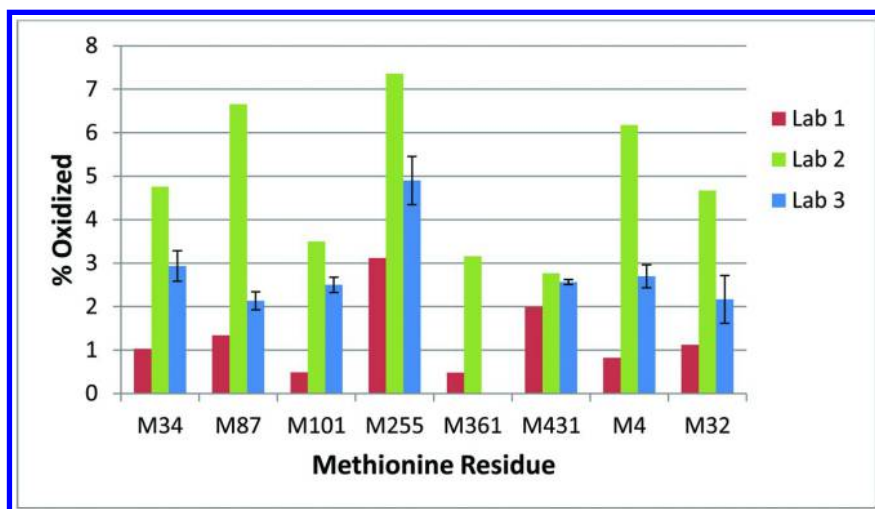
### *Oxidation*

Peptide mapping revealed that a number of Met residues were oxidized to relatively low levels in the native NISTmAb sample as demonstrated in Table 8 and Figure 5. All three laboratories identified nearly the same oxidized residues other than the M361, which was not observed in Lab 3. M255 was the most highly oxidized Met residue across all three laboratories, which is in agreement with the literature for other IgG1 $\kappa$  mAbs (121–123).

**Table 8. Identification and Quantification of NISTmAb Methionine Oxidation After Tryptic Digestion of Control Sample**

Chain	Residue and Position	Lab 1	Lab 2	Lab 3 <sup>a</sup>
H	M34	1.03	4.76	2.93 (±0.35)
H	M87	1.34	6.66	2.13 (±0.21)
H	M101	0.49	3.5	2.5 (±0.17)
H	M255	3.12	7.36	4.9 (±0.56)
H	M361	0.48	3.16	---
H	M431	1.99	2.77	2.57 (±0.06)
L	M4	0.82	6.18	2.7 (±0.26)
L	M32	1.12	4.67	2.17 (±0.55)

<sup>a</sup> Value in parenthesis represents ± 1 standard deviation for three separate digestions and analyses.



*Figure 5. Relative abundance as a function of Met residue observed in the NISTmAb. Error bars represent ± 1 standard deviation for three separate digestions and analysis. (see color insert)*

Met residues throughout the heavy and light chains showed percent relative abundance values that differed between the three laboratories; however, the same general trend for the percent oxidized (Lab 2 > Lab 3 > Lab 1) was observed for all residues other than M361, which was not identified by Lab 3. The trend in the interlaboratory data suggests method-related artifacts as a cause of interlaboratory variability. A recent study reported a correlation between

increased digest-induced oxidation and the presence of higher metal content originating from buffer components used during digestion (228). In general, phosphate buffer was found to contain higher levels of residual iron than Tris, and the addition of EDTA to phosphate-based digestions was shown to reduce artifactual oxidation (228). In the current experiment, each of the laboratories performed digestion using EDTA during the denaturation step, so this explanation does not seem to account for differences observed. Therefore an alternative explanation seems likely, although digest-induced oxidation cannot be ruled out completely in the current example. Oxidation may have been introduced by many factors such as elevated temperature, longer exposure to air, and oxidant or trace metal contamination in digestion reagents or analytical instrument surfaces (228).

Continued discussion of the interlaboratory differences in reported oxidation levels revealed that Lab 2 also observed higher levels of in-source oxidation. Fortunately, in-source oxidation artifacts always “co-elute” with native peaks because they are generated after LC separation, whereas real oxidation peaks generally elute earlier when using RP chromatography. Though most software tools are not programmed to distinguish true oxidation peaks from in-source oxidations, this rule was easily applied by manually authenticating the values reported for the oxidation products listed in Table 8. For example, the Lab 2 original software-indicated oxidation level for the peptide containing M255 was approximately 10%. It was thought that in-source oxidation contributed to the higher values. Indeed, when manual XIC analysis was performed on the same data set and in-source oxidation removed from the calculation (in-source oxidation was assumed for oxidized peptides having the same retention time as the unmodified peptide), the value was reduced to 7.36% as reported in Table 8.

Collectively, neither the potential difference in initial metal content nor in-source oxidation seem to reliably account for variations between the individual laboratories. However, this step-by-step evaluation of various analysis parameters highlights the common troubleshooting that one might undertake in an effort to minimize artifactual oxidation during digest optimization. Although time did not allow a full evaluation of all parameters here, additional method differences between laboratories that one might consider systematically testing for their effect on oxidation may include the use of microwave digestion (Lab 1) versus digestion without a microwave (Labs 2 and 3), additional raw materials (trypsin, buffer components, and other materials), and other factors that may have differed between the individual laboratories. Stable isotope labeling strategies also have been used to circumvent this rather tedious optimization process altogether. For example, Liu et al. fully reacted a recombinant mAb with  $^{18}\text{O}$ -enriched peroxide prior to any sample handling to prevent Met residues from oxidizing during any following experimental preparation or analytical processes. The isotopic label provided a mass tag to distinguish the forced oxidized residues from those that were present in the native mAb prior to sample handling because these Met residues were oxidized with  $^{16}\text{O}$ . Finally, the differential 2 Da shift between labeled and unlabeled peptides provided a means to accurately quantify the relative levels of initially modified  $^{16}\text{O}$  residues via XIC (229).

Although the specific cause for variability in reported Met oxidation levels between the laboratories could not be verified, what can be ascertained is that the



intra-laboratory variability (triplicate analysis in Lab 3) was much smaller than the interlaboratory variability. The CVs between laboratories ranged from 17 to 104%, whereas the Lab 3 reproducibility was considerably lower, with CVs of only 2 to 25%. These results indicate that even in the potential presence of method-related oxidation, good reproducibility can be attained when all digestion, method, and analysis protocols are consistently performed.

Low levels of Trp oxidation were also reported at various sites in the heavy chain (Table 9). As mentioned previously, oxidation of Trp may occur due to light exposure or the generation of free radicals, but in most cases, the levels are too low to raise concerns. Recently, one example reported a combination effect whereby light exposure was reported to generate a buffer-derived His photosensitizer that further degraded the mAb (230). Such a situation cannot be verified from the current data but is worth noting because the NISTmAb is also formulated in His.

**Table 9. Identification and Quantification of NISTmAb Tryptophan Oxidation After Tryptic Digestion of Control Sample**

<i>Chain</i>	<i>Residue and Position</i>	<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3<sup>a</sup></i>
H	W161 Double Oxidation	---	0.58	---
H	W280 Double Oxidation	---	0.49	0.13 ( $\pm 0.06$ )
H	W316 Double Oxidation	0.20	0.43	1.47 ( $\pm 0.12$ )
H	W384	---	0.75	---
H	W420	0.16	0.23	---

<sup>a</sup> Value in parenthesis represents  $\pm 1$  standard deviation for three separate digestions and analyses.

### *Glycation*

Although trypsin is the most commonly used enzyme for PTM analysis, it is certainly not the ideal enzyme when targeting glycation for quantitation. The specificity of trypsin itself, as opposed to experimental conditions, is the cause of its ineffectiveness as it has been shown to have a high propensity for missed cleavage at glycated lysines. The resulting inconsistency is significant enough that Lab 3 did not search for glycation in the current study. Despite this limitation, a data set search allowing for missed cleavage can yield identification of glycation sites. Glycation was observed at many Lys sites in the NISTmAb, but some obvious variances between Labs 1 and 2 are recorded (Table 10) and are likely due to the various limitations of trypsin digestion for this PTM, differences in the search parameters used and/or data quality of the MS/MS spectra. Lab 1 observed relatively consistent glycations across the molecule with an average of

approximately 0.5%. In contrast, Lab 2 observed more glycation sites (K66 and K225 of the heavy chain as well as K52, K102, K168, K182, and K206 of the light chain), but the average is higher ( $\approx 1\%$ ) and the range is wider ( $\approx 0\%$  to  $3\%$ ).

**Table 10. Identification and Quantification of NISTmAb Glycation After Tryptic Digestion of Control Sample**

Chain	Lys Position	Approx. Relative Abundance of Glycation (%)		
		Lab 1	Lab 2	Lab 3
H	K58	0.12	---	No report because Lab 3 does not routinely use trypsin peptide mapping for glycation quantification.
H	K66	---	0.90	
H	K77	0.68	1.06	
H	K136	0.11	1.28	
H	K225	---	0.38	
H	K249 or K251	0.23	1.13	
H	K291 or K293	1.87	0.55	
H	K329	0.16	---	
L	K38, K41, or K44	0.24	0.51	
L	K52	---	2.05	
L	K102	---	1.74	
L	K148	0.11	0.61	
L	K168	---	0.41	
L	K182	---	2.87	
L	K187	---	3.79	
L	K189	0.15	0.13	
L	K206	---	0.41	

Due to the common use of trypsin in analysis, it is worth discussing the data processing limitations associated with glycation quantitation in this analysis as well. As expected, the smallest peptide identified for each modification in Table 10 contains at least one missed cleavage because trypsin activity at the glycated Lys is blocked in the modified peptide. On the other hand, the majority of the unmodified peptide will exist as a properly cleaved (C-terminal to the Lys) species. As a result, the percent relative abundance can only be calculated by using the value for the missed cleavage glycated peptides relative to a properly cleaved unmodified counterpart. A subjective decision must therefore be made as to which of the unmodified peptides “halves” to use when calculating total peptide AUC.

For example:

$$\% \text{ Relative glycation} = \left[ \frac{DTSkNQVVLK}{DTSkNQVVLK + DTSK} \right] \times 100$$

or

$$\% \text{ Relative glycation} = \left[ \frac{DTSkNQVVLK}{DTSkNQVVLK + NQVVLK} \right] \times 100$$

where  $k$  = glycated Lys.

A scrutiny of the quantification in the current experiment reveals that the percentage calculation formulas used in Lab 1 and Lab 2 are different: in Lab 1, the total peak areas used for the percentage glycation calculation include the unmodified peptide with the potential glycation site (i.e., the first formula, above), whereas Lab 2's data processing software only uses the longer of the two properly cleaved unmodified peptides for total peak area calculation (i.e., the second formula above). The difference in length, chemical properties, particularly ionization efficiencies, and so forth between the unmodified peptide forms chosen for the total peptide calculation, as well as differences between the glycated and unmodified peptides themselves, could contribute significantly to the variances observed. Pursuit of a more comprehensive coverage of this PTM and/or more accurate quantitative values would be achieved through the use of an alternative enzyme such as Asp-N or Glu-C.

### *Overview of NISTmAb PTMs*

Figure 6 depicts a summation of the PTMs observed in the NISTmAb by the three laboratories involved in this project. Figure 6A lists the summation of PTMs identified in an effort to provide the most comprehensive list of potential modification sites in the NISTmAb. Figure 6B lists only those PTMs identified by all three laboratories. During the development of a therapeutic protein, years of analytical data are collected with a variety of orthogonal assays. Validation of the inherent presence of a given PTM would require a rigorous correlation of peptide mapping data with the totality of evidence, especially for those deemed to be critical quality attributes. For example, the combined level of deamidation and sialic acid content may be correlated to the quantity of acidic peak variants observed in a charge-based assay such as cIEF. It is expected that mass spectrometer settings, digest conditions, and other peptide mapping variables would also be optimized for the specific product during a method development protocol. Targeted multiple reaction monitoring (MRM) assays using internal standards also might be developed to more accurately quantify the presence of a given modification. It is expected that in the years to come, a more rigorous validation of the PTMs intrinsic to the NISTmAb may become a reality. In the absence of time to perform such a full validation, Figure 6 is presented as a visual representation of how variations in analytical protocols for peptide

mapping may result in slight differences in PTMs reported. Neither figure should be regarded as a comprehensive list of the PTMs present on the NISTmAb; Figure 6A represents a relatively comprehensive list of PTMs one might observe, however, and Figure 6B represents those for which the interlaboratory data suggests are most likely to be observed.

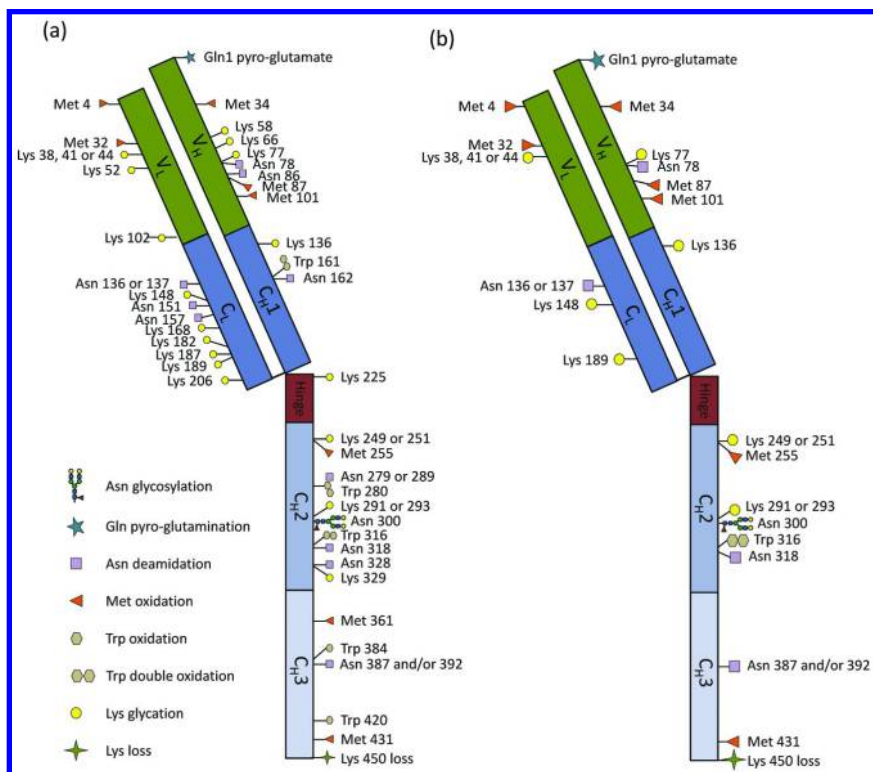


Figure 6. List of (A) all post-translational modifications (PTMs) identified and (B) those PTMs identified by all three laboratories in the control sample. The reader is referred to the Primary Structure chapter/Volume 2, Chapter 1 for the full amino acid sequence of the NISTmAb. (see color insert)

It is likely that future peptide mapping studies may identify the same and/or different PTMs in the NISTmAb. Various factors were demonstrated above that may increase or decrease the relative quantity of a particular PTM reported, or result in false positive or negative PTM reporting. The optimized strategy for digestion of a given therapeutic protein typically involves a significant amount of testing to evaluate sample preparation conditions, HPLC conditions, MS settings, MS/MS software identifications, and data processing for quantitative results. For example, digestion time-course studies could be performed to identify an optimal digestion time that is long enough to minimize missed cleavages while also being short enough to limit artifactual deamidation. Specific digestion conditions (e.g., low pH, metal chelating agents, different buffer compositions) also may be employed for digests targeted at analysis of a specific PTM. The longitudinal

availability of this material will therefore provide an optimum platform for additional round robin studies targeted at specific PTM measurement science questions. For example, a single data set may be circulated to compare data analysis, a pre-digested sample could be used to test instrumentation, and a full protocol from digest to analysis could be used to assess the entirety of peptide mapping.

Globally, Figure 6 and the previous discussion highlight that reported PTMs and their relative abundances are dependent on a variety of sample processing, analytical method, and data analysis factors. The interlaboratory data suggest that it is not uncommon for reported values to span up to a 5% relative abundance range in the absence of a completely defined method. When all appropriate controls are in place, however, PTMs can be reported with a high level of intra-laboratory precision.

### Importance of Forced Degradation Analysis

When characterizing therapeutic proteins, it is also important to consider the behavior of the material under any external stresses it may endure during production, shipment, storage, and ultimately, patient delivery. PTMs may be induced in the protein when it comes into contact with various surfaces (production equipment, packaging material, administration devices) and/or when it is exposed to sub-optimal temperatures, mechanical stresses, light exposure, pH levels, freeze/thaw cycles, and so forth.

Forced degradation (or stress testing as referred to in ICH Q1A(R2) (231)) refers to the intentional application of one or more stressors (in excess of conditions encountered for clinical material) to artificially induce changes in product attributes. Understanding changes that may result after exposure to certain stress factors is useful for elucidating critical quality attributes and potential degradation pathways. This information can in turn be used to design production, processing, formulation, and other strategies that will mitigate degradation as much as possible. Furthermore, forced degradation protocols can be used to produce materials containing relevant product-related impurities that are suitable for challenging analytical methods during qualification exercises. The aforementioned applications and relevance of forced degradation studies throughout the drug development lifecycle have recently been reviewed (232).

It is important to highlight the difference between forced degradation and stability testing. Forced degradation is a relatively broad term that encompasses stress applied to the drug substance or drug product *in excess* of that necessary for stability testing. For example, forced degradation may encompass incubation at elevated temperature and high pH simultaneously in an attempt to introduce high levels of degradation. Forced degradation investigations may need to be repeated at different points throughout the development following changes in concentration, formulation, and container closure or when stability-indicating capabilities of new or updated analytical methods requires assessment.

Stability testing is most often performed later in the development lifecycle after potential degradation pathways have been identified through forced degradation analysis, and tentative container/closure, formulation, validated

assays, and so forth have been established. Stability testing refers to assessing drug substance or drug product of the *intended formulation* according to regulatory guidelines such as those released by the International Conference on Harmonisation (ICH). Stability testing of drug substances and drug products must be performed on multiple production lots and is covered in general in ICH Q1A(R2). These guidance were further refined for biotechnology products in ICH Q5C (233).

ICH Q1A(R2) recommends real-time stability testing under the intended storage/temperature conditions for a minimum of 1 year, with a commitment from the manufacturer that real-time stability testing will continue to cover the proposed shelf-life of the material. ICH Q5C recognizes the complexity of biopharmaceutical stability monitoring and provides for a minimum of 6 months of data, or the proposed maximum storage period, and accounts for the fact that more complex analytical and potency assays may be required. ICH also recommends accelerated stability testing, which consists of an elevated temperature relative to real-time stability (all other conditions remaining consistent) and higher sampling frequency to monitor potential changes in the product under the given formulation conditions.

Stability and accelerated stability can in general be considered evaluation of a proposed drug substance and/or drug product formulation with the intent of establishing a suitable shelf-life for the given material. The specific application of forced degradation in this chapter is representative of a stage prior to formal stability testing. Forced degradation is used here on the NISTmAb to demonstrate its utility for molecule characterization and identification of potential degradation pathways. The exaggerated stress conditions are intended to produce larger than normal quantities of altered material with the intent of identifying all relevant potential degradation pathways and products, many of which are variants containing altered levels of one or more intrinsically present PTMs.

### **PTM Level Changes in Stressed NISTmAb Samples**

The quantitative assessment of the NISTmAb control samples discussed above readily demonstrated that quantification of PTMs can vary between laboratories due to inherent differences in sample preparation; however, consistent quantification through XICs can be achieved when all sample preparation remains consistent. For this reason, a single laboratory (Lab 3) conducted forced degradation analysis on the NISTmAb. The forced degradation study was conducted because this technique readily induces changes in PTM levels that can be used to elucidate potential degradation pathways, identify critical quality attributes for structure/function studies, and produce product-related impurities relevant for challenging the methods used for qualification exercises. To this end, NISTmAb samples were subjected to a series of forced degradation protocols known to introduce PTM changes (Table 11). The forced degradation conditions chosen for the current study, each of which are described in more detail below, are representative of those which would be performed during the early development stage of a new biopharmaceutical candidate.

**Table 11. NISTmAb Forced Degradation Conditions<sup>a</sup>**

<i>Sample Name</i>	<i>Condition</i>	<i>Potential Modifications</i>
Control	4 °C, 1 week (PBS, pH 7.4)	----
Heat (Thermal) Stress	40 °C, 1 week (PBS, pH 7.4)	Oxidation, deamidation, unfolding, aggregation, fragmentation
High pH (8.9)	37 °C, 1 week (phosphate, pH 8.9),	Deamidation (Asn→Asu, Asp, iso-Asp)
Low pH (3.5)	25 °C, 6 hours (ammonium acetate, pH 3.5)	Deamidation (Asn→Asp), fragmentation
Oxidative Stress	25 °C, 6 hours (0.1% H <sub>2</sub> O <sub>2</sub> )	Oxidation (Met)
Oxidation in the Presence of Fe(II)	25 °C, 2 weeks (30 μM ammonium iron sulfate)	Oxidation (Met, His, Cys, Trp)

<sup>a</sup> NISTmAb samples were subjected to the stress conditions shown. The most prevalent post-translational modifications (PTMs) expected to result from this exposure are listed. Although unfolding, aggregation, and fragmentation are expected results of stress conditions, these properties cannot be evaluated by peptide mapping and therefore were not included in this study. High pH, low pH, oxidative stress, and Fe oxidation stress conditions were performed at moderately elevated temperatures to accelerate the reactions. These elevated temperatures may in themselves slightly increase oxidation and deamidation levels across the molecule as compare to the control. Asu = succinimide, iso-Asp = isospartic acid, PBS = 14190-DPBS obtained from Life Technologies.

All forced degradation analyses were performed by dialyzing the original NISTmAb sample against the appropriate incubation buffer for the allotted time. Following incubation, the samples were dialyzed back into PBS and immediately processed for analysis as described above for Lab 3. PBS (pH 7.4) refers to 14190-DPBS obtained from Life Technologies (Grand Island, NY), which contains 2.666 mmol/L KCl, 1.471 mmol/L KH<sub>2</sub>PO<sub>4</sub>, 137.931 mmol/L NaCl, and 8.060 mmol/L Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O.

### *Thermal Stress*

Thermal stress evaluations are important in determining the appropriate storage and handling conditions for a given molecule. They also provide an understanding of the effects that higher than recommended storage temperatures may have on product stability. Although every effort is made to ensure appropriate storage, it is possible the product will encounter elevated temperatures during production, shipment, storage, and/or patient administration. Elevated temperatures can lead to conformational unfolding of the molecule, aggregation, and/or fragmentation (83, 232). The thermal unfolding of various IgG domains can be monitored using differential scanning calorimetry (DSC), differential

scanning fluorimetry (DSF), circular dichroism (CD), and Fourier-transform infrared spectroscopy (FTIR) (as discussed in more detail in the Biophysical chapter/Volume 2, Chapter 6 and the Developability chapter/Volume 2, Chapter 7). As the temperature is elevated, conformational unfolding of a given domain begins at a temperature dependent on the given protein ( $T_{\text{onset}}$ ,  $\approx 60$  °C for NISTmAb in formulation buffer), and the point at which half of the protein molecules are unfolded is defined as the thermal transition temperature, or melting temperature ( $T_m$ ,  $\approx 69$  °C for NISTmAb in formulation buffer). In the current chapter, the interest is the induction of chemical modifications that may occur *prior to* unfolding of the protein but which also may result in decreased product stability and/or activity. Therefore, the temperature for such thermal forced degradation is often chosen to be below the  $T_{\text{onset}}$  and well below the  $T_m$ . Elevated temperature for thermal forced degradation is therefore somewhat molecule dependent but can be generalized for molecules of the same classes. Under such conditions the kinetics of all potential pathways (e.g., unfolding, aggregation) may be increased, but the intent is to observe chemical modification associated with oxidation and deamidation without inducing significant changes in the overall conformational structure.

### *pH Stress*

In addition to thermal stress, forced degradation analysis may utilize extremes in pH at ambient or elevated temperatures (83). Therapeutic proteins such as IgGs nearly always encounter some form of elevated and/or lowered pH during downstream processing, such as during elution from a Protein A column during the initial product capture step. A number of additional pH changes may be encountered during formulation, freeze/thaw cycles due to cryo-concentration, and/or dilution during patient infusion. Assessing the protein's stability in response to pH excursions can be a measure of protein robustness during formulation/processing conditions. For example, at elevated pH and temperature, the rate of deamidation is increased as a result of increased deprotonation of the backbone nitrogen (see Table 1) to initiate the succinimide intermediate formation. Acidic conditions, on the other hand, induce deamidation as a result of direct hydrolysis of Asn to produce mainly Asp. In addition to deamidation, acidic conditions may also result in fragmentation of backbone peptide bonds, such as the conserved SCDKTH region of IgG1, which was shown to be prone to cleavage (Asp-Lys bond) in four IgG1 antibodies upon storage under acidic conditions (234).

### *Oxidative Stress*

Oxidation of specific amino acid residues may affect the conformation of the given protein and/or its activity. Biopharmaceuticals may be exposed to oxidizing conditions such as residual cleaning agents, excipients, and stainless steel components during production. Forced oxidation of therapeutic proteins is



a particularly important chemical modification as a number of residues may be sensitive to oxidation, including Met, His, Cys, Tyr, and Trp. H<sub>2</sub>O<sub>2</sub> is the most commonly used chemical oxidant and predominantly results in oxidation of Met residues. Oxidizing agents such as t-BHP also have been suggested to allow more selective oxidation of only surface-exposed residues. Metal ions, both in the presence or absence of peroxides, may be used to evaluate the susceptibility of protein to metal-catalyzed oxidation and may result in oxidation of additional amino acids, including His, Cys, and Trp.

### *NISTmAb Stress Results*

The forced degradation conditions used in this study are expected to have greater impact on oxidation and deamidation, with little to no impact on the levels of other PTMs. A selected number of peptides were chosen to demonstrate the type of data and interpretation sought during forced degradation analysis, the results of which are shown in Figure 7. These peptides included residues of interest that have previously been shown to be prone to modification upon stress in mAbs of the IgG1 $\kappa$  class, as well as NISTmAb-specific CDR regions.

Two peptides were chosen as representative sites for monitoring deamidation under forced degradation conditions, including the H(305-320) and H(374-395) peptides of the C<sub>H</sub>2, and C<sub>H</sub>3 regions, respectively. The H(374-395) PENNYK peptide was discussed in detail previously, highlighting the complexity of quantifying deamidation due to the fact that it contains three potential sites of deamidation, each of which may form a variety of deamidation products (Asp, Asu, iso-Asp) and may have slightly different and/or overlapping retention times depending on the chromatographic methods used. For this reason, all potential deamidation products, regardless of retention time, were summed in the current example.

Figure 7 shows the relative abundance of deamidation products identified for the H(374-395) peptide in the forced degradation samples. As expected, the low pH and thermal stress samples showed relatively minor increases in deamidation. Fe oxidation and H<sub>2</sub>O<sub>2</sub> samples also were slightly higher, likely due to the elevated temperature (25 °C versus control at 4 °C) during the treatment. On the other hand, the high pH stress sample showed a large increase in deamidation from 0.7% (control) to 29.5% (high pH), a change of 28.8%. This is indicative of a site that is relatively prone to chemical deamidation. As discussed previously, the H(374-395) PENNYK peptide is conserved among IgG1 mAbs and has been previously reported as having relatively high susceptibility to deamidation (79). This conserved motif provides a class-specific “marker” for chemical deamidation, and can therefore be used as an internal standard of sorts to gauge the relative susceptibility of other motifs in the molecule.

The second peptide monitored, H(305-320), also displays a relatively small increase in deamidation under low pH and thermal stress conditions. Differing from the previous example, this peptide shows relatively low levels of deamidation under the high pH stress conditions as well (1.1% [control] to 5.6% [high pH], a change of 4.5%). This peptide is part of the conserved region

of the CH<sub>2</sub> domain and has previously been reported in other antibodies to have a slower rate of deamidation than the PENNYK peptide, thought to be due in part to additional conformational constraints imparted by the protein's secondary structure in this region (79). H(305-320) serves as an example of a peptide that is relatively "protected" from artificial deamidation conditions in this experiment

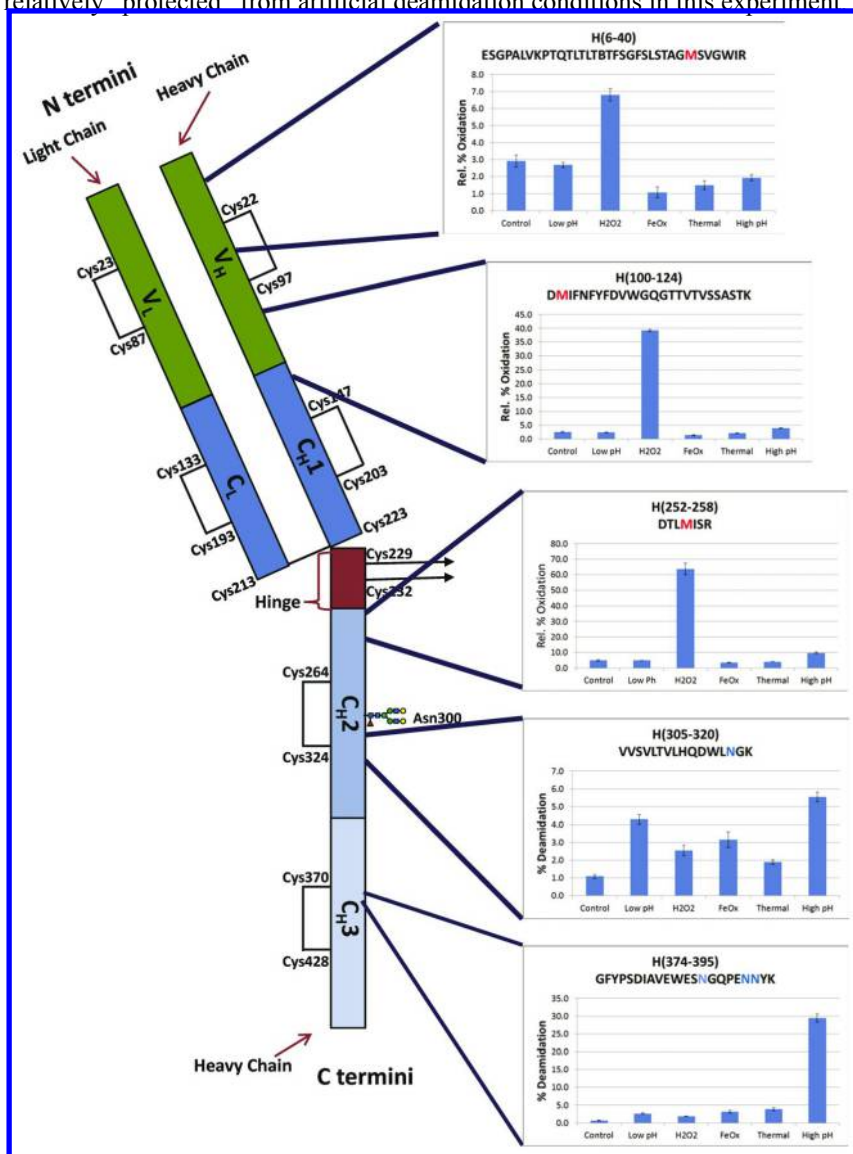


Figure 7. Percent relative abundance of selected peptides containing potential chemical modification sites. Relative abundance calculations on each forced degradation sample were collected as described previously by Lab 3. Error bars represent  $\pm 1$  standard deviation (SD) for triplicate analyses (three individual digestions of the same forced degradation sample). (see color insert)

Three peptides containing potentially oxidizable residues also were chosen for analysis, including the NISTmAb CDR1-containing peptide H(6-40); the NISTmAb CDR3-containing peptide H(100-124); and H(252-258), which has been demonstrated to be highly prone to chemical oxidation in mAbs of the same IgG1k class.

Beginning the discussion with H(252-258) allows visualization of results typical of chemical oxidation during forced degradation for highly prone sites. This motif is conserved in all IgGs and provides a “marker” for chemical oxidation; it therefore can be used as an internal standard of sorts to gauge the relative susceptibility of other motifs in the molecule. Figure 7 shows the chemical oxidation in the H<sub>2</sub>O<sub>2</sub>-stressed sample, increased from 4.9% (control) to 67.3% (H<sub>2</sub>O<sub>2</sub>), a change of 62.4%. Oxidation under each of the other stress conditions remained relatively consistent. As mentioned previously, the M255 site has been shown to be highly prone to chemical oxidation. This residue is just below the highly flexible hinge region, and therefore it is not unexpected that it would be rather solvent exposed and more prone to oxidation. From a risk assessment point of view, the impact of oxidation at this site may initially be viewed as relatively low as it is located far from the antigen binding regions of the molecule. However, it has been linked to altered FcRn binding, which consequently can affect the pharmacokinetics of the therapeutic mAb (235). The combination of its frequency of occurrence in all IgG1 antibodies and susceptibility to oxidation makes it an attribute to monitor during process optimization where variations in oxidative insults during production may affect levels of oxidation in the final product.

The next two peptides chosen to highlight the results of oxidative stress are H(6-40) and H(100-124), as shown in Figure 7. Both of these peptides contain a potentially oxidizable Met, as well as an oxidizable Trp. Under the experimental conditions, the Met is expected to be the predominantly oxidizable residue, and MS/MS site analysis was used along with retention time to confirm Met oxidation. Both peptides also are within a CDR region, with M34 being part of CDR1 and M101 being part of CDR3. Their native composition and stability are therefore important to consider as potential critical quality attributes. Upon stress, the H(6-40) peptide was shown to have a small increase in relative oxidation level from 2.9% (control) to 6.8% (H<sub>2</sub>O<sub>2</sub>), a change of 3.9%. This indicates that M34 has relatively low susceptibility for chemical modification. On the other hand, H(100-124) was shown to have significantly increased oxidation from 2.5% (control) to 39.3% (H<sub>2</sub>O<sub>2</sub>), a change of 36.8%. The trend in oxidation under various stress conditions is very similar to that shown for M255, discussed in the previous paragraph, and is indicative of a site prone to chemical oxidation.

Ultimately, both the M34 and M101 residues are potential critical quality attributes as both are indeed contained in a CDR region of the NISTmAb; the criticality of the M101 residue is somewhat higher, however, due to its higher susceptibility to chemical oxidation. Further studies would be required in conjunction with activity-based measurements to elucidate any structure–function relationships that may result from chemical oxidation at this site. Manufacturing consistency also would be of importance during process optimization and lifecycle management because this site has been shown to be susceptible to chemical oxidation.

In the case of all three peptides monitored for oxidation in the above NISTmAb example, good intra-laboratory reproducibility was demonstrated on the triplicate samples, as indicated by the error bars in Figure 7. In addition to evaluating intra-laboratory reproducibility, it was of interest to determine if despite the differences in digestion conditions between laboratories noted for the control samples, each laboratory could confidently identify changes in PTM levels as a result of forced degradation. For this reason, a single NISTmAb oxidative stress sample was prepared in Lab 1. In this case, the sample (8 mL of 10 mg/mL NISTmAb in 25 mmol/L His, pH 6.0) was adjusted by adding 10  $\mu$ L of 30%  $H_2O_2$ , resulting in a final concentration of 0.04%  $H_2O_2$ . The sample was maintained in the dark at 37  $^{\circ}C$  for 4 hours, and then quenched by adding 96 mg of solid Met. The sample was sent to each of the three laboratories for analysis, and each laboratory was asked to monitor the three peptides H(6-40), H(100-124), and H(252-258).

Figure 8 demonstrates that despite differences in values of oxidation observed due to sample preparation, all three laboratories were able to confidently identify relative changes in oxidation upon peroxide stress. All three laboratories identified the relative susceptibility to oxidation in the order of H(252-258) > H(100-124) > H(6-40). These findings indicate that despite minor differences in oxidation due to sample preparation, changes in oxidation as measured in a forced degradation analysis can confidently identify hot spots prone to chemical degradation. These observed differences in total oxidation are indicative of their respective solvent accessibility and provide insight as to which residues are most prone to oxidative degradation.

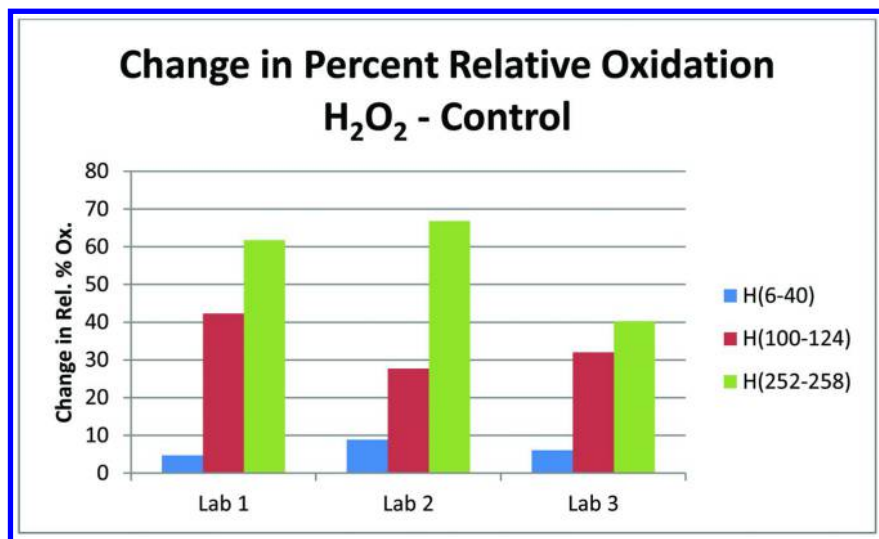


Figure 8. Comparison of relative post-translational modification (PTM) levels between control and  $H_2O_2$ -oxidized samples analyzed in three laboratories. (see color insert)

## Conclusions for the Interlaboratory Study

This seminal cross-site study demonstrates the use of peptide mapping for monitoring and quantifying frequently occurring attributes. Further, the comparison of results across laboratories highlights the dependence of outcomes on a myriad parameters, including sample storage and preparation, instrumentation, and data analysis and interpretation methods. The largest variances in results obtained across different platforms are likely to be observed for low-abundance PTMs simply due to the increased difficulties associated with accurately detecting species at or near interfering background signals. Peptide mapping is an intricate technique requiring skilled analysts with thorough knowledge regarding not only the nature of the sample being analyzed and its potential PTMs, but also the strengths and weaknesses of the instrumentation and methods employed for data collection and analysis.

Although minor variations in relative abundance values were reported, similar trends were identified in all laboratories. In addition, the power of peptide mapping is clearly evident in that each laboratory in this study chose some form of this experiment as the platform method to analyze PTM levels and gain an understanding of which residues may be more inclined toward modification and to what relative degree. In a typical development environment, the residues identified as potential degradation sites may be further evaluated for their tendency to degrade under certain conditions as well as the effect that such product-related variants may have on safety and/or efficacy. If the attribute level is too low to be accurately quantified with high-resolution MS, product-specific MRM methods on a triple quadrupole instrument may be developed for peptides showing a propensity for degradation. Such experiments would serve to validate tentatively identified degradation pathways as well as establish a more traceable analytical method for understanding the kinetics of these processes.

Development of targeted MRM methods for every candidate molecule is not feasible due to the high volume of samples produced throughout the drug development process. It is imperative to develop robust peptide mapping methods that are not overly sensitive to inevitable variations such as those that may occur when two samples are prepared by two different analysts or by the same analyst on different days even when the same methods are used. Although challenging, every effort should be made to optimize methods to produce the most accurate and reproducible values possible without the introduction of confounding variations from processing and handling. Finally, the reliability and reproducibility of results, whether qualitative or quantitative, further demand the use of sensitive instruments capable of high resolution/accurate mass measurements. By adhering to these principles and continuing to strive for improvements as technology develops, analytical confidence in peptide identification, method robustness, and accuracy of quantitation will continue to increase.

## Closing Remarks

PTMs are major contributors to the heterogeneity of recombinant proteins, and the problem is compounded in the case of mAbs due to their large size and the presence of multiple subunits. They can be introduced during any stage of production, from cell culture to purification and formulation of the final product. Rigorous discovery, tracking, and evaluation of protein changes are critical for optimization of the manufacturing process in order to deliver a safe and efficacious drug product. In addition to high-abundance N- and C-terminal variants, a variety of low-abundance deamidation, oxidation, and glycation sites were identified in the control NISTmAb sample.

Various techniques can be used to quantify PTMs, and we have demonstrated their effectiveness with a cross-site effort; it remains a challenging task to ensure the precision and accuracy of analytical techniques, however, especially for quantifying minor PTMs. Sample preparation, analytical methods (both LC and MS), and data analysis were identified as potential steps that may result in interlaboratory variation in the PTMs identified. Software versus manual reconstruction of chromatograms was shown in a few examples throughout this study to contribute to differences in the relative abundance values reported for a particular PTM. This emphasizes the importance of understanding the algorithms behind the software, especially their limitations (e.g., most software will not use retention time information to distinguish true oxidation versus in-source oxidation). Despite this, rapid software evaluation of data is absolutely critical to ensure high levels of efficiency in drug development pipelines. Routine analysis of a plethora of samples is simply not possible without such software. In addition, as mentioned before, it is consistency (with both data collection and analysis) within an organization that is most important for ensuring consistency of the product and identifying changes from one material to the next. As observed in our study, the interlaboratory CVs are always higher than the intra-laboratory CVs, highlighting the effect of consistency in sample preparation and analysis on quantified PTM levels. Although detailed intermediate precision assessment across laboratories was not possible during the course of this study, it is most certainly a future interest and future application of the NISTmAb.

A forced degradation analysis also was performed in Lab 3, and data were reported for a select number of peptides to highlight representative results from such studies. Forced deamidation under elevated pH and temperature, for example, was observed at different relative levels for different peptides. The peptide H(305-320) showed low levels of induced deamidation, whereas the PENNYK peptide [H(374-395)] was used to highlight results for a more susceptible region, as well as the difficulty in assessing chemical deamidation compounded by the presence of multiple potential deamidation sites. Oxidation prone residues also were highlighted in the NISTmAb following exaggerated chemical oxidation conditions. Salient examples include M101 and M255 of the heavy chain, which demonstrate results commensurate with residues more highly susceptible to oxidation under chemical stress conditions, as well as M34 of the heavy chain, which is only moderately oxidizable in comparison. Although levels of reported modification relative abundance were shown to be

dependent on sample preparation and analytical method, each laboratory was capable of reporting consistent trends in the relative susceptibility to degradation as demonstrated with a second interlaboratory chemical oxidation study.

The importance of robust methods that do not themselves introduce PTMs and thereby confound the true levels of modification was apparent in the current study. Consistent sample processing, analysis, and data processing were all identified as critical stages for control. Future interlaboratory peptide mapping studies will likely be designed to selectively target particular stages of the work flow. For example, pre-digested samples may be distributed with or without defined analytical methodology to highlight those aspects associated with instrumentation and/or data processing without confounding modifications from differences in sample processing. Standard reference data in the form of a common raw data file also may be a potential application of the NISTmAb to selectively evaluate XIC settings for PTM quantification. The widespread availability of the NISTmAb IgG1κ reference material (RM) will afford future opportunities for such interlaboratory studies to further elucidate method-dependent variability and guide innovation in PTM characterization methodology.

## Disclaimer

NIST Disclaimer: Commercial equipment, instruments, and materials are identified to adequately specify the experimental procedure and do not imply a recommendation or endorsement by NIST.

## References

- Walsh, C. *Post-Translational Modification of Proteins: Expanding Nature's Inventory*; Roberts and Co. Publishers: Englewood, CO, 2006; p 490.
- Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 7342–7372.
- Creasy, D. M.; Cottrell, J. S. *Proteomics* **2004**, *4*, 1534–1536.
- Consortium, I. H. G. S. *Nature* **2004**, *431*, 931–945.
- Jensen, O. N. *Curr. Opin. Chem. Biol.* **2004**, *8*, 33–41.
- Kohler, G.; Milstein, C. *Nature* **1975**, *256*, 495–497.
- Ansar, W.; Ghosh, S. *Ind. J. Clin. Med.* **2013**, *4*, 9–21.
- Edwards, P. A. *Biochem. J.* **1981**, *200*, 1–10.
- Li, J.; Zhu, Z. *Acta Pharmacol. Sin.* **2010**, *31*, 1198–1207.
- Tyagi, S.; Sharma, P. K.; Kumar, N.; Visht, S. *Int. J. PharmTech Res.* **2011**, *3*, 459–463.
- Zola, H. Monoclonal Antibodies *eLS* [Online], 2010. <http://www.els.net> (accessed January 2015).
- Smith, S. L. *J Transpl Coord* **1996**, *6*, 109–119quiz 120-101.
- Genazzani, A. A.; Biggio, G.; Caputi, A. P.; Del Tacca, M.; Drago, F.; Fantozzi, R.; Canonico, P. L. *BioDrugs* **2007**, *21*, 351–356.
- Patient Protection and Affordable Care Act*; Public Law 111–148; United States Government, 2010; pp 119, 804–821.

15. EMA Adopts Guideline on Biosimilar Monoclonal Antibodies *GaBi Online* [Online], 2011. <http://www.gabionline.net/Guidelines/EMA-adopts-guideline-on-biosimilar-monoclonal-antibodies> (accessed January 2015).
16. Schneider, C. K.; Vleminckx, C.; Gravanis, I.; Ehmann, F.; Trouvin, J. H.; Weise, M.; Thirstrup, S. *Nat. Biotechnol.* **2012**, *30*, 1179–1185.
17. Tsiftoglou, A. S.; Ruiz, S.; Schneider, C. K. *BioDrugs* **2013**, *27*, 203–211.
18. Rockoff, J. D. European Commission Approves Biosimilar of J&J and Merck's Remicade *The Wall Street Journal* [Online], 2013. <http://online.wsj.com/news/articles/SB10001424127887324094704579067360123704206> (accessed January 2015).
19. FDA Approves First Biosimilar Product Zarxio, 2015. U.S. Food and Drug Administration. <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm436648.htm> (accessed September 2015).
20. Du, Y.; Walsh, A.; Ehrick, R.; Xu, W.; May, K.; Liu, H. *mAbs* **2012**, *4*, 578–585.
21. Srebalus Barnes, C. A.; Lim, A. *Mass Spectrom. Rev.* **2007**, *26*, 370–388.
22. Whitelegge, J. *Expert Rev. Proteomics* **2013**, *10*, 127–129.
23. Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. *Curr. Protoc. Mol. Biol.* **2009** Chapter 10, Unit10.25.
24. Wu, S.; Lourette, N. M.; Tolic, N.; Zhao, R.; Robinson, E. W.; Tolmachev, A. V.; Smith, R. D.; Pasa-Tolic, L. *J. Proteome Res.* **2009**, *8*, 1347–1357.
25. Chevreux, G.; Tilly, N.; Bihoreau, N. *Anal. Biochem.* **2011**, *415*, 212–214.
26. Marvin, L. F.; Roberts, M. A.; Fay, L. B. *Clin. Chim. Acta* **2003**, *337*, 11–21.
27. Beck, A.; Bussat, M. C.; Zorn, N.; Robillard, V.; Klinguer-Hamour, C.; Chenu, S.; Goetsch, L.; Corvaia, N.; Van Dorsselaer, A.; Haeuw, J. F. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *819*, 203–218.
28. Quiles, S.; Raisch, K. P.; Sanford, L. L.; Bonner, J. A.; Safavy, A. *J. Med. Chem.* **2010**, *53*, 586–594.
29. Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. *Anal. Chem.* **2014**, *86*, 317–320.
30. Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 4239–4246.
31. Valeja, S. G.; Kaiser, N. K.; Xian, F.; Hendrickson, C. L.; Rouse, J. C.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 8391–8395.
32. Brady, L. J.; Valliere-Douglass, J.; Martinez, T.; Balland, A. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 502–509.
33. Zhang, J.; Liu, H.; Katta, V. *J. Mass Spectrom.* **2010**, *45*, 112–120.
34. Debaene, F.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. *Anal. Chem.* **2013**, *85*, 9785–9792.
35. Fenn, L. S.; McLean, J. A. *Methods Mol. Biol.* **2013**, *951*, 171–194.
36. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Jr. *J. Mass Spectrom.* **2008**, *43*, 1–22.
37. Bagal, D.; Valliere-Douglass, J. F.; Balland, A.; Schnier, P. D. *Anal. Chem.* **2010**, *82*, 6751–6755.



38. 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.
39. 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.
40. Li, W.; Song, C.; Bailey, D. J.; Tseng, G. C.; Coon, J. J.; Wysocki, V. H. *Anal. Chem.* **2011**, *83*, 9540–9545.
41. Wiesner, J.; Premisler, T.; Sickmann, A. *Proteomics* **2008**, *8*, 4466–4483.
42. Abès, R.; Teillaud, J.-L. *Pharmaceuticals* **2010**, *3*, 146–157.
43. Huhn, C.; Selman, M. H.; Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. *Proteomics* **2009**, *9*, 882–913.
44. Siberil, S.; Dutertre, C. A.; Fridman, W. H.; Teillaud, J. L. *Crit. Rev. Oncol. Hematol.* **2007**, *62*, 26–33.
45. Takahashi, M.; Kuroki, Y.; Ohtsubo, K.; Taniguchi, N. *Carbohydr. Res.* **2009**, *344*, 1387–1390.
46. Abes, R.; Dutertre, C. A.; Agnelli, L.; Teillaud, J. L. *Expert. Rev. Clin. Immunol.* **2009**, *5*, 735–747.
47. Dick, L. W., Jr.; Qiu, D.; Mahon, D.; Adamo, M.; Cheng, K. C. *Biotechnol. Bioeng.* **2008**, *100*, 1132–1143.
48. Luo, J.; Zhang, J.; Ren, D.; Tsai, W. L.; Li, F.; Amanullah, A.; Hudson, T. *Biotechnol. Bioeng.* **2012**, *109*, 2306–2315.
49. Antes, B.; Amon, S.; Rizzi, A.; Wiederlum, S.; Kainer, M.; Szolar, O.; Fido, M.; Kircheis, R.; Nechansky, A. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *852*, 250–256.
50. Khawli, L. A.; Goswami, S.; Hutchinson, R.; Kwong, Z. W.; Yang, J.; Wang, X.; Yao, Z.; Sreedhara, A.; Cano, T.; Tesar, D.; Nijem, I.; Allison, D. E.; Wong, P. Y.; Kao, Y. H.; Quan, C.; Joshi, A.; Harris, R. J.; Motchnik, P. *mAbs* **2010**, *2*, 613–624.
51. Dick, L. W., Jr.; Kim, C.; Qiu, D.; Cheng, K. C. *Biotechnol. Bioeng.* **2007**, *97*, 544–553.
52. Liu, Y. D.; Goetze, A. M.; Bass, R. B.; Flynn, G. C. *J. Biol. Chem.* **2011**, *286*, 11211–11217.
53. Schilling, S.; Wasternack, C.; Demuth, H. U. *Biol. Chem.* **2008**, *389*, 983–991.
54. Li, P.; Roller, P. P. *Curr. Top. Med. Chem.* **2002**, *2*, 325–341.
55. Rink, R.; Arkema-Meter, A.; Baudoin, I.; Post, E.; Kuipers, A.; Nelemans, S. A.; Akanbi, M. H.; Moll, G. N. *J. Pharmacol. Toxicol. Methods* **2010**, *61*, 210–218.
56. Chelius, D.; Jing, K.; Lueras, A.; Rehder, D. S.; Dillon, T. M.; Vizel, A.; Rajan, R. S.; Li, T.; Treuheit, M. J.; Bondarenko, P. V. *Anal. Chem.* **2006**, *78*, 2370–2376.
57. Liu, H.; Gaza-Bulseco, G.; Chumsae, C. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 4081–4088.
58. McAdam, S. N.; Fleckenstein, B.; Rasmussen, I. B.; Schmid, D. G.; Sandlie, I.; Bogen, B.; Viner, N. J.; Sollid, L. M. *J. Exp. Med.* **2001**, *193*, 1239–1246.

59. Terashima, I.; Koga, A.; Nagai, H. *Anal. Biochem.* **2007**, *368*, 49–60.
60. Yang, H.; Zubarev, R. A. *Electrophoresis* **2010**, *31*, 1764–1772.
61. Robinson, A. B.; Scotchler, J. W.; McKerrow, J. H. *J. Am. Chem. Soc.* **1973**, *95*, 8156–8159.
62. Scotchler, J. W.; Robinson, A. B. *Anal. Biochem.* **1974**, *59*, 319–322.
63. Tyler-Cross, R.; Schirch, V. *J. Biol. Chem.* **1991**, *266*, 22549–22556.
64. Wright, H. T. *Protein Eng.* **1991**, *4*, 283–294.
65. Sreedhara, A.; Cordoba, A.; Zhu, Q.; Kwong, J.; Liu, J. *Pharm Res.* **2012**, *29*, 187–197.
66. Brennan, T. V.; Clarke, S. *Int. J. Pept. Protein Res.* **1995**, *45*, 547–553.
67. Chu, G. C.; Chelius, D.; Xiao, G.; Khor, H. K.; Coulibaly, S.; Bondarenko, P. V. *Pharm Res.* **2007**, *24*, 1145–1156.
68. Harris, R. J.; Kabakoff, B.; Macchi, F. D.; Shen, F. J.; Kwong, M.; Andya, J. D.; Shire, S. J.; Bjork, N.; Totpal, K.; Chen, A. B. *J. Chromatogr. B: Biomed. Sci. Appl.* **2001**, *752*, 233–245.
69. Cacia, J.; Keck, R.; Presta, L. G.; Frenz, J. *Biochemistry* **1996**, *35*, 1897–1903.
70. Rehder, D. S.; Chelius, D.; McAuley, A.; Dillon, T. M.; Xiao, G.; Crouse-Zineddini, J.; Vardanyan, L.; Perico, N.; Mukku, V.; Brems, D. N.; Matsumura, M.; Bondarenko, P. V. *Biochemistry* **2008**, *47*, 2518–2530.
71. Wakankar, A. A.; Borchardt, R. T.; Eigenbrot, C.; Shia, S.; Wang, Y. J.; Shire, S. J.; Liu, J. L. *Biochemistry* **2007**, *46*, 1534–1544.
72. Wakankar, A. A.; Liu, J.; Vandervelde, D.; Wang, Y. J.; Shire, S. J.; Borchardt, R. T. *J. Pharm. Sci.* **2007**, *96*, 1708–1718.
73. Liu, H.; Gaza-Bulsecu, G.; Faldu, D.; Chumsae, C.; Sun, J. J. *Pharm. Sci.* **2008**, *97*, 2426–2447.
74. Wakankar, A. A.; Borchardt, R. T. *J. Pharm. Sci.* **2006**, *95*, 2321–2336.
75. Patel, K.; Borchardt, R. T. *J. Parenter. Sci. Technol.* **1990**, *44*, 300–301.
76. Peters, B.; Trout, B. L. *Biochemistry* **2006**, *45*, 5384–5392.
77. Geiger, T.; Clarke, S. *J. Biol. Chem.* **1987**, *262*, 785–794.
78. Aswad, D. W.; Paranandi, M. V.; Schurter, B. T. *J. Pharm. Biomed. Anal.* **2000**, *21*, 1129–1136.
79. Chelius, D.; Rehder, D. S.; Bondarenko, P. V. *Anal. Chem.* **2005**, *77*, 6004–6011.
80. Li, B.; Gorman, E. M.; Moore, K. D.; Williams, T.; Schowen, R. L.; Topp, E. M.; Borchardt, R. T. *J. Pharm. Sci.* **2005**, *94*, 666–675.
81. Xiao, G.; Bondarenko, P. V.; Jacob, J.; Chu, G. C.; Chelius, D. *Anal. Chem.* **2007**, *79*, 2714–2721.
82. Huang, L.; Lu, J.; Wroblewski, V. J.; Beals, J. M.; Rigglin, R. M. *Anal. Chem.* **2005**, *77*, 1432–1439.
83. Liu, H.; Gaza-Bulsecu, G.; Sun, J. J. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2006**, *837*, 35–43.
84. Perkins, M.; Theiler, R.; Lunte, S.; Jeschke, M. *Pharm. Res.* **2000**, *17*, 1110–1117.
85. Zhang, W.; Czupryn, M. J. *J. Pharm. Biomed. Anal.* **2003**, *30*, 1479–1490.
86. Oliyai, C.; Borchardt, R. T. *Pharm. Res.* **1993**, *10*, 95–102.
87. Smyth, D. G.; Stein, W. H.; Moore, S. *J. Biol. Chem.* **1963**, *238*, 227–234.

88. Johnson, B. A.; Aswad, D. W. *Anal. Biochem.* **1991**, *192*, 384–391.
89. Potter, S. M.; Henzel, W. J.; Aswad, D. W. *Protein Sci.* **1993**, *2*, 1648–1663.
90. Lehmann, W. D.; Schlosser, A.; Erben, G.; Pipkorn, R.; Bossemeyer, D.; Kinzel, V. *Protein Sci.* **2000**, *9*, 2260–2268.
91. Cournoyer, J. J.; Lin, C.; Bowman, M. J.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 48–56.
92. Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *Anal. Chem.* **2006**, *78*, 1264–1271.
93. Cournoyer, J. J.; Pittman, J. L.; Ivleva, V. B.; Fallows, E.; Waskell, L.; Costello, C. E.; O'Connor, P. B. *Protein Sci.* **2005**, *14*, 452–463.
94. O'Connor, P. B.; Cournoyer, J. J.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 15–19.
95. Harris, R. J.; Shire, S. J.; Winter, C. *Drug Dev. Res.* **2004**, *61*, 137–154.
96. Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. *Anal. Chem.* **2006**, *78*, 6645–6650.
97. Gaza-Bulsecu, G.; Li, B.; Bulsecu, A.; Liu, H. C. *Anal. Chem.* **2008**, *80*, 9491–9498.
98. Li, X.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 855–864.
99. Bhandary, B.; Marahatta, A.; Kim, H. R.; Chae, H. J. *Int. J. Mol. Sci.* **2012**, *14*, 434–456.
100. Farber, J. L. *Environ. Health Perspect.* **1994**, *102* (Suppl 10), 17–24.
101. Harper, M. E.; Bevilacqua, L.; Hagopian, K.; Weindruch, R.; Ramsey, J. J. *Acta Physiol. Scand.* **2004**, *182*, 321–331.
102. Schrader, M.; Fahimi, H. D. *Biochim. Biophys. Acta* **2006**, *1763*, 1755–1766.
103. Davies, M. J. *Biochim. Biophys. Acta* **2005**, *1703*, 93–109.
104. Manke, A.; Wang, L.; Rojanasakul, Y. *Biomed. Res. Int.* **2013**, *2013*, 942916.
105. Stadtman, E. R. *Annu. Rev. Biochem.* **1993**, *62*, 797–821.
106. Stadtman, E. R.; Berlett, B. S. *J. Biol. Chem.* **1991**, *266*, 17201–17211.
107. Vallyathan, V.; Shi, X. *Environ. Health Perspect.* **1997**, *105* (Suppl 1), 165–177.
108. Kim, Y. H.; Berry, A. H.; Spencer, D. S.; Stites, W. E. *Protein Eng.* **2001**, *14*, 343–347.
109. Liu, H.; Gaza-Bulsecu, G.; Xiang, T.; Chumsae, C. *Mol. Immunol.* **2008**, *45*, 701–708.
110. Hawkins, C. L.; Davies, M. J. *Biochim. Biophys. Acta* **2001**, *1504*, 196–219.
111. Strlic, M.; Kolar, J.; Pihlar, B. *Acta Chim. Slov.* **1999**, *46*, 555–566.
112. Neumann, N. P. *Methods Enzymol.* **1972**, *25*, 393–400.
113. Kowalik-Jankowska, T.; Ruta, M.; Wisniewska, K.; Lankiewicz, L.; Dyba, M. *J. Inorg. Biochem.* **2004**, *98*, 940–950.
114. Li, S.; Nguyen, T. H.; Schoneich, C.; Borchardt, R. T. *Biochemistry* **1995**, *34*, 5762–5772.
115. Uchida, K.; Enomoto, N.; Itakura, K.; Kawakishi, S. *Agric Biol. Chem.* **1989**, *53*, 3285–3292.
116. Zhao, F.; Ghezzi-Schoneich, E.; Aced, G. I.; Hong, J.; Milby, T.; Schoneich, C. *J. Biol. Chem.* **1997**, *272*, 9019–9029.

117. Amano, M.; Kobayashi, N.; Yabuta, M.; Uchiyama, S.; Fukui, K. *Anal. Chem.* **2014**, *86*, 7536–7543.
118. Lu, H. S.; Fausset, P. R.; Narhi, L. O.; Horan, T.; Shinagawa, K.; Shimamoto, G.; Boone, T. C. *Arch. Biochem. Biophys.* **1999**, *362*, 1–11.
119. Mulinacci, F.; Capelle, M. A.; Gurny, R.; Drake, A. F.; Arvinte, T. *J. Pharm. Sci.* **2011**, *100*, 451–463.
120. Taggart, C.; Cervantes-Laurean, D.; Kim, G.; McElvaney, N. G.; Wehr, N.; Moss, J.; Levine, R. L. *J. Biol. Chem.* **2000**, *275*, 27258–27265.
121. Chumsae, C.; Gaza-Bulseco, G.; Sun, J.; Liu, H. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *850*, 285–294.
122. Lam, X. M.; Yang, J. Y.; Cleland, J. L. *J. Pharm. Sci.* **1997**, *86*, 1250–1255.
123. Shen, F. J.; Kwong, M. Y.; Keck, R. G.; Harris, R. J. *The Application of Tert-Butylhydroperoxide Oxidation to Study Sites of Potential Methionine Oxidation in a Recombinant Antibody*; Marshak, D. R., Ed.; Academic Press, Inc.: San Diego, CA, 1996; Vol. VII.
124. 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.
125. Bertolotti-Ciarlet, A.; Wang, W.; Lownes, R.; Pristatsky, P.; Fang, Y.; McKelvey, T.; Li, Y.; Li, Y.; Drummond, J.; Prueksaritanont, T.; Vlasak, J. *Mol. Immunol.* **2009**, *46*, 1878–1882.
126. Pan, H.; Chen, K.; Chu, L.; Kinderman, F.; Apostol, I.; Huang, G. *Protein Sci.* **2009**, *18*, 424–433.
127. Wang, W.; Vlasak, J.; Li, Y.; Pristatsky, P.; Fang, Y.; Pittman, T.; Roman, J.; Wang, Y.; Prueksaritanont, T.; Ionescu, R. *Mol. Immunol.* **2011**, *48*, 860–866.
128. Ji, J. A.; Zhang, B.; Cheng, W.; Wang, Y. J. *J. Pharm. Sci.* **2009**, *98*, 4485–4500.
129. Duenas, E. T.; Keck, R.; De Vos, A.; Jones, A. J.; Cleland, J. L. *Pharm. Res.* **2001**, *18*, 1455–1460.
130. Keck, R. G. *Anal. Biochem.* **1996**, *236*, 56–62.
131. Wagner, R. M.; Fraser, B. A. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 69–72.
132. Lagerwerf, F. M.; van de Weert, M.; Heerma, W.; Haverkamp, J. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1905–1910.
133. Gevaert, K.; Van Damme, P.; Ghesquiere, B.; Vandekerckhove, J. *Biochim. Biophys. Acta* **2006**, *1764*, 1801–1810.
134. Srikanth, R.; Wilson, J.; Bridgewater, J. D.; Numbers, J. R.; Lim, J.; Olbris, M. R.; Kettani, A.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1499–1506.
135. Wong, C.; Strachan-Mills, C.; Burman, S. *J. Chromatogr. A* **2012**, *1270*, 153–161.
136. Sreedhara, A.; Lau, K.; Li, C.; Hosken, B.; Macchi, F.; Zhan, D.; Shen, A.; Steinmann, D.; Schoneich, C.; Lentz, Y. *Mol. Pharm.* **2013**, *10*, 278–288.
137. Okajima, T.; Kawata, Y.; Hamaguchi, K. *Biochemistry* **1990**, *29*, 9168–9175.
138. Donbrow, M.; Azaz, E.; Pillersdorf, A. *J. Pharm. Sci.* **1978**, *67*, 1676–1681.
139. Ha, E.; Wang, W.; Wang, Y. J. *J. Pharm. Sci.* **2002**, *91*, 2252–2264.

140. Yao, J.; Dokuru, D. K.; Noestheden, M.; Park, S. S.; Kerwin, B. A.; Jona, J.; Ostovic, D.; Reid, D. L. *Pharm. Res.* **2009**, *26*, 2303–2313.
141. Lam, X. M.; Lai, W. G.; Chan, E. K.; Ling, V.; Hsu, C. C. *Pharm. Res.* **2011**, *28*, 2543–2555.
142. Grosvenor, A. J.; Morton, J. D.; Dyer, J. M. *Amino Acids* **2010**, *39*, 285–296.
143. Simat, T. J.; Steinhart, H. J. *Agric. Food Chem.* **1998**, *46*, 490–498.
144. Bienvenut, W. V.; Deon, C.; Pasquarello, C.; Campbell, J. M.; Sanchez, J. C.; Vestal, M. L.; Hochstrasser, D. F. *Proteomics* **2002**, *2*, 868–876.
145. Li, Y.; Polozova, A.; Gruia, F.; Feng, J. *Anal. Chem.* **2014**, *86*, 6850–6857.
146. Roeser, J.; Permentier, H. P.; Bruins, A. P.; Bischoff, R. *Anal. Chem.* **2010**, *82*, 7556–7565.
147. Wei, Z.; Feng, J.; Lin, H. Y.; Mullapudi, S.; Bishop, E.; Tous, G. I.; Casas-Finet, J.; Hakki, F.; Strouse, R.; Schenerman, M. A. *Anal. Chem.* **2007**, *79*, 2797–2805.
148. Qi, P.; Volkin, D. B.; Zhao, H.; Nedved, M. L.; Hughes, R.; Bass, R.; Yi, S. C.; Panek, M. E.; Wang, D.; Dalmonte, P.; Bond, M. D. *J. Pharm. Sci.* **2009**, *98*, 3117–3130.
149. Reubsæet, J. L.; Beijnen, J. H.; Bult, A.; van Maanen, R. J.; Marchal, J. A.; Underberg, W. J. *J. Pharm. Biomed. Anal.* **1998**, *17*, 979–984.
150. Yang, J.; Wang, S.; Liu, J.; Raghani, A. *J. Chromatogr. A* **2007**, *1156*, 174–182.
151. Hensel, M.; Steurer, R.; Fichtl, J.; Elger, C.; Wedekind, F.; Petzold, A.; Schlothauer, T.; Molhoj, M.; Reusch, D.; Bulau, P. *PLoS One* **2011**, *6*, e17708.
152. Mouls, L.; Silajdzic, E.; Haroune, N.; Spickett, C. M.; Pitt, A. R. *Proteomics* **2009**, *9*, 1617–1631.
153. Todorovski, T.; Fedorova, M.; Hoffmann, R. *J. Mass Spectrom.* **2011**, *46*, 1030–1038.
154. Van de Weert, M.; Lagerwerf, F. M.; Haverkamp, J.; Heerma, W. *J. Mass Spectrom.* **1998**, *33*, 884–891.
155. Perdivara, I.; Deterding, L. J.; Przybylski, M.; Tomer, K. B. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1114–1117.
156. Maillard, L. C. *Compte Rendu* **1912**, *154*, 66–68.
157. 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.
158. Quan, C. P.; Wu, S.; Dasovich, N.; Hsu, C.; Patapoff, T.; Canova-Davis, E. *Anal. Chem.* **1999**, *71*, 4445–4454.
159. Iberg, N.; Fluckiger, R. *J. Biol. Chem.* **1986**, *261*, 13542–13545.
160. Shapiro, R.; McManus, M. J.; Zalut, C.; Bunn, H. F. *J. Biol. Chem.* **1980**, *255*, 3120–3127.
161. Kuhn, R.; Weygand, F. *Ber. Dtsch. Chem. Ges. A/B* **1937**, *70*, 769–772.
162. Amadori, M. *Atti R. Accad. Naz. Lincei* **1929**, *9*, 226–230.
163. Amadori, M. *Atti R. Accad. Naz. Lincei* **1929**, *9*, 68–73.
164. Takatsy, A.; Boddi, K.; Nagy, L.; Nagy, G.; Szabo, S.; Marko, L.; Wittmann, I.; Ohmacht, R.; Ringer, T.; Bonn, G. K.; Gjerde, D.; Szabo, Z. *Anal. Biochem.* **2009**, *393*, 8–22.

165. Miller, A. K.; Hambly, D. M.; Kerwin, B. A.; Treuheit, M. J.; Gadgil, H. S. *J. Pharm. Sci.* **2011**, *100*, 2543–2550.
166. Chumsae, C.; Gifford, K.; Lian, W.; Liu, H.; Radziejewski, C. H.; Zhou, Z. *S. Anal. Chem.* **2013**, *85*, 11401–11409.
167. Iijima, K.; Murata, M.; Takahara, H.; Irie, S.; Fujimoto, D. *Biochem. J.* **2000**, *347* (Pt 1), 23–27.
168. Kislinger, T.; Humeny, A.; Peich, C. C.; Becker, C. M.; Pischetsrieder, M. *Ann. N. Y. Acad. Sci.* **2005**, *1043*, 249–259.
169. Frolov, A.; Schmidt, R.; Spiller, S.; Greifenhagen, U.; Hoffmann, R. *J. Agric. Food Chem.* **2014**, *62*, 3626–3635.
170. Zhang, Y.; Cocklin, R. R.; Bidasee, K. R.; Wang, M. *J. Biomol. Tech.* **2003**, *14*, 224–230.
171. Zeng, J.; Davies, M. J. *Chem. Res. Toxicol.* **2005**, *18*, 1232–1241.
172. Mostafa, A. A.; Randell, E. W.; Vasdev, S. C.; Gill, V. D.; Han, Y.; Gadag, V.; Raouf, A. A.; El Said, H. *Mol. Cell. Biochem.* **2007**, *302*, 35–42.
173. Haberger, M.; Bomans, K.; Diepold, K.; Hook, M.; Gassner, J.; Schlothauer, T.; Zwick, A.; Spick, C.; Kepert, J. F.; Hienz, B.; Wiedmann, M.; Beck, H.; Metzger, P.; Molhoj, M.; Knoblich, C.; Grauschopf, U.; Reusch, D.; Bulau, P. *mAbs* **2014**, *6*, 327–339.
174. Quan, C.; Alcalá, E.; Petkovska, I.; Matthews, D.; Canova-Davis, E.; Taticek, R.; Ma, S. *Anal. Biochem.* **2008**, *373*, 179–191.
175. Zhang, J.; Zhang, T.; Jiang, L.; Hewitt, D.; Huang, Y.; Kao, Y. H.; Katta, V. *Anal. Chem.* **2012**, *84*, 2313–2320.
176. Brady, L. J.; Martinez, T.; Balland, A. *Anal. Chem.* **2007**, *79*, 9403–9413.
177. Yuk, I. H.; Zhang, B.; Yang, Y.; Dutina, G.; Leach, K. D.; Vijayasankaran, N.; Shen, A. Y.; Andersen, D. C.; Snedecor, B. R.; Joly, J. C. *Biotechnol. Bioeng.* **2011**, *108*, 2600–2610.
178. Liu, H. *Study of Glycation and Advanced Glycation on a Humanized Monoclonal Antibody*; Dissertation, University of Rhode Island, 2013.
179. Goetze, A. M.; Liu, Y. D.; Arroll, T.; Chu, L.; Flynn, G. C. *Glycobiology* **2012**, *22*, 221–234.
180. Harris, R. J.; Shire, S. J.; Winter, C. *Drug Dev. Res.* **2004**, *61*, 137–154.
181. Shilton, B. H.; Walton, D. J. *J. Biol. Chem.* **1991**, *266*, 5587–5592.
182. Watkins, N. G.; Thorpe, S. R.; Baynes, J. W. *J. Biol. Chem.* **1985**, *260*, 10629–10636.
183. Nacharaju, P.; Acharya, A. S. *Biochemistry* **1992**, *31*, 12673–12679.
184. Gadgil, H. S.; Bondarenko, P. V.; Pipes, G.; Rehder, D.; McAuley, A.; Perico, N.; Dillon, T.; Ricci, M.; Treuheit, M. *J. Pharm. Sci.* **2007**, *96*, 2607–2621.
185. Kaschak, T.; Boyd, D.; Yan, B. *Anal. Biochem.* **2011**, *417*, 256–263.
186. Zhang, Q.; Tang, N.; Brock, J. W.; Mottaz, H. M.; Ames, J. M.; Baynes, J. W.; Smith, R. D.; Metz, T. O. *J. Proteome Res.* **2007**, *6*, 2323–2330.
187. Zhang, Q.; Tang, N.; Schepmoes, A. A.; Phillips, L. S.; Smith, R. D.; Metz, T. O. *J. Proteome Res.* **2008**, *7*, 2025–2032.
188. Kotia, R. B.; Raghani, A. R. *Anal. Biochem.* **2010**, *399*, 190–195.
189. Arena, S.; Salzano, A. M.; Renzone, G.; D’Ambrosio, C.; Scaloni, A. *Mass Spectrom. Rev.* **2014**, *33*, 49–77.

190. Gadgil, H. S.; Bondarenko, P. V.; Treuheit, M. J.; Ren, D. *Anal. Chem.* **2007**, *79*, 5991–5999.
191. Kijewska, M.; Kuc, A.; Kluczyk, A.; Waliczek, M.; Man-Kupisinska, A.; Lukaszewicz, J.; Stefanowicz, P.; Szewczuk, Z. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 966–976.
192. Arena, S.; Renzone, G.; Novi, G.; Paffetti, A.; Bernardini, G.; Santucci, A.; Scaloni, A. *Proteomics* **2010**, *10*, 3414–3434.
193. Zhang, Q.; Frolov, A.; Tang, N.; Hoffmann, R.; van de Goor, T.; Metz, T. O.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 661–666.
194. Lapolla, A.; Fedele, D.; Reitano, R.; Arico, N. C.; Seraglia, R.; Traldi, P.; Marotta, E.; Tonani, R. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 496–509.
195. Zhang, Z. *Anal. Chem.* **2009**, *81*, 8354–8364.
196. Humeny, A.; Kislinger, T.; Becker, C. M.; Pischetsrieder, M. *J. Agric. Food Chem.* **2002**, *50*, 2153–2160.
197. Liu, H.; Ponniah, G.; Neill, A.; Patel, R.; Andrien, B. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2014**, *958*, 90–95.
198. 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.
199. Meadows, L.; Wang, W.; den Haan, J. M.; Blokland, E.; Reinhardus, C.; Drijfhout, J. W.; Shabanowitz, J.; Pierce, R.; Agulnik, A. I.; Bishop, C. E.; Hunt, D. F.; Goulmy, E.; Engelhard, V. H. *Immunity* **1997**, *6*, 273–281.
200. Pierce, R. A.; Field, E. D.; den Haan, J. M.; Caldwell, J. A.; White, F. M.; Marto, J. A.; Wang, W.; Frost, L. M.; Blokland, E.; Reinhardus, C.; Shabanowitz, J.; Hunt, D. F.; Goulmy, E.; Engelhard, V. H. *J. Immunol.* **1999**, *163*, 6360–6364.
201. Alexander, A. J.; Hughes, D. E. *Anal. Chem.* **1995**, *67*, 3626–3632.
202. Cohen, S. L.; Price, C.; Vlasak, J. *J. Am. Chem. Soc.* **2007**, *129*, 6976–6977.
203. Jiskoot, W.; Beuvery, E. C.; de Koning, A. A.; Herron, J. N.; Crommelin, D. J. *Pharm. Res.* **1990**, *7*, 1234–1241.
204. Vlasak, J.; Ionescu, R. *mAbs* **2011**, *3*, 253–263.
205. Gu, S.; Wen, D.; Weinreb, P. H.; Sun, Y.; Zhang, L.; Foley, S. F.; Kshirsagar, R.; Evans, D.; Mi, S.; Meier, W.; Pepinsky, R. B. *Anal. Biochem.* **2010**, *400*, 89–98.
206. Kshirsagar, R.; McElearney, K.; Gilbert, A.; Sinacore, M.; Ryll, T. *Biotechnol. Bioeng.* **2012**, *109*, 2523–2532.
207. Li, N.; Fort, F.; Kessler, K.; Wang, W. *J. Pharm. Biomed. Anal.* **2009**, *50*, 73–78.
208. Gorres, K. L.; Raines, R. T. *Crit Rev Biochem Mol Biol* **2010**, *45*, 106–124.
209. Valtavaara, M.; Szpirer, C.; Szpirer, J.; Myllylä, R. *J. Biol. Chem.* **1998**, *273*, 12881–12886.
210. Zhou, M.; Gucinski, A.; Boyne, M. In *System Suitability Metrics for Analysis of Protein Therapeutics by LC-MS*, Proceedings of the 62nd ASMS Conference on Mass Spectrometry and Allied Topics, Baltimore, MD, June 15–19; Baltimore, MD, 2014; p MOE pm.
211. Banerjee, S.; Mazumdar, S. *Int. J. Anal. Chem.* **2012**, *2012*, 282574.

212. Yu, X. C.; Joe, K.; Zhang, Y.; Adriano, A.; Wang, Y.; Gazzano-Santoro, H.; Keck, R. G.; Deperalta, G.; Ling, V. *Anal. Chem.* **2011**, *83*, 5912–5919.
213. 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.
214. Rohrer, J. S.; Cooper, G. A.; Townsend, R. R. *Anal. Biochem.* **1993**, *212*, 7–16.
215. Rehder, D. S.; Dillon, T. M.; Pipes, G. D.; Bondarenko, P. V. *J. Chromatogr. A* **2006**, *1102*, 164–175.
216. Sagi, D.; Peter-Katalinic, J.; Conradt, H. S.; Nimtz, M. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1138–1148.
217. Kailemia, M. J.; Ruhaak, L. R.; Lebrilla, C. B.; Amster, I. J. *Anal. Chem.* **2014**, *86*, 196–212.
218. Shah, B.; Jiang, X. G.; Chen, L.; Zhang, Z. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 999–1011.
219. Wang, L.; Amphlett, G.; Lambert, J. M.; Blattler, W.; Zhang, W. *Pharm. Res.* **2005**, *22*, 1338–1349.
220. Lundell, N.; Schreitmuller, T. *Anal. Biochem.* **1999**, *266*, 31–47.
221. Stroop, S. D. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 830–836.
222. Formolo, T.; Heckert, A.; Phinney, K. W. *Anal. BioAnal. Chem.* **2014**, *406*, 6587–6598.
223. Lopez-Ferrer, D.; Capelo, J. L.; Vazquez, J. *J. Proteome Res.* **2005**, *4*, 1569–1574.
224. Lopez-Ferrer, D.; Petritis, K.; Hixson, K. K.; Heibeck, T. H.; Moore, R. J.; Belov, M. E.; Camp, D. G., 2nd; Smith, R. D. *J. Proteome Res.* **2008**, *7*, 3276–3281.
225. Srzentic, K.; Fornelli, L.; Laskay, U. A.; Monod, M.; Beck, A.; Ayoub, D.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86*, 9945–9953.
226. Dasari, S.; Wilmarth, P. A.; Reddy, A. P.; Robertson, L. J.; Nagalla, S. R.; David, L. L. *J. Proteome Res.* **2009**, *8*, 1263–1270.
227. Sargaeva, N. P.; Goloborodko, A. A.; O'Connor, P. B.; Moskovets, E.; Gorshkov, M. V. *Electrophoresis* **2011**, *32*, 1962–1969.
228. Zang, L.; Carlage, T.; Murphy, D.; Frenkel, R.; Bryngelson, P.; Madsen, M.; Lyubarskaya, Y. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2012**, *895–896*, 71–76.
229. Liu, H.; Ponniah, G.; Neill, A.; Patel, R.; Andrien, B. *Anal. Chem.* **2013**, *85*, 11705–11709.
230. Stroop, S. D.; Conca, D. M.; Lundgard, R. P.; Renz, M. E.; Peabody, L. M.; Leigh, S. D. *J. Pharm. Sci.* **2011**, *100*, 5142–5155.
231. Stability Testing of New Drug Substances and Products, Q1a(R2). In *ICH Harmonised Tripartite Guideline*, Step 4 version ed.; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use: Geneva, Switzerland, 2003.
232. Hawe, A.; Wiggenhorn, M.; van de Weert, M.; Garbe, J. H.; Mahler, H. C.; Jiskoot, W. *J. Pharm. Sci.* **2012**, *101*, 895–913.
233. Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products, Q5c. In *ICH Harmonised Tripartite Guideline*, Step 4 version ed.; International Conference on Harmonisation of Technical



Requirements for Registration of Pharmaceuticals for Human Use: Geneva, Switzerland, 1995.

234. Cordoba, A. J.; Shyong, B. J.; Breen, D.; Harris, R. J. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *818*, 115–121.
235. Gao, X.; Ji, J. A.; Veeravalli, K.; Wang, Y. J.; Zhang, T.; McGreevy, W.; Zheng, K.; Kelley, R. F.; Laird, M. W.; Liu, J.; Cromwell, M. *J. Pharm. Sci.* **2015**, *104*, 368–377.

## Chapter 4

# Orthogonal Technologies for NISTmAb N-Glycan Structure Elucidation and Quantitation

**Justin M. Prien,<sup>1</sup> Henning Stöckmann,<sup>2</sup> Simone Albrecht,<sup>2</sup>  
Silvia M. Martin,<sup>2</sup> Matthew Varatta,<sup>1</sup> Marsha Furtado,<sup>1</sup>  
Stephen Hosselet,<sup>1</sup> Meiyao Wang,<sup>3</sup> Trina Formolo,<sup>3</sup>  
Pauline M. Rudd,<sup>2</sup> and John E. Schiel<sup>\*3</sup>**

<sup>1</sup>Analytical Sciences, Amgen Inc., 40 Technology Way,  
West Greenwich, Rhode Island 02817, United States

<sup>2</sup>NIBRT GlycoScience Group, NIBRT (The National Institute for  
Bioprocessing Research and Training), Fosters Avenue, Mount Merrion,  
Blackrock, Co. Dublin, Ireland

<sup>3</sup>Biomolecular Measurement Division, National Institute of Standards and  
Technology, 100 Bureau Dr., Stop 8314,  
Gaithersburg, Maryland 20899, United States

\*E-mail:john.schiel@nist.gov

N-linked glycosylation is a common post-translational modification that imparts structural heterogeneity to recombinant monoclonal antibody (mAb) therapeutics. The various oligosaccharides attached to the C<sub>H</sub>2 domains of an IgG can impact the efficacy, safety, pharmacokinetics, and pharmacodynamics of the product. Depending on the mechanism of action of the biotherapeutic, specific glycan moieties may be deemed critical quality attributes (CQAs) and require appropriate process control and analytical strategies to ensure product quality consistency. Glycan analysis presents a significant challenge because the N-glycosylation biosynthetic pathway results in a heterogeneous population of complex branched structures at each glycosylation site. Consequently, extensive sample handling is required for analysis. At various stages of developing and manufacturing of a mAb, several levels of glycan analysis are required: (i) profiling, which can

provide a quick comparative overview of the glycosylation; (ii) characterization of the glycans by type to provide particular CQAs; and (iii) full, detailed analysis, including monosaccharide sequence and linkage of the glycans, which usually requires orthogonal technologies. Glycoanalytical standards are of critical importance in evaluating the suitability of methods intended to monitor glycosylation through the lifecycle of the product. The intact NISTmAb reference material may provide an additional supplement to in-house reference standards for system suitability, operator training, and analytical method evaluation. The NISTmAb glycan population was therefore analyzed using several orthogonal approaches. The comprehensive workstreams presented here demonstrate state-of-the art technologies that are suitable for sequencing the glycans attached to mAbs throughout the lifecycle of the product.

## Introduction

N-linked glycosylation is a common post-translational modification that imparts structural heterogeneity to each of the C<sub>H</sub>2 domains in the Fc region of both native IgG and recombinant monoclonal antibody (mAb) therapeutics (Mechanism Action Ch.) (1). N-linked glycans are modifications of an asparagine (Asn) residue side chain at the consensus sequence Asn-Xaa-Ser/Thr, where Xaa can be any amino acid except proline. Glycan processing is regulated by a variety of stepwise removals and additions of monosaccharide residues to the dolichol phosphate pre-cursor (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>) in the endoplasmic reticulum and Golgi apparatus. These enzyme actions result in complex, hybrid, or high mannose glycans that all share the common trimannosyl core sequence (Man<sub>3</sub>GlcNAc<sub>2</sub>) depicted in (Figure 1) (2). The most common class of N-glycans attached to mAbs is that of the complex bi-antennary type; however, all classes have been detected at some level. In Figure 1, a representative N-glycan from each class according to the Consortium for Functional Glycomics (CFG) (3) system, as well as the Oxford system (4), is depicted. Both methods provide a relatively simple mechanism for conveying individual monosaccharide residues via shapes, and their linkages based on additional text (CFG) or linkage orientation/line style (Oxford).

N-linked glycosylation can impact the efficacy, safety, pharmacokinetics, and pharmacodynamics of recombinant mAbs by modulating the effector functions of the Fc region (see Mechanism Action Chapt. for additional discussion). For example, afucosylated complex glycans support antibody-dependent cell-mediated cytotoxicity (ADCC) (5, 6) by modulating FcγR binding. Glycans carrying terminal galactose or GlcNAc residues affect complement-dependent cytotoxicity (CDC) by modulating C1q and mannose-binding lectin, respectively (7–9), whereas the presence of high mannose glycans can lead to increased drug clearance *in vivo* (10). These critical effects on pharmacological behavior

have resulted in efforts to engineer more tailored glycoprofiles through molecular engineering of biosynthetic pathways and optimization of bioreactor process parameters (11, 12). For both glyco-engineered and non-engineered biotherapeutic candidates, it is the intent to produce product with consistent glycosylation. Depending on the mechanism of action of the biotherapeutic, specific glycan moieties may also be deemed critical quality attributes and require an appropriate process control strategy to ensure product quality consistency (13). Robust product quality analytical methods to monitor N-linked glycosylation are therefore essential.

Mass spectrometry (MS), high-performance liquid chromatography (HPLC), capillary electrophoresis, and/or hyphenated combinations of these techniques (LC-MS) are commonly used for studying protein glycosylation (13–26). High-resolution MS has become a popular tool to match medium-to-high abundance mAb glycoprotein compositions directly with intact mAbs or subunits/fragments thereof, as demonstrated in the Primary Structure chapter/Volume 2, Chapter 1. However, comprehensive identification and characterization is achieved through release of the N-glycan population followed by directed analysis. Removal of mammalian N-glycan species is most commonly achieved using peptide N-glycosidase F (PNGase F), an enzyme that selectively cleaves between the Asn residue and the proximal GlcNAc (both  $\alpha$ -[1-6]-core fucosylated and afucosylated) of high mannose, hybrid, and complex oligosaccharides (27) commonly found on mAbs. Plant-derived antibodies require the use of peptide N-glycosidase A to release glycans containing  $\alpha$ -(1-3)-linked core fucose, and endoglycosidase enzymes can be used to release specific glycan moieties (28–30). Released glycans have a free reducing terminus that can be further modified prior to analysis with labels to improved detectability, ionization efficiency, and/or chromatographic retention. Recent advancements in column resin chemistries, ultra-performance liquid chromatography (UPLC) systems, highly sensitive fluorescence detectors, and state-of-the-art mass spectrometers increasingly provide the means to characterize released and labeled glycan species rapidly. MS yields confident, rapid identification of glycan composition, whereas sequence and linkage information can be achieved through a variety of sample preparation and fragmentation techniques that result in both glycosidic and cross-ring fragmented product ions, as depicted for the core structure in Figure 1. High-resolution LC technologies using high-sensitivity fluorescent labels are quantitative. Combined with exoglycosidase array digestions, LC provides monosaccharide sequence and linkage information as well. The LC may be used either coupled or uncoupled to MS for further structural information (19).

The current chapter uses the NISTmAb reference material as a surrogate to demonstrate a variety of glycoanalytical workflows in accordance with their intended purpose throughout the commercialization lifecycle of a biotherapeutic. The glycoanalytical techniques discussed can also be applied to other protein therapeutic classes as well as during the development of follow-on (biosimilar) products (19, 31). First, we discuss an automated, high-throughput glycan analytical method suitable for process support. In addition to monitoring product quality attributes for process support, comprehensive product characterization is a requirement for regulatory filing. Therefore, a comprehensive N-linked glycan

analysis workflow purposed for biological license application of monoclonal antibodies is outlined in this chapter. A common aspect to both workflows, as well as being widespread throughout industry, is the fluorescent labeling and analysis of released N-glycans. A separate section has therefore been dedicated to highlighting various experimental factors that may be considered when implementing this technique. Finally, an interlaboratory comparison of each laboratory's platform LC-based glycoanalytical mapping results is presented.

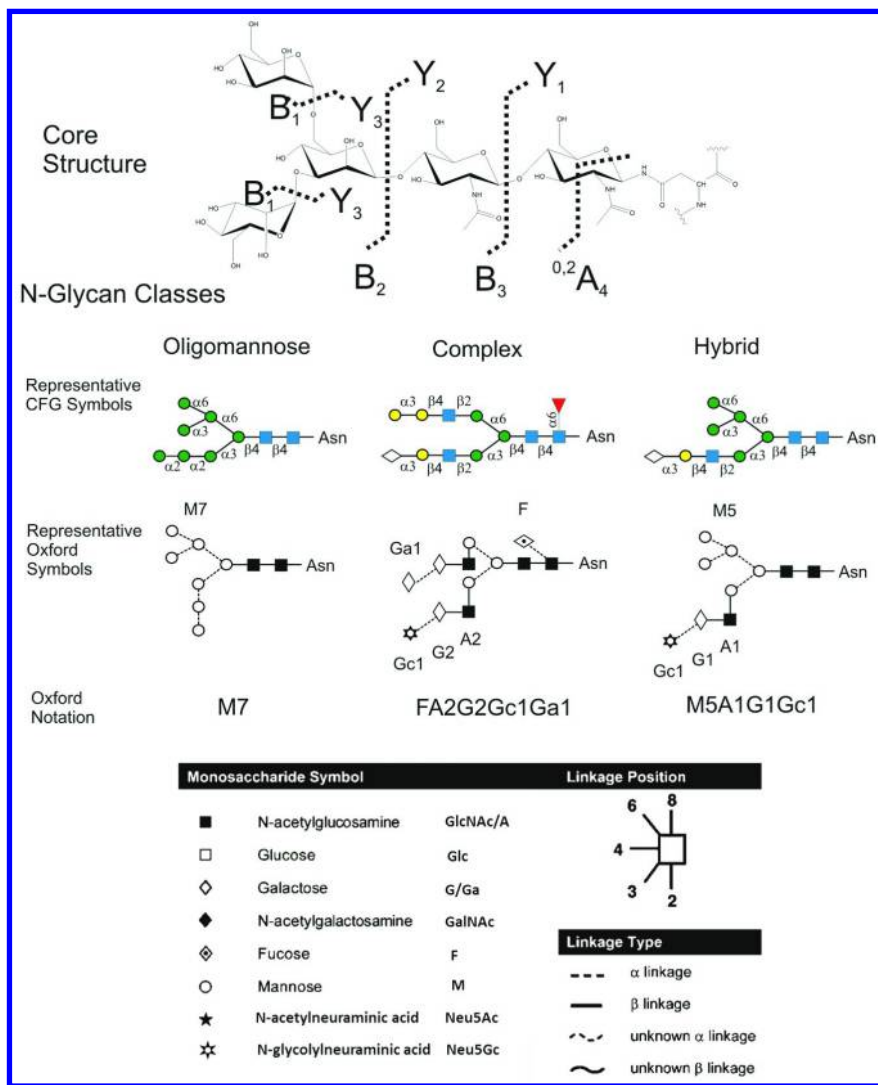


Figure 1. Core chemical structure (with representative glycosidic and cross-ring fragments) and representative classes of N-linked glycans. Linkage annotations are made according to putative biosynthetic pathways. Reference table is included describing the Oxford Nomenclature Symbols (4). (see color insert)

## High-Throughput N-linked Glycan Profiling for Process Support

During process optimization campaigns, operational parameters are intentionally varied to determine the effect on product quality attributes and process performance. To complete these studies, high volumes of test samples are submitted for analysis to the analytical testing groups. Automated, high-throughput, qualified product quality and purity assays are ideal to support process optimization activities. Fully humanized monoclonal antibodies contain relatively straightforward N-linked glycan populations, predominantly asialylated, core-fucosylated, bi-antennary glycans possessing 0, 1, or 2 terminal galactose residues and modest levels of hybrid and high mannose glycans. Sialylated glycans are typically in low abundance in recombinant IgG1 and IgG2 modalities. The collection of N-linked glycan species is relatively constant, albeit the relative proportions of the glycan species may differ significantly. The consistent population of glycan species justifies the development of a platform glycan profiling assay, commonly referred to as a glycan map, to support a mAb biotherapeutic pipeline. Recently, the National Institute for Bioprocessing Research and Training (NIBRT) introduced an automated, high-throughput glycomics platform that is low-cost, reliable, and suitable as a platform for glycan mapping in biotherapeutics development (Figure 2).

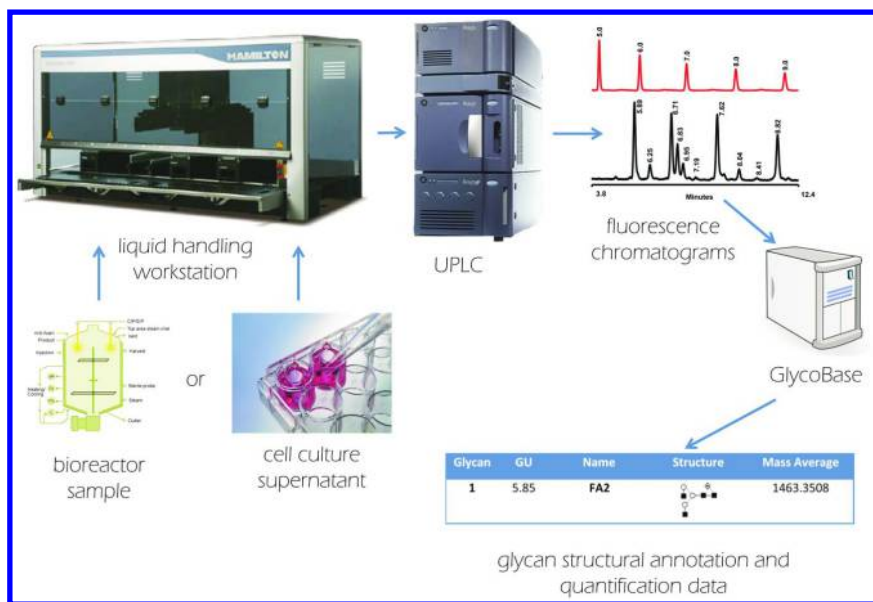


Figure 2. National Institute for Bioprocessing Research and Training (NIBRT) glycomics platform, consisting of an automated glycan sample preparation module, glycan analysis technology and bioinformatics tools for glycan structure elucidation. (see color insert)

Briefly, the glycoanalysis platform consists of sample preparation using a liquid-handling robot in a 96-well or 384-well format (Appendix A) followed by hydrophilic interaction liquid chromatography (HILIC) separation with fluorescence detection and use of a reference database, GlycoBase ([https://glycibase.nibrt.ie/glycibase/show\\_nibrt.action](https://glycibase.nibrt.ie/glycibase/show_nibrt.action)), to facilitate rapid peak assignment (26, 32). Some key features of the robotic sample preparation consist of rapid IgG affinity purification and sample concentration, protein denaturation and enzymatic glycan release on a multiwell filtration device, glycan purification on a solid hydrazide support, and fluorescent labeling with 2-aminobenzamide (2-AB), allowing the preparation of 96 samples in 22 hours (Figure 2). The fluorescently labeled glycans are analyzed on UPLC instruments equipped with a HILIC BEH130 column (150 mm × 2.1 mm, 1.7 μm).

GlycoBase is a database for the storage, classification, and visualization of glycan structures and their associated experimental parameters obtained from various analytical techniques such as HPLC, UPLC, and capillary electrophoresis (CE). GlycoBase is an open-access resource that contains glycan data normalized to chromatographic retention time data, expressed as Glucose Unit (GU) values, for more than 700 unique glycan structures (labeled with 9-aminopyrene-1,4,6-trisulfonic acid [APTS] and/or 2-AB). These values were experimentally obtained by detailed analysis of released glycans from a diverse set of glycoproteins. A dedicated biotherapeutics sub-collection is available through the Waters UNIFI 1.7 platform, which contains GU values for oligosaccharides released from therapeutically relevant glycoproteins such as Erythropoietin, Herceptin, and Infliximab. HILIC combined with fluorescence detection, exoglycosidase sequencing, and mass spectrometry were used to generate this high-confidence glycan library. The resulting database has been made accessible through a customized Web application containing a simple graphical user interface for assignment and confirmation of glycan structures (32, 33).

Figure 3 represents the 2-AB-labeled, N-linked glycan profile of the NISTmAb prepared on the NIBRT-automated platform. Chromatographic peaks are labeled with proposed glycan species assigned utilizing the GlycoBase database and serial exoglycosidase treatment. Table 1 represents 2-AB-labeled glycan species that are ≥ 0.1% of the total peak area based on GlycoBase assignment. When utilizing GlycoBase, a dextran ladder should be incorporated into the sample sequence, allowing the assignment of a GU value to each of the peaks and assisting in peak annotation. The use of standard GUs renders these values independent of the instrument and running conditions, which allows for the direct comparison of the normalized peak GU unit to the curated GlycoBase reference database. For example, the first peak in the NISTmAb glycan map has an associated GU value of  $4.74 \pm 0.05$  as determined from three independent UPLC analyses (Figure 3 and Table 1). A GlycoBase search for a GU value range of 4.69 to 4.79 returns the glycans A1 and FM3 (Figure 4). Using automated sample preparation, standardized chromatographic conditions, and GlycoBase for peak annotation and data storage, a single glycan map may be used to rapidly identify and quantify glycan species for the purpose of monitoring process consistency.

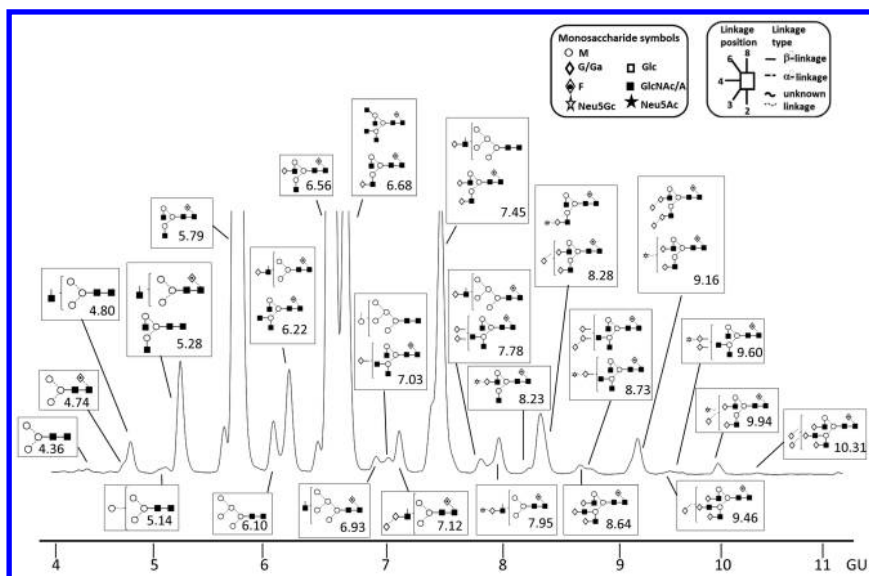


Figure 3. Hydrophilic interaction liquid chromatography (HILIC) fluorescence chromatogram of the NIST monoclonal antibody glycans from automated sample preparation and assignment with normalized GU units.

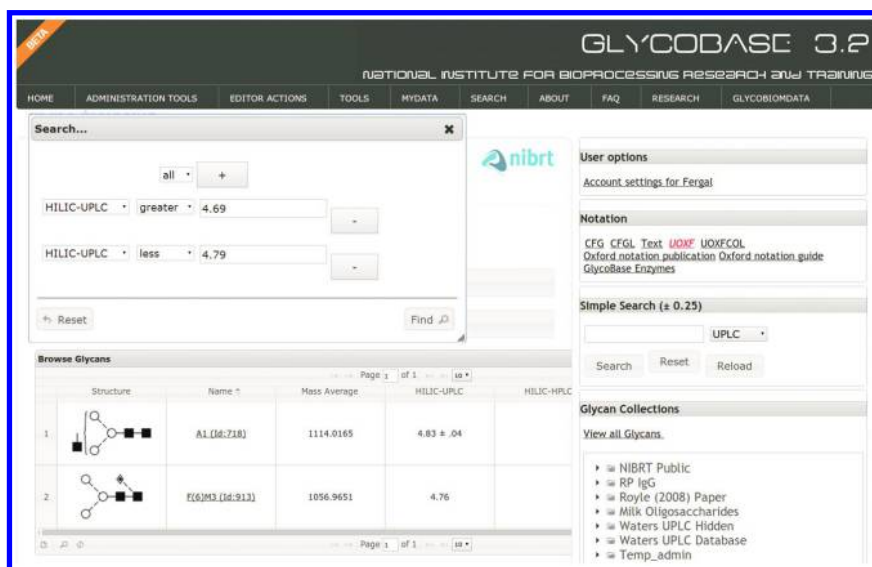


Figure 4. GlycoBase Web interface and GlycoBase search for a representative NISTmAb glycan peak with a glucose unit (GU) value range between 4.69–4.79, returning the glycans A1 and FM3. (see color insert)



**Table 1. Hydrophilic Interaction Liquid Chromatography (HILIC) Glycan Peaks  $\geq 0.1\%$  of the Total Peak Area with Assignment Based on GlycoBase**

Peak Assignment	Empirical Glucose Units		Theoretical Glucose Units		Retention Time			Relative Peak Area		
	Value	SD	Value	Relative % Difference	AV (min)	SD	CV	AV (%)	SD	CV
<b>FM3</b>	4.74	0.05	4.76	0.35	3.48	0.06	1.63	0.12	0.01	8.96
<b>A1</b>	4.80	0.05	4.83	0.62	3.55	0.06	1.63	0.70	0.04	4.98
<b>M4</b>	5.14	0.05	5.11	0.52	4.00	0.07	1.68	0.20	0.03	14.21
<b>FA1</b> <b>A2</b>	5.28	0.06	5.27 5.42	0.13 2.64	4.18	0.07	1.70	2.58	0.06	2.37
<b>FA2</b>	5.79	0.06	5.87	1.42	4.91	0.08	1.67	39.81	0.31	0.77
<b>M5</b>	6.10	0.06	6.19	1.45	5.39	0.09	1.60	1.17	0.06	4.99
<b>FA1G1</b> <b>FA3</b>	6.22	0.06	6.23 6.24	0.11 0.27	5.57	0.10	1.73	2.57	0.10	4.03
<b>FA2G1-a</b>	6.56	0.07	6.70	2.04	6.11	0.10	1.66	28.05	0.23	0.82
<b>FA2G1-b</b> <b>FA4</b>	6.68	0.07	6.80 6.95	1.86 3.84	6.30	0.10	1.66	10.10	0.07	0.66
<b>FM4A1G1</b>	6.93	0.07	7.04	1.61	6.70	0.11	1.61	0.41	0.05	11.89
<b>M6</b> <b>FA3G1</b>	7.03	0.07	7.02 ND	0.09 –	6.86	0.11	1.61	0.41	0.03	8.05
<b>FA1G1Ga1</b>	7.12	0.08	7.11	0.09	7.00	0.12	1.66	1.00	0.05	5.27

Peak Assignment	Empirical Glucose Units		Theoretical Glucose Units		Retention Time			Relative Peak Area		
	Value	SD	Value	Relative % Difference	AV (min)	SD	CV	AV (%)	SD	CV
<b>FA2G2</b> <b>M5A1G1</b>	7.45	0.08	7.60 7.54	1.97 1.19	7.53	0.12	1.62	7.55	0.24	3.22
<b>FM5A1G1</b> <b>FA3G2</b>	7.78	0.09	7.95 ND	2.10 –	8.05	0.13	1.58	0.36	0.02	5.65
<b>FA1G1Gc1</b>	7.95	0.08	7.98	0.33	8.32	0.10	1.19	0.83	0.09	10.88
<b>FA2G1Gc1-a</b>	8.23	0.10	8.37	1.67	8.73	0.13	1.49	0.14	0.03	21.30
<b>FA2G1Gc1-b</b> <b>FA2G2Ga1</b>	8.28	0.10	8.38 8.4	1.15 1.15	8.82	0.14	1.55	1.85	0.05	2.96
<b>FA3G3</b>	8.64	0.11	8.61	0.31	9.34	0.14	1.50	0.27	0.02	7.32
<b>FA3G2Ga1</b> <b>FA3G1Gc1</b>	8.73	0.11	ND ND	– –	9.48	0.12	1.24	0.15	0.02	11.13
<b>FA2G2Ga2</b> <b>FA2G2Gc1</b>	9.16	0.12	9.33 9.17	1.82 0.11	10.09	0.15	1.45	1.11	0.03	2.92
<b>FA3G3Ga1</b>	9.46	0.12	ND	–	10.52	0.14	1.34	0.12	0.01	8.21
<b>FA2G2Ga1Gc1</b>	9.94	0.12	10.03	0.86	11.17	0.12	1.12	0.27	0.02	8.75
<b>**FA3G2Gc1</b>	9.60	0.11	ND	–	10.71	0.12	1.09	0.03	0.02	47.17
<b>**FA3G3Ga2</b>	10.31	0.14	10.40	0.87	11.65	0.16	1.36	0.04	0.02	41.42

AV = average, CV = coefficient of variance, SD = standard deviation. \*\* Indicates species identified as separate peaks with < 0.1% abundance. ND represents glycans for which theoretical GU units were not available.

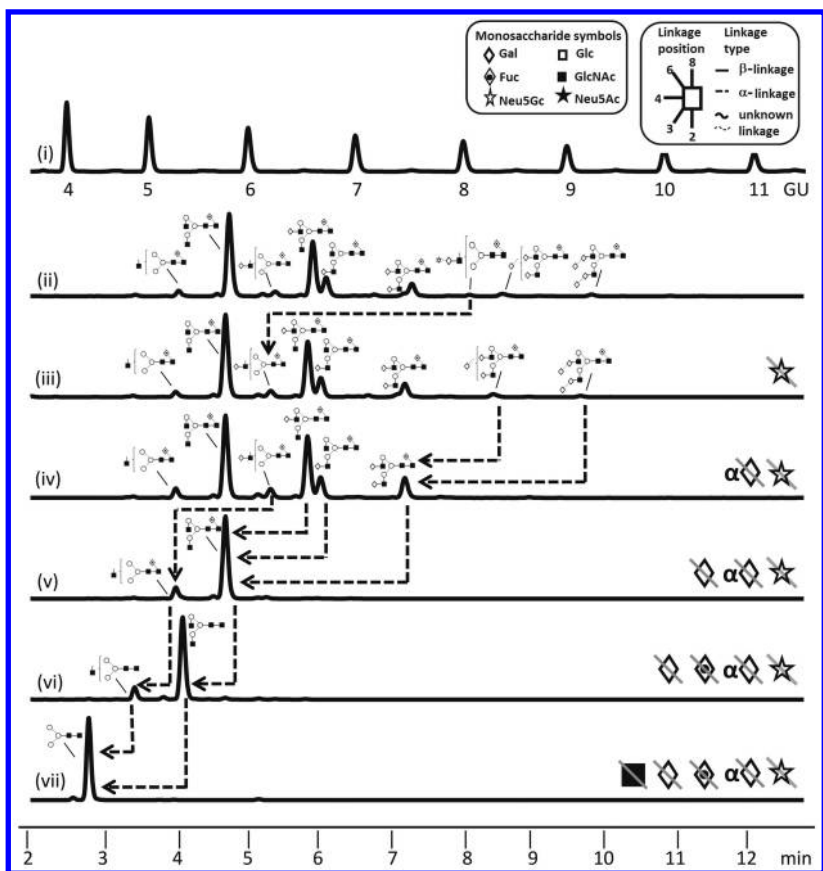


Figure 5. Glycan sequencing panel of 2-aminobenzamide (2-AB)-labeled NIST monoclonal antibody (mAb) glycans analyzed by ultra-performance liquid chromatography (UPLC) systems with fluorescence detection and dextran ladder for GU definition. (i) Dextran ladder; (ii) undigested glycans; (iii) ABS (*Athrobacter ureafaciens* sialidase) releases  $\alpha$ -(2-3,6,8)-linked sialic acids; (iv) CBG (green coffee bean  $\alpha$ -galactosidase) releases  $\alpha$ -(1-3,4,6)-galactose; (v) BTG (bovine testes  $\beta$ -galactosidase) releases  $\beta$ -(1-3) and  $\beta$ -(1-4)-linked galactose residues; (vi) BKF (bovine kidney fucosidase) releases  $\alpha$ -(1-2,6)-linked fucose; (vii) GUH (*Streptococcus pneumoniae* hexosaminidase) releases  $\beta$ -linked N-acetylglucosamine. Dotted arrows indicate glycan peak shifts upon exoglycosidase treatment. The proposed structural annotation for selected peaks are shown.

An important component in the development of a robust platform glycan map is peak assignment. For high confidence peak assignment, a combination of glycan sequencing through serial exoglycosidase digestion and acquiring GU data to index the elution position of the sequenced glycan is recommended. Serial or parallel exoglycosidase array digestions are a robust analytical approach

for glycan species assignment, including linkage and monosaccharide sequence determination. Exoglycosidase treatment can therefore assist in determining the identities of co-eluting species when they occur. A full understanding of co-eluting glycans under a given HILIC peak is important to monitor the changes in N-glycosylation that result from planned or unplanned perturbations in the manufacturing process. For the purpose of peak assignment, the entire population of glycans can be digested without the need for pre-fractionation of individual peaks. Exoglycosidase digestion results in characteristic peak shifts whose magnitude depends on the nature and the number of monosaccharides removed.

Figure 5 shows the complete exoglycosidase digestion scheme for the NISTmAb (34). Treatment of the glycans with sialidase resulted in the disappearance of the peak at GU = 8.01 and a corresponding increase in the peak at GU = 6.22. The observed GU-shift of approximately 1.8 is indicative of the loss of N-glycolylneuraminic acid (Neu5Gc), so the peak at GU = 8.01 must be a glycan with one terminal Neu5Gc (Figure 5, trace (ii) to (iii), shift marked with a dotted arrow). The glycan pool obtained after sialidase treatment was then sequentially digested with  $\alpha$ -galactosidase, fucosidase, galactosidase, and hexosaminidase, resulting in one single peak at GU = 4.38, consistent with the elution position of the putative N-linked core. The detailed analysis of all GU shifts allows for rapid annotation of the glycan peaks from the NISTmAb. Through utilization of the GlycoBase database and serial exoglycosidase digestion, all peaks above 0.5% of the total peak area were assigned as a specific glycan species (Figure 3 and Table 1).

## Qualification of Glycan Assays for Process Optimization Support

International Conference on Harmonisation (ICH) guideline Q2(R1) discusses performance parameters that should be considered during the validation of analytical methods used for commercial product release (35); however, specific ICH guidelines do not exist for qualifying an analytical method used during the development phases of a molecule's lifecycle. Nonetheless, it is an industry expectation that product quality assays are qualified to demonstrate acceptable method performance for their intended purpose consistent with the product's phase of development.

For the purposes of this chapter, a complete method qualification was not performed; however, repeatability was evaluated. NISTmAb samples were prepared and analyzed on 6 separate days using the NIBRT glycomics platform and repeatability was assessed with regard to relative peak percentage areas, retention time, and GU values. Each of the qualification target acceptance criteria for repeatability were met, as demonstrated in Table 2. This limited qualification demonstrates the repeatability of an automated glycan mapping platform that is suitable for monitoring N-linked glycosylation for process support.

**Table 2. Performance Characteristics for NIST Monoclonal Antibody (mAb) Sample Run on Automated Glycomics Platform**

<i>Performance Characteristic</i>	<i>Qualification Target</i>	<i>Qualification Result</i>
<b>Precision- Repeatability</b>	Relative peak area standard deviation $\leq 0.4\%$ for all identified peaks $\geq 0.5\%$ of the total peak area	Standard deviation $\leq 0.4\%$ for all identified peaks $\geq 0.5\%$ of the total peak area, except the FM4A1G1 peak
	Relative peak area coefficient of variance $\leq 10.0\%$ for all peaks $\geq 0.5\%$ of the total peak area	Coefficient of variance $\leq 10.0\%$ for all peaks $\geq 0.5\%$ of the total peak area, except the FM4A1G1 peak
	Retention time coefficient of variance $\leq 5\%$ for all peaks $\geq 0.5\%$ of the total peak area	Coefficient of variance $\leq 5\%$ for all peaks $\geq 0.5\%$ of the total peak area
<b>Peak Assignment Accuracy</b>	GU standard deviation $\leq 0.4\%$ for identified peaks $\geq 0.5\%$ of the total peak area	Standard deviation $\leq 0.4\%$ for all peaks for identified peaks $\geq 0.5\%$ of the total peak area
	Relative percent difference from theoretical GU unit $\leq 5\%$ for all species	Relative percent difference from theoretical GU units $\leq 5\%$ for all peaks

Through method qualification, appropriate system suitability and assay acceptance criteria are typically established. To ensure that assay performance criteria are met, a standardized sample sequence is suggested. Table 3 highlights a suggested sample sequence used for running test samples to support process optimization. Two positive controls, a working in-house reference standard and a dextran ladder, are purposefully placed within the sequence. Multiple injections of the working in-house reference standard are dispersed throughout the sequence to monitor system suitability and assay performance. The in-house reference standard also serves as an intra-assay control against which test samples may be evaluated. The ability to detect unexpected peaks is an important component in the development of a quantitative purity test to monitor process consistency. A dextran ladder is incorporated into the sample sequence to aid in the identification of a new peak if one is detected above an established threshold. The NISTmAb reference material also may be incorporated into such a workflow as a platform system suitability metric. Such an external control would ensure that should deviations occur, both sample preparation and instrument performance conform to expectations for each experimental sequence and/or assist with troubleshooting various parameters.

**Table 3. Suggested Sample Sequence for Routine Sample Testing Using the Automated Glycomics Platform**

<i>Description</i>	<i>Number of Injections</i>
<b>Water Injection</b>	2
<b>Reagent Blank Injection</b>	1
<b>Dextran Ladder</b>	1
<b>Working In-House Reference standard (initial)</b>	3
<b>Samples 1 to 20</b>	1
<b>Working In-House Reference standard (bracketing)</b>	2
<b>Dextran Ladder</b>	1
<b>Samples 21 to 40</b>	1
<b>Working In-House Reference standard (bracketing end)</b>	2
<b>Dextran Ladder</b>	1

## **Detailed Glycan Characterization Workflow for Marketing Application**

Tremendous analytical resources are allocated to determining comprehensively the biochemical, biophysical, and biological function and structure of a recombinant mAb biotherapeutic. This expansive body of work is encapsulated within the elucidation of structure section of the marketing application for the product. N-linked glycosylation imparts structural heterogeneity to the two C<sub>H2</sub> domains in the Fc of mAbs. Due to the potential biological and pharmacokinetic impact that certain glycan moieties can exert on the mechanism of action and clearance of a product, the quantitation and sequencing of individual glycan species is necessary. In this section, we outline a detailed glycan characterization workflow purposed for data collection slated for the use in a marketing application.

Using the NISTmAb reference material as a surrogate mAb therapeutic, the following glycoanalytical workflows (for which detailed methods are described in Appendix B) will be discussed:

- Tryptic peptide mapping to identify and determine the occupancy level of the N-linked glycosylation consensus site.
- Species identification and quantitation using HILIC with fluorescence and on-line mass spectrometry detection.
- HILIC peak fractionation followed by second dimension reversed-phase liquid chromatography with fluorescence and on-line mass spectrometry for detection and quantitation of co-eluting species.

- Permethylation followed by sequential mass spectrometry (MS<sup>n</sup>) for structural confirmation and linkage analysis of each HILIC peak fraction.
- Strategic exoglycosidase treatment followed by HILIC with fluorescence and on-line mass spectrometry detection for orthogonal structural confirmation and linkage analysis.
- Weak anion exchange (WAX) chromatography followed by HILIC with fluorescence for identification and quantitation of low-abundance acidic glycan species.

## Site Determination and Occupancy

The NISTmAb underwent enzymatic deglycosylation using peptide N-Glycosidase F (PNGase F) followed by denaturation, reduction and alkylation, and subsequent digestion by trypsin. According to the NISTmAb theoretical heavy chain sequence, a single conserved N-glycosylation site, N<sup>300</sup>ST, exists in each of the C<sub>H</sub>2 domains. The tryptic peptide TH(296-304) (EEQYNSTYR) contains the only site with an N-glycosylation consensus sequence, Asn-Xaa-Ser/Thr. PNGase F is an amidase which cleaves high mannose, hybrid, and complex glycan moieties between the reducing terminal *N*-acetylglucosamine residue of the N-glycan and the side-chain of the asparagine residue in the peptide backbone (36). A byproduct of the enzymatic cleavage is the conversion of asparagine to aspartate, leading to a nominal mass shift of 1 Da in the deglycosylated TH(296-304) peptide (EEQYDSTYR). After PNGase F treatment, the TH(296-304) peptide exists as the following: (1) aglycosylated TH(296-304), which naturally lacks glycosylation; and (2), deglycosylated TH(296-304), which lacks glycosylation due to enzymatic removal. The nominal mass shift between these analogous peptides allows for N-linked site determination and relative quantitation of N-linked site occupancy based on mass spectral peak integration.

The NISTmAb peptides generated from tryptic digestion with and without PNGase F treatment were subjected to a high resolution/accurate mass peptide map analysis to confirm the N-glycosylation attachment sites and determine the percent site occupancy. The instrument method was set up in a data-independent fashion in which the mass spectrometer software drives the data acquisition. For example, as the chromatographic separation occurs, the mass spectrometer sequentially cycles through three scan modes: (1) first-order MS survey scan; (2) zoom scan on the *n*th-most abundant species in the first-order survey scan; and (3) gas-phase dissociation (i.e., MS/MS fragmentation) of the selected species. A criterion was established in which the mass accuracy error tolerance between the theoretical and empirical peptide masses had to be ≤ 10 ppm to trigger MS/MS fragmentation. High-resolution/accurate mass data easily identified the aglycosylated and deglycosylated TH(296-304) peptides with better than 10 ppm mass accuracy (Table 4). The MS fragmentation results shown in Figure 6 confirm the presence of an Asp residue corresponding to the site where the *N*-glycosylated Asn<sup>300</sup> of the heavy chain resided prior to PNGase F treatment.

**Table 4. Theoretical and Observed Peptide Masses for the Deglycosylated and Aglycosylated TH(296-304) Peptide**

Peptide	Theoretical Mass <sup>†</sup>	Theoretical $m/z$ [M+2H] <sup>2+</sup> [M+1H] <sup>+</sup>	Observed $m/z$	Observed Mass <sup>†</sup>	Mass Accuracy (ppm)
Aglycosylated TH(296-304)	1188.5047	595.2596	595.2578	1188.5000	4.00
		1189.5120	1189.5087	1188.5009	3.22
Deglycosylated TH(296-304)	1189.4887	595.7516	595.7499	1189.4842	3.83
		1190.4960	1190.4916	1189.4838	4.14

<sup>†</sup> Denotes monoisotopic mass.

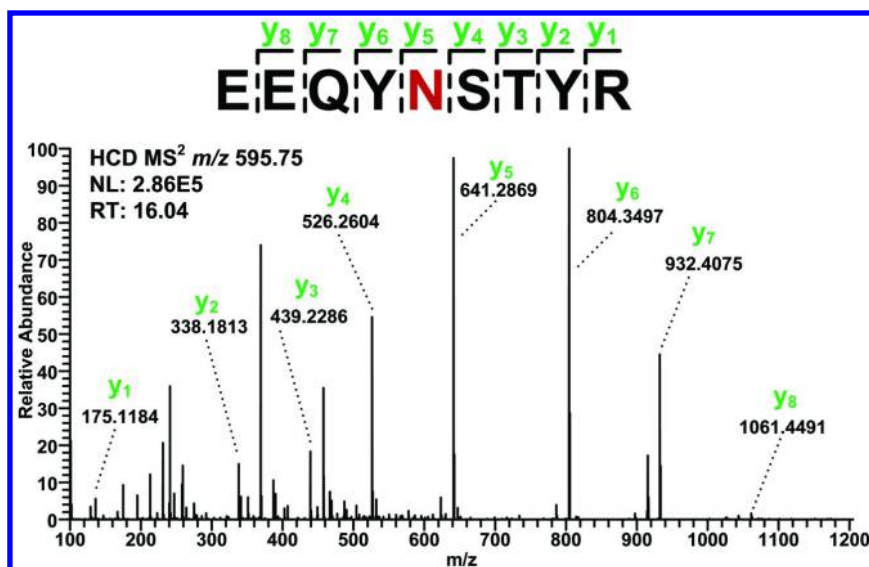


Figure 6. Tandem mass spectrometry (MS/MS) higher energy collisional dissociation (HCD) spectrum of the deglycosylated TH(296-304) peptide confirming the sequence position of the N-glycosylation site at Asn<sup>300</sup>. The red N denotes conversion of the Asn to Asp as a result of N-glycosidase F (PNGase F) release. (see color insert)

The exceptional dynamic range (> 5000 relative ion intensity) and resolving power of the Orbitrap enables a peptide ion to be distinguished from noise and its mass accurately measured at a level of 2 to 3 signal-to-noise (S/N) (37). As such, low-abundance or suppressed co-eluting species, such as the aglycosylated TH(296-304) peptide, can be confidently identified. However, the <sup>13</sup>C isotopes of the aglycosylated peptide cannot be spectrally resolved



from the deglycosylated monoisotope and may result in inaccurate quantitation. Chromatographic resolution of the two species is therefore important for accurate quantitation via extracted ion chromatograms (EICs). The extent of occupancy of the N-glycosylation site at Asn<sup>300</sup> was determined from the LC-MS-based tryptic peptide map following deglycosylation with PNGase F. Figure 7 demonstrates baseline chromatographic resolution of the aglycosylated and deglycosylated TH(296-304) peptides and, therefore, EICs can be used for accurate evaluation of percent site occupancy.

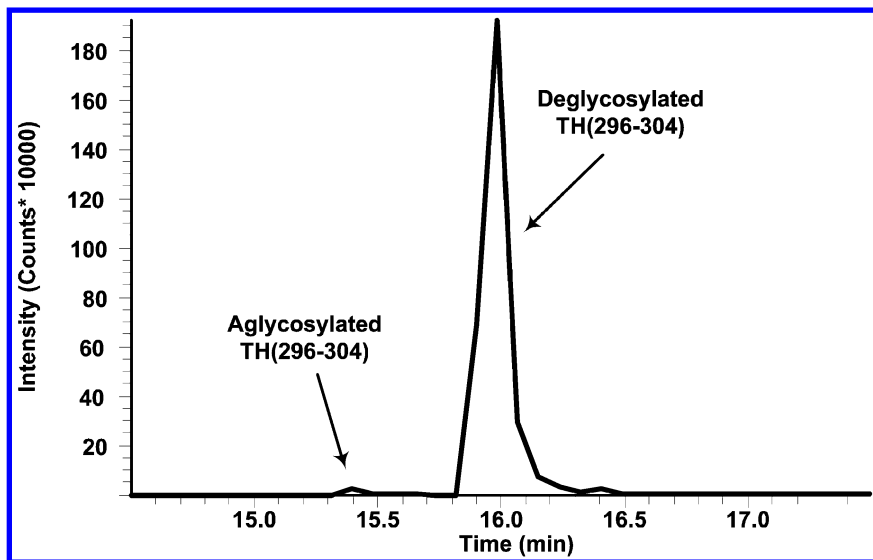


Figure 7. Extracted ion chromatogram of deglycosylated and aglycosylated TH(296-304) peptide.

The percent occupancy of the N-glycosylation site was determined from the absolute peak areas of the EIC traces (sum of 1+ and 2+ monoisotopic peaks) for the aglycosylated and deglycosylated TH(296-304) peptides, as presented in Figure 7. The calculation assumed that the ionization efficiency of the aglycosylated and deglycosylated species were equivalent; in practice, given that the aglycosylated peptide contains an Asn residue whereas the deglycosylated peptide contains an Asp residue at position 300, the ionization efficiency of the aglycosylated peptide is likely to be slightly higher than the deglycosylated peptide. Accordingly, the proportion of aglycosylated peptide is probably slightly over-reported. The calculated percentages demonstrate that the N-glycosylation sites are approximately 99.0% occupied in the native NISTmAb. These values are in excellent agreement with those reported in the Separation chapter/Volume 2, Chapter 5 using orthogonal capillary sodium dodecylsulfate electrophoresis (cSDS) methods.

Although the NISTmAb contains a single consensus sequence in the conserved Fc region of each heavy chain, glycosylation on the Fab region of

the molecule may also be present. N-glycosylation can occur at low levels (0.5 to 2.0%) on C<sub>H1</sub> domain Asn residues not residing within a consensus site motif (Asn-Xaa-Ser/Thr) (38, 39). This modification, commonly known as non-consensus glycosylation, is present on both recombinant human IgG and antibodies derived from human donors. The extent of non-consensus glycosylation may be assessed quantitatively using reduced capillary electrophoresis based on the number of occupied glycosylation sites (0, 1, and 2) (38, 39). Heavy chain modified with both consensus and non-consensus glycosylation migrates to a position past the putative heavy chain modified with only consensus glycosylation. For the purposes of this chapter, non-consensus N-glycosylation of NISTmAb was not pursued.

## Species Identification and Quantitation

The use of high-throughput chromatography with fluorescence detection alone is a powerful technique to monitor N-glycosylation for process support and to annotate peaks based on the GU unit system; however, for the purposes of product characterization, higher resolving chromatographic methods coupled to mass spectrometric detection are necessary. Liquid chromatography/mass spectrometry (LC-MS)-based methods provide exquisite sensitivity, rapid data acquisition, and fundamental physical measurement-based analysis, enabling the detection of low-abundance, co-eluting glycomers as well as structural isomers that resolve on the LC. In addition, orthogonal mass spectrometry detection can discriminate between a fluorescently labeled glycan species and a fluorescence emitting sample artifact that elutes at the same chromatographic elution index (see the LC-fluorescence detection [FLD]-MS/MS method development considerations section, below). Significant effort is made to characterize the minor peaks in a chromatographic profile for deeper product and process understanding. A glycan characterization workflow with the capability of sequencing glycan species, including co-eluting structural isomers at  $\geq 0.1\%$  of the total chromatographic peak area, will be discussed in the next section.

Figure 8 depicts the glycan characterization workflow used to identify glycan species present on the NISTmAb. Briefly, N-linked glycans were released enzymatically from the protein backbone using PNGase F, fluorescently labeled, and chromatographically separated by HILIC with fluorescence detection (Steps 1–2a; Figure 8). The back end of the chromatographic instrument was coupled to an Advion Triversa Nanomate<sup>®</sup> operating in simultaneous LC fraction collection/MS Nano-infusion mode (NSI-MS), infusing sample directly into an ion trap mass spectrometer for preliminary glycan identification (Steps 2b-c; Figure 8). Each collected HILIC fraction was enriched and subjected to both off-line second dimension reversed-phase separation (LC-MS) to quantitate co-eluting species and permethylated sequential mass spectrometry (NSI-MS<sup>n</sup>) to sequence the glycan species (Step 3a and 3b; Figure 8). Although not shown in Figure 8, strategic exoglycosidase treatment and WAX were used as an orthogonal confirmation of species assignment and quantitation of acidic species.

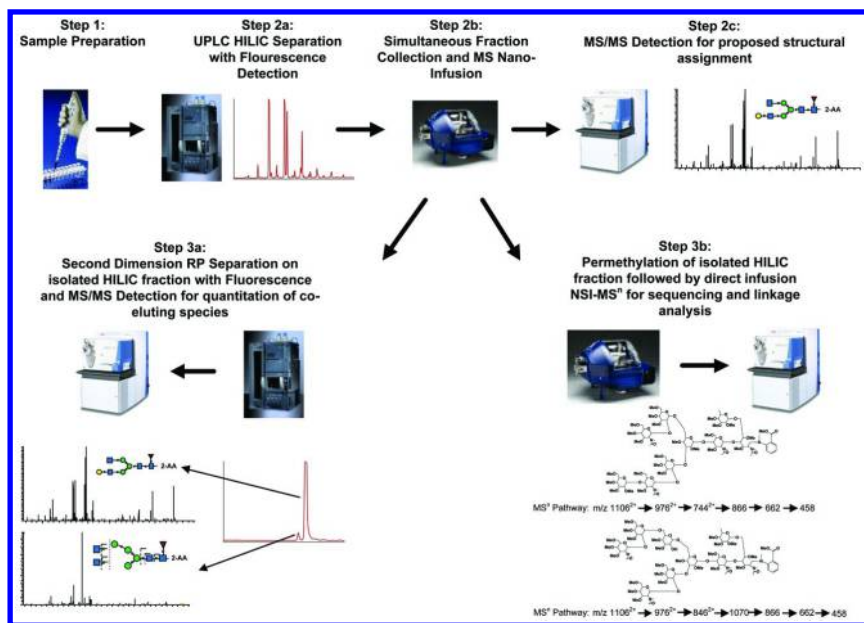


Figure 8. Glycan characterization workflow used for product characterization studies. Workflow steps include UPLC (ultra-performance liquid chromatographic systems) HILIC (hydrophilic interaction liquid chromatography) separation, MS (mass spectrometry), tandem mass spectrometry (MS/MS), RP (reversed phase) separation, and NSI-MS<sup>n</sup> (sequential MS Nano-infusion mode). (see color insert)

HILIC separation is the result of a polar stationary phase and mobile phases with high organic and salt buffer content (40). The strongly retained polar glycans are eluted with an increasing aqueous gradient. Fluorescently labeled glycans are predictably retained according to the number and accessibility of polar hydroxyl groups extending from the hexose constituents of the glycomer (41). Interestingly, fucose and *N*-acetylglucosamine residues exhibit lower hydrophilicity compared to hexose residues and contribute significantly less to on-column retention (40). Due to the limited influence of fucose and *N*-acetylglucosamine residues on retention, the opportunity for co-elution of species is possible using HILIC chromatography. Often a low flow rate and extended elution period aids in reducing the overlap of co-eluting species; therefore, an HILIC method with a lower flow rate and extended elution period was used for detailed product characterization. The fluorescent label 2-aminobenzoic acid (2-AA) was used in this characterization workflow. The 2-AA label has been reported to confer advantages such as an increase in reactivity in aqueous conditions to improve sample workflow efficiency and an increase in fluorescence intensity to improve chromatographic sensitivity for low-abundance glycan species (42, 43). Intriguingly, 2-AA affords even greater advantage with respect to mass spectrometric analysis such as improved ionization efficiency in the negative mode, allowing improved sensitivity for detection of low-abundance species

and preferential deprotonation of the reducing termini, promoting sequential disassembly from the nonreducing termini to the reducing termini and allowing for improved glycan sequencing (17, 44–49). Figure 9 depicts the fluorescence profile of released 2-AA labeled glycans from the NISTmAb.

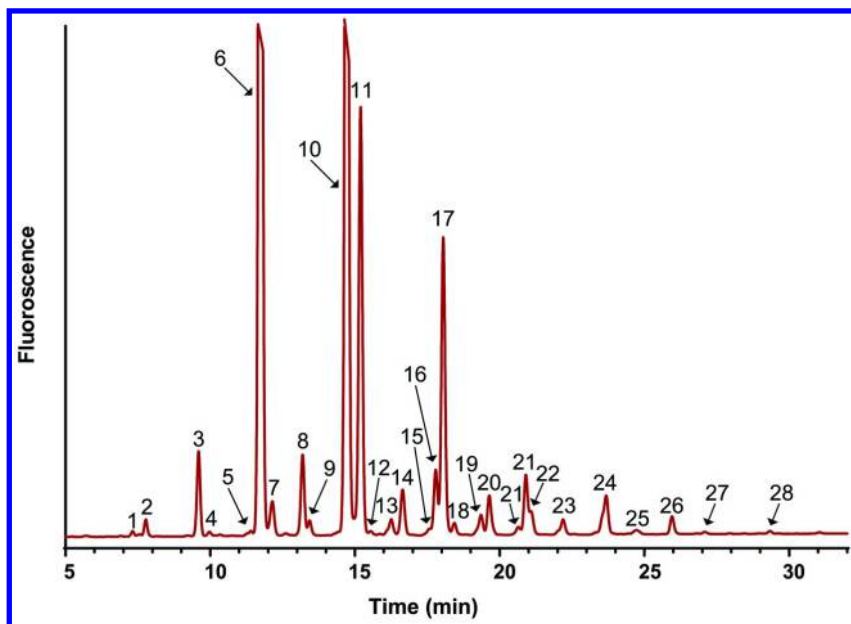


Figure 9. Representative zoomed glycan map of NIST IgG1. (see color insert)

The compositions of species in the N-linked glycan map (Figure 9) were determined by coupling the outlet of the chromatographic instrument to an LTQ Velos Pro ion trap mass spectrometer via the Advion Triversa Nanomate Nano-electrospray source set in LC mode. Composition assignments for the observed  $m/z$  were made using the theoretical mass values of non-specific monosaccharide residues: hexose (H), *N*-acetylhexosamine (N), deoxyhexose (F), and Neu5Gc (Sg). The identification of each peak composition  $\geq 0.1\%$  of the total peak area is summarized in Table 5.

The percent relative abundance value for individual glycan species are slightly different between the higher resolving HILIC map used for product characterization compared to the high-throughput HILIC map, mentioned in the previous section, used for process support. These discrepancies can be accounted for by the differences in elution gradient and mobile phase content that result in the co-elution of species. For example, in the high-throughput map, FA1 and A2, co-elute, whereas FA1 and A2 are well-resolved in the product characterization HILIC map (Figures 2 and Figures 14, respectively), leading to the discrepancy in the species quantitation between the two methods. Importantly, the high-throughput map is used as a precision-based assay to monitor process consistency, whereas the higher resolving map is used for product characterization. Each method is fit for its intended purpose.

**Table 5. Identification of N-linked Glycan Compositions  $\geq 0.1\%$  of the Total Peak Area**

<i>Peak Label</i>	<i>Approximate Retention Time (Min)</i>	<i>% Peak Area</i>	<i>Observed Mass</i>	<i>Theoretical Mass<sup>†</sup></i>	<i>Mass Accuracy (ppm)</i>	<i>Empirical Formula<sup>†</sup></i>	<i>Proposed Composition</i>	<i>Proposed Structure<sup>‡</sup></i>
<b>1</b>	7.32	0.11	1177.50	1177.44	49	C <sub>47</sub> H <sub>75</sub> N <sub>3</sub> O <sub>31</sub>	H <sub>3</sub> N <sub>2</sub> F <sub>1</sub> -2AA	FM3
<b>2</b>	7.78	0.35	1234.54	1234.46	63	C <sub>49</sub> H <sub>78</sub> N <sub>4</sub> O <sub>32</sub>	H <sub>3</sub> N <sub>3</sub> -2AA	A1
<b>3</b>	9.60	1.81	1380.63	1380.52	78	C <sub>55</sub> H <sub>88</sub> N <sub>4</sub> O <sub>36</sub>	H <sub>3</sub> N <sub>3</sub> F <sub>1</sub> -2AA	FA1
<b>4</b>	9.97	0.11	1437.64	1437.54	71	C <sub>57</sub> H <sub>91</sub> N <sub>5</sub> O <sub>37</sub>	H <sub>3</sub> N <sub>4</sub> -2AA	A2
<b>5</b>	11.40	0.15	1565.72	1565.59	84	C <sub>63</sub> H <sub>99</sub> N <sub>5</sub> O <sub>40</sub>	H <sub>3</sub> N <sub>4</sub> F <sub>1</sub> -2AA-H <sub>2</sub> O	FA2-H <sub>2</sub> O
<b>6</b>	11.74	38.06	1583.74	1583.60	87	C <sub>63</sub> H <sub>101</sub> N <sub>5</sub> O <sub>41</sub>	H <sub>3</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2
<b>7</b>	12.14	0.94	1355.61	1355.49	87	C <sub>53</sub> H <sub>85</sub> N <sub>3</sub> O <sub>37</sub>	H <sub>5</sub> N <sub>2</sub> -2AA	M5
<b>8</b>	13.19	1.85	1542.71	1542.57	98	C <sub>61</sub> H <sub>98</sub> N <sub>4</sub> O <sub>41</sub>	H <sub>4</sub> N <sub>3</sub> F <sub>1</sub> -2AA	FA1G1
<b>9</b>	13.43	0.35	1786.96	1786.68	159	C <sub>71</sub> H <sub>114</sub> N <sub>6</sub> O <sub>46</sub>	H <sub>3</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA3
<b>10</b>	14.72	28.49	1745.78	1745.65	75	C <sub>69</sub> H <sub>111</sub> N <sub>5</sub> O <sub>46</sub>	H <sub>4</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G1-a
<b>11</b>	15.20	10.11	1745.80	1745.65	87	C <sub>69</sub> H <sub>111</sub> N <sub>5</sub> O <sub>46</sub>	H <sub>4</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FM4A2
								FA2G1-b
<b>12</b>	15.54	0.10	1517.29	1517.54	166	C <sub>59</sub> H <sub>95</sub> N <sub>3</sub> O <sub>42</sub>	H <sub>6</sub> N <sub>2</sub> -2AA	M6
<b>13</b>	16.25	0.45	1948.68	1948.81	64	C <sub>67</sub> H <sub>108</sub> N <sub>4</sub> O <sub>46</sub>	H <sub>4</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA3G1-a

Peak Label	Approximate Retention Time (Min)	% Peak Area	Observed Mass	Theoretical Mass <sup>†</sup>	Mass Accuracy (ppm)	Empirical Formula <sup>‡</sup>	Proposed Composition	Proposed Structure <sup>‡</sup>
14	16.65	1.15	1704.60	1704.62	9	C <sub>67</sub> H <sub>108</sub> N <sub>4</sub> O <sub>46</sub>	H <sub>5</sub> N <sub>3</sub> F <sub>1</sub> -2AA	FA1G1Ga1
			1948.60	1948.81	106	C <sub>67</sub> H <sub>108</sub> N <sub>4</sub> O <sub>46</sub>	H <sub>4</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FM4A1G1
15	17.57	0.15	1721.14	1720.61	305	C <sub>67</sub> H <sub>108</sub> N <sub>4</sub> O <sub>47</sub>	H <sub>6</sub> N <sub>3</sub> -2AA	FA3G1-b
16	17.79	1.59	1907.76	1907.70	33	C <sub>75</sub> H <sub>121</sub> N <sub>5</sub> O <sub>51</sub>	H <sub>5</sub> N <sub>4</sub> F <sub>1</sub> -2AA	M5A1G1
17	18.05	7.14	1907.84	1907.70	74	C <sub>75</sub> H <sub>121</sub> N <sub>5</sub> O <sub>51</sub>	H <sub>5</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G1Ga1-a
18	18.43	0.32	1908.14	1907.70	233	C <sub>75</sub> H <sub>121</sub> N <sub>5</sub> O <sub>51</sub>	H <sub>5</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G2
19	19.35	0.60	2110.51	2110.78	127	C <sub>83</sub> H <sub>134</sub> N <sub>6</sub> O <sub>56</sub>	H <sub>5</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA2G1Ga1-b
20	19.64	1.04	1849.78	1849.66	66	C <sub>72</sub> H <sub>115</sub> N <sub>5</sub> O <sub>50</sub>	Sg <sub>1</sub> H <sub>4</sub> N <sub>3</sub> F <sub>1</sub> -2AA	FA3G2-a
			2111.62	2110.78	396	C <sub>83</sub> H <sub>134</sub> N <sub>6</sub> O <sub>56</sub>	H <sub>5</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA1G1Gc1
21	20.65	0.21	2053.34	2052.74	290	C <sub>80</sub> H <sub>128</sub> N <sub>6</sub> O <sub>55</sub>	Sg <sub>1</sub> H <sub>4</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA3G2-b
			1866.38	1865.66	384	C <sub>72</sub> H <sub>115</sub> N <sub>5</sub> O <sub>51</sub>	Sg <sub>1</sub> H <sub>5</sub> N <sub>3</sub> -2AA	FA2G1Gc1-a
			1882.62	1882.67	29	C <sub>73</sub> H <sub>118</sub> N <sub>4</sub> O <sub>52</sub>	H <sub>7</sub> N <sub>3</sub> -2AA	M4A1G1Gc1
22	20.90	1.53	2070.38	2069.71	325	C <sub>81</sub> H <sub>131</sub> N <sub>5</sub> O <sub>56</sub>	H <sub>7</sub> N <sub>3</sub> -2AA	M5A1G1Ga1
23	21.08	0.52	2053.42	2052.74	329	C <sub>80</sub> H <sub>128</sub> N <sub>6</sub> O <sub>55</sub>	H <sub>6</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G2Ga1
24	22.19	0.48	2273.66	2272.84	359	C <sub>89</sub> H <sub>144</sub> N <sub>6</sub> O <sub>61</sub>	Sg <sub>1</sub> H <sub>4</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G1Gc1-b
							H <sub>6</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA3G3

Continued on next page.

**Table 5. (Continued). Identification of N-linked Glycan Compositions  $\geq 0.1\%$  of the Total Peak Area**

<i>Peak Label</i>	<i>Approximate Retention Time (Min)</i>	<i>% Peak Area</i>	<i>Observed Mass</i>	<i>Theoretical Mass<sup>†</sup></i>	<i>Mass Accuracy (ppm)</i>	<i>Empirical Formula<sup>†</sup></i>	<i>Proposed Composition</i>	<i>Proposed Structure<sup>‡</sup></i>
<b>25</b>	23.68	1.41	2215.54	2214.79	336	C <sub>86</sub> H <sub>138</sub> N <sub>6</sub> O <sub>60</sub>	Sg <sub>1</sub> H <sub>5</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G2Gc1
			2232.58	2231.81	343	C <sub>87</sub> H <sub>141</sub> N <sub>5</sub> O <sub>61</sub>	H <sub>7</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G2Ga2
<b>26</b>	24.71	0.17	2435.80	2434.89	375	C <sub>95</sub> H <sub>154</sub> N <sub>6</sub> O <sub>66</sub>	H <sub>7</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA3G3Ga1
<b>27</b>	25.95	0.45	2376.78	2376.85	31	C <sub>92</sub> H <sub>148</sub> N <sub>6</sub> O <sub>65</sub>	Sg <sub>1</sub> H <sub>6</sub> N <sub>4</sub> F <sub>1</sub> -2AA	FA2G2Ga1Gc1
<b>28</b>	27.18	0.06	2614.38	2612.94	524	C <sub>101</sub> H <sub>164</sub> N <sub>6</sub> O <sub>72</sub>	H <sub>9</sub> N <sub>5</sub> -2AA	FA3G3Ga2
<b>29</b>	29.33	0.07	2759.64	2758.99	237	C <sub>107</sub> H <sub>174</sub> N <sub>6</sub> O <sub>76</sub>	H <sub>9</sub> N <sub>5</sub> F <sub>1</sub> -2AA	FA3G3Ga3

<sup>†</sup> Theoretical mass and chemical formula include the 2-aminobenzoic acid label (2-AA). <sup>‡</sup> Proposed structures are based on composition analysis and known glycosylation biosynthetic pathways. Exact glycan structures have not been empirically determined at the time of publication. Structural isomers are noted as -a and -b.

Even with the higher resolving HILIC method, co-elution of species occurs. As such, a second dimension chromatographic separation using UPLC-based reversed-phase (RP LC-MS) chromatography was employed for the quantitation of co-eluting species. Under RP chromatographic conditions, glycans elute in a structurally related group pattern in the following order: (1) high mannose species; (2) sialylated afucosylated hybrid and complex species; (3) afucosylated hybrid species; (4) afucosylated complex species; (5) sialylated fucosylated hybrid and complex species; (6) asialylo, fucosylated hybrid species; and (7) asialylo, fucosylated complex species (18, 41, 50, 51). The orthogonal chromatographic properties of the UPLC RP method were effectively utilized to isolate the NISTmAb glycan species that co-elute under HILIC conditions. Additionally, the mobile phases of the UPLC RP method, 0.1% formic acid and acetonitrile with 0.1% formic acid, are highly compatible with on-line mass spectrometric analysis compared to the HILIC mobile phases, which require a higher salt content for adequate chromatographic separation (15, 41).

Peak 11 shown in Figure 9 and Table 5 will be used to exemplify the ability of the analytical workflow to characterize and quantitate low-abundance co-eluting species. Figure 10 depicts the positive and negative mode spectra associated with Peak 11. The positive mode MS/MS spectrum product ions  $m/z$  1543.60 (neutral loss of a terminal HexNAc),  $m/z$  1381.52 (neutral loss of a Hex-HexNAc disaccharide),  $m/z$  1258.43 (reducing end 2-AA labeled Fuc-HexNAc disaccharide), and  $m/z$  1178.43 (neutral loss of two HexNAc and one hexose residue) are consistent with fragments originating from a FA2G1 structure represented as a cartoon in Panel A of Figure 10. The negative mode MS/MS spectrum product ions  $m/z$  790.76<sup>2-</sup> (neutral loss of terminal hexose residue),  $m/z$  770.42<sup>2-</sup> (neutral loss of terminal HexNAc residue),  $m/z$  689.26<sup>2-</sup> (neutral loss of a Hex-HexNAc disaccharide),  $m/z$  487.25 (reducing end 2-AA labeled Fuc-HexNAc disaccharide), and  $m/z$  364.17 (terminal Hex-HexNAc disaccharide) are consistent with fragments stemming from a FA2G1 structure and further corroborate the presence of an FA2G1 species eluting at Peak 11 (Panel B, Figure 10).

When evaluating MS/MS spectra, it is important to probe for product ions that are not compatible with fragments originating from the dissociation of the proposed structure. The presence of aberrant product ions may indicate the existence of a co-eluting structural isomer. For example, in Panel A of Figure 10, product ions  $m/z$  1340.52 (neutral loss of two terminal HexNAc residues) is inconsistent with the proposed FA2G1 structure but do, however, align with fragment ions stemming from a glycan species having two terminally exposed HexNAc residues. Monomer rearrangement has been reported when disassembling neutral glycan species under positive mode, protonated, low-energy collision-induced dissociation (CID) conditions. As such,  $m/z$  1340.52 may be an artifact of gas-phase rearrangement. Interestingly, monomer rearrangement has not been reported under negative mode, deprotonated, low-energy CID conditions. The negative mode CID spectrum of Peak 11 contains the doubly deprotonated product ion,  $m/z$  669.62<sup>2-</sup> (Panel B of Figure 10). This corresponds to the neutral loss of two terminal HexNAc residues, suggesting that a structural



isomer having two terminal HexNAc residues co-elutes with FA2G1 in Peak 11, which is further confirmed with orthogonal techniques below.

In addition to positive and negative ionization modes, data confirmation using orthogonal methodologies is important when evaluating low-abundance species. In this workstream, we employed the use of exoglycosidases either individually or in a specific combination to orthogonally verify the presence of low-abundance glycans that were preliminarily detected by LC-MS methodologies. The NISTmAb-released glycans were digested with  $\beta$ -(1-4)-galactosidase to identify any potential glycan structures containing a terminal  $\beta$ -(1-4)-galactose residue (Figure 11). Following  $\beta$ -(1-4)-galactosidase treatment, a minor residual peak remained at the Peak 11 elution position. This information suggests either the incomplete digestion of the putative FA2G1 structure, the digestion product of a high-order glycan now eluting at the Peak 11 elution index, the presence of an agalactosylated structural isomer unaffected by  $\beta$ -(1-4)-galactosidase treatment, or a combination of the aforementioned. Next, the glycan pool was treated with  $\beta$ -N-acetylhexosaminidase, which removes terminal  $\beta$ -N-Acetylglucosamine residues. As shown in Figure 11, Peak 11 was completely reduced to baseline after  $\beta$ -N-acetylhexosaminidase treatment signifying the removal of terminal N-acetylhexosamines from glycan structures residing at the Peak 11 elution position. These exoglycosidase data, coupled with the MS data shown in Figure 10, suggest the co-elution of the putative FA2G1 structure and a structural isomer containing two terminally exposed HexNAc residues.

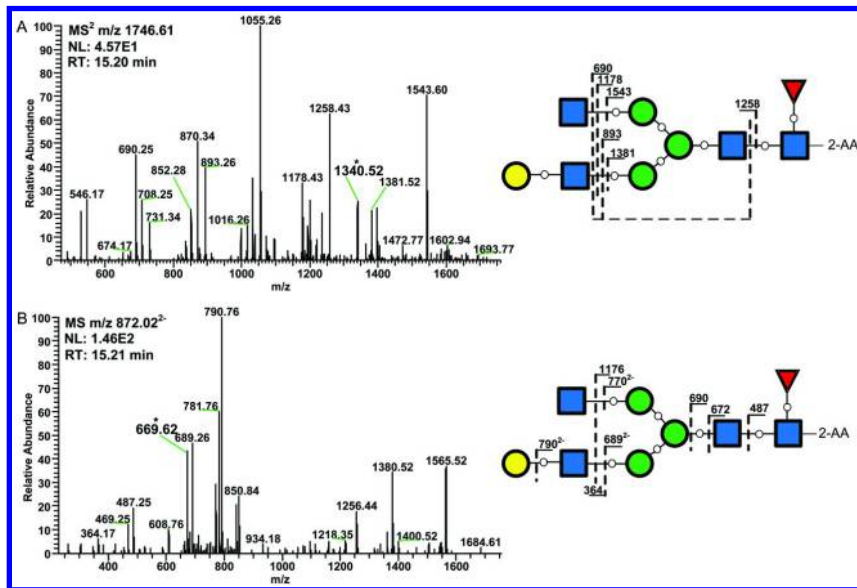


Figure 10. Tandem mass spectrometry (MS/MS) positive (A) and negative (B) mode spectra from hydrophilic interaction liquid chromatography (HILIC) Peak 11 with putative structure assignments shown. The peaks marked with an asterisk represent product ions inconsistent with the putative structure. (see color insert)

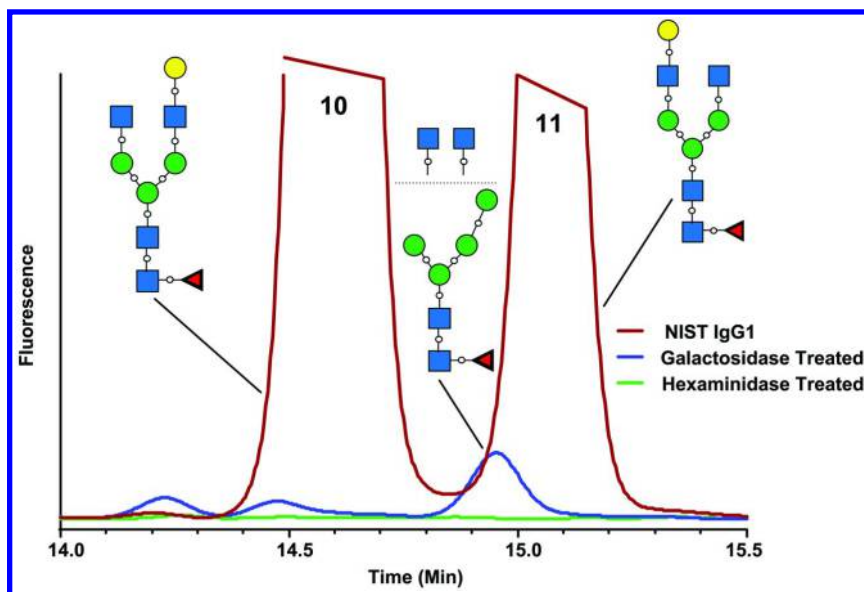


Figure 11. Overlay of NIST monoclonal antibody (mAb) glycan map with and without  $\beta$ -(1-4)-galactosidase and  $\beta$ -N-acetylhexosaminidase treatment, demonstrating the presence of co-eluting species. (see color insert)

The Peak 11 fraction was also enriched using the Advion Triversa Nanomate<sup>®</sup> and then subjected to off-line second dimension RP chromatographic separation coupled to a linear ion trap mass spectrometer. Figure 12 portrays the RP separation of HILIC Peak 11 and demonstrates two distinct, highly resolved peaks, labeled Peak A and Peak B, eluting at 13.71 minutes and 14.21 minutes, respectively. The associated MS spectra from both Peak A and Peak B exhibit a doubly deprotonated molecular ion  $[M-2H]^{2-}$  species,  $m/z$  872.13 and  $m/z$  872.18, respectively, which is consistent with structural isomers composed of a  $Hex_4HexNAc_4Deoxy_1$  composition (data not shown). The Peak A product ion,  $m/z$  668.92<sup>2-</sup>, corresponds to the neutral loss of two terminal HexNAc residues, which is consistent with an agalactosylated structure, whereas the Peak B deprotonated MS/MS spectrum exhibits a product ion distribution that is consistent with the expected galactosylated FA2G1 structure (Figure 12). As such, the RP LC-MS data demonstrates the co-elution of  $Hex_4HexNAc_4Deoxy_1$  structural isomers at the Peak 11 HILIC elution index. Based on the relative proportion of the Peak A and Peak B peak areas compared to the total RP peak area for both peaks multiplied by the total HILIC percent area of Peak 11, it was determined that both isomeric species exceed the experimental reportable criteria of  $\geq 0.1\%$  of the total HILIC peak area used for the current study (Table 6).

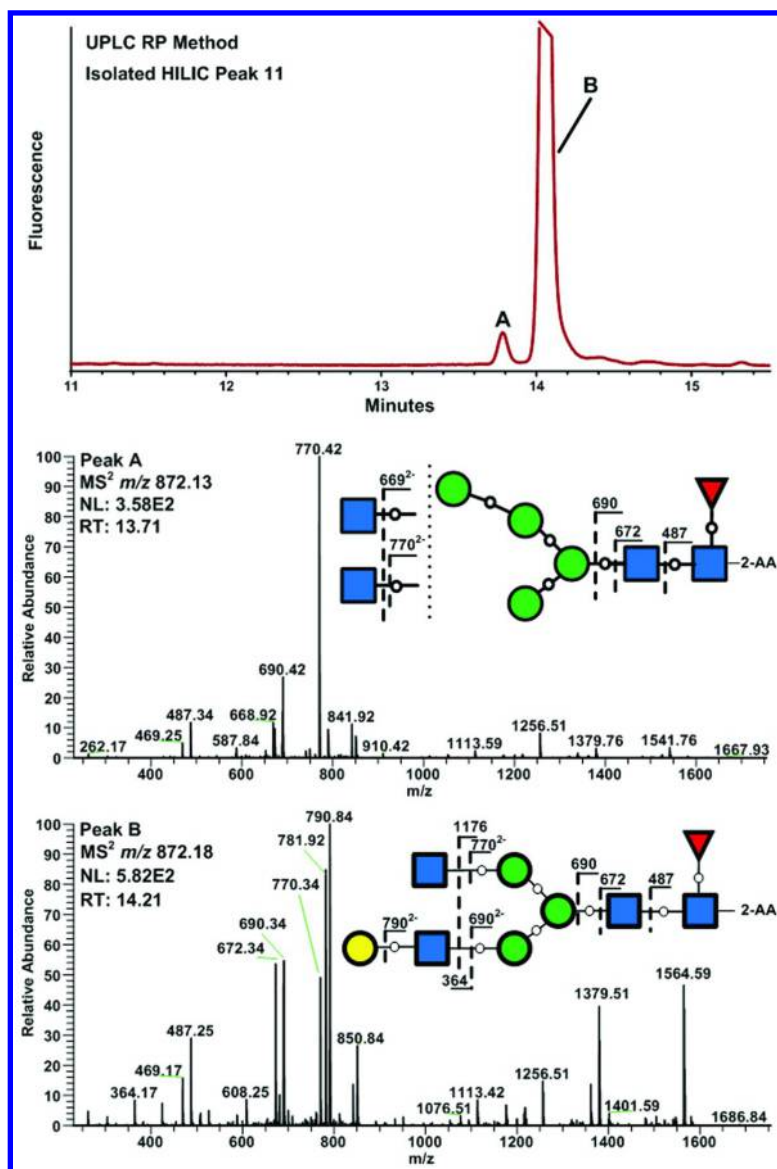


Figure 12. Second dimension ultra-performance liquid chromatography reversed phase (UPLC RP) separation of hydrophilic interaction liquid chromatography (HILIC) Peak 11 and tandem mass spectrometry (MS/MS) spectra of resolved ions. (see color insert)

**Table 6. Hydrophilic Interaction Liquid Chromatography (HILIC) Peak 11 Co-Eluting Species and Relative Percentages**

<i>Species Name</i>	<i>UPLC RP Peak Area</i>	<i>UPLC RP % Area</i>	<i>HILIC % Area</i> <sup>†</sup>
<b>HILIC Peak 11</b>	9,151,265	N/A	10.11
<b>FM4A2</b>	394,381	4.31	0.44
<b>FA2G1-b</b>	8,756,884	95.69	9.67

<sup>†</sup> HILIC percent area for each glycan species was calculated based on the relative proportion of the glycan ultra-performance liquid chromatography reversed phase (UPLC RP) peak area to the total UPLC RP peak area multiplied by the total HILIC percent area of Peak 11.











Numerous reports have shown that glycan permethylation followed by nanospray sequential mass spectrometry (NSI-MS<sup>n</sup>) provides highly informative spectral data suitable for structural elucidation (16, 51–67). In short, several pertinent benefits of glycan permethylation with respect to structural characterization include: (1) increased oligosaccharide ionization efficiency; (2) increased abundance of diagnostic cross-ring fragments that can be utilized to determine linkage position; and (3) upon fragmentation, terminal and internal residues exhibit mass differences (14 Da) depending on the number of exposed hydroxyl groups, commonly known as “scars.” Glycan structures are sequenced by following diagnostic MS<sup>n</sup> precursor-product pathways. Table 7 lists the fragments and associated mass values used for structure determination.

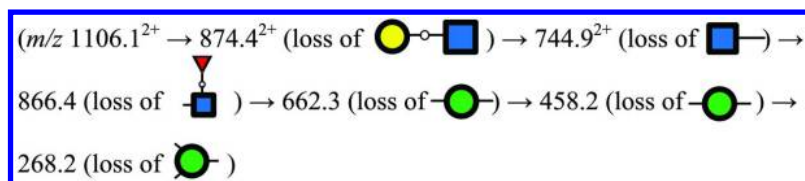
The Peak 11 fraction was purified, permethylated, and subjected to NSI-MS<sup>n</sup> as a fourth tier of confirmatory characterization. The NSI-MS<sup>n</sup> precursor-product pathway shown in Scheme 1 was followed, and consistent with the putative FA2G1 glycan topology (Figure 13).

In this instance, the fully permethylated, doubly charged, sodium-adducted parent ion ( $m/z$  1106.1) was selected and subjected to CID, resulting in the fragmentation pattern in Figure 13, Panel A. Successive rounds of isolation and dissociation of specific product fragments at each subsequent step (as shown in Figure 13, Panels B through F) indicate the manner in which a glycan structure is sequentially determined. The formation of both glycosidic cleavage and cross-ring product ions at each disassembly stage affords unambiguous assignment of glycan structures.

Permethylation and MS<sup>n</sup> allows verification of many specific linkages in the putative structures. For example, the fragment ion,  $m/z$  486.3, from the MS<sup>2</sup>  $m/z$  1106.1<sup>2+</sup> spectrum (Figure 13A) was selected for subsequent MS<sup>3</sup> analysis. Figure 14 exhibits the MS<sup>3</sup>  $m/z$  486.3 spectrum, representing a terminal B<sub>2</sub>-ion disaccharide. Several product ions,  $m/z$  241.1,  $m/z$  259.1,  $m/z$  268.1, and  $m/z$  329—representing a B<sub>1</sub>-ion, C<sub>1</sub>-ion, B/Y-ion, and a <sup>3,5</sup>A-ion, respectively—are present and diagnostic of a 1-4-linked lactosamine extension. Here we demonstrate that permethylation followed by NSI-MS<sup>n</sup> empirically sequenced the FA2G1 topology and verified the terminal lactosamine linkage without the need to leverage glycosylation biosynthetic understanding.

**Table 7. Ion Mass Values for Permethylated Fragments of Glycan Species (16) (see color insert)**

Fragment Symbol	Fragment Identity	Neutral Loss (Nominal Singly-Charged Precursor)	Nominal Sodiated Mass/Charge $[M+Na]^+$
	Terminal mannose, B type	218	241
	Terminal galactose, B type	218	241
	Terminal N-acetylglucosamine, B type	259	282
	Internal mannose, B/Y type	204	227
	Internal N-acetylglucosamine, B/Y type	245	268
	Bifurcating internal mannose, B/Y/Y type	190	213
	2-aminobenzoic labeled, N-acetylglucosamine reducing end terminus	426	449
	2-aminobenzoic labeled, fucosylated N-acetylglucosamine reducing end terminus	605	628
	Terminal lactosamine extension, B type	463	486
	Core mannose-(1,4)-N-acetylglucosamine, B/Y/Y type	N/A	458



*Scheme 1. NSI-MS<sup>n</sup> precursor-product pathway consistent with putative FA2G1 glycan topology. (see color insert)*

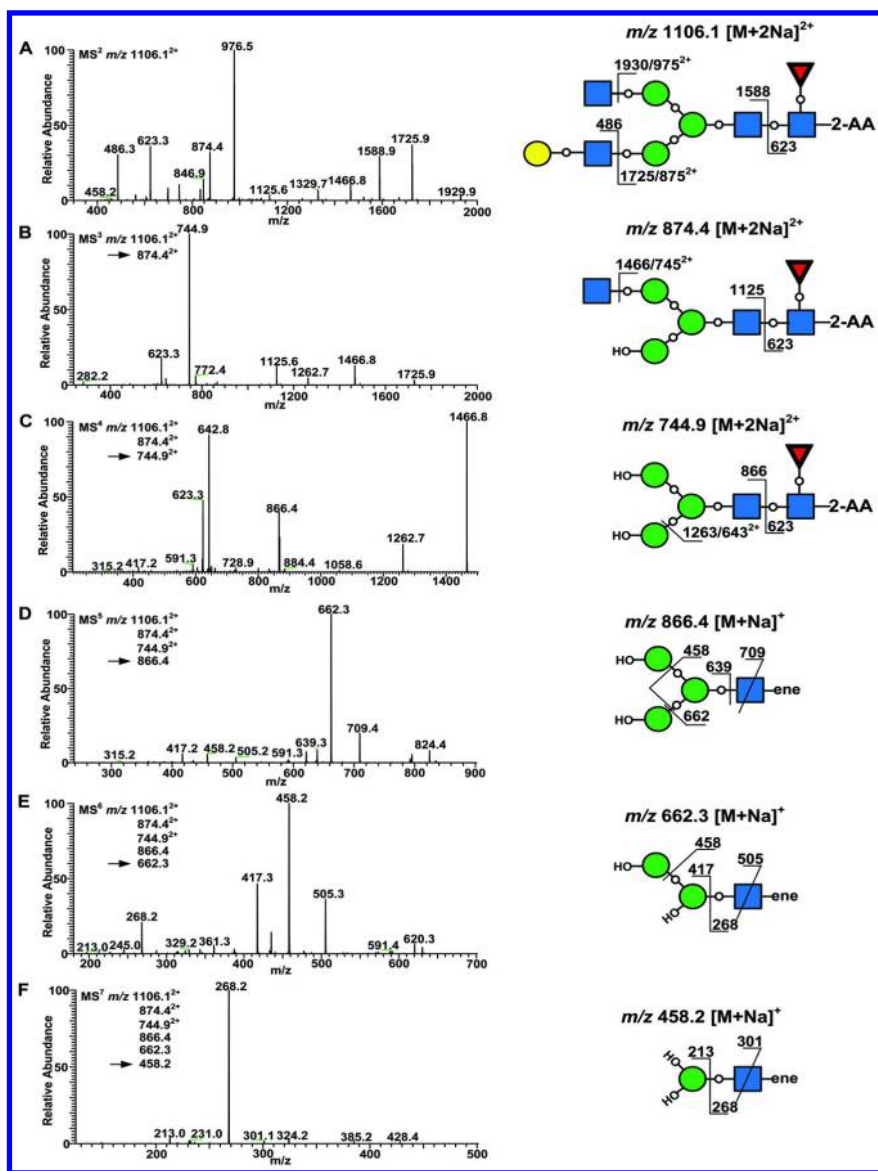


Figure 13. Sequential mass spectrometry ( $MS^n$ ) spectrum for putative FA2G1 structure from Peak 11 fraction. (see color insert)

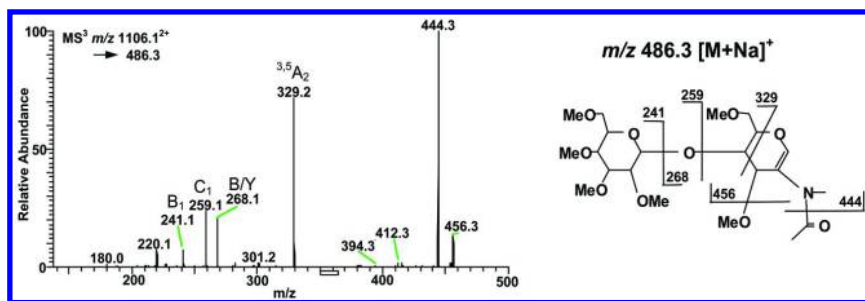


Figure 14. MS<sup>3</sup> m/z 486 spectrum confirming a β-(1-4)-linked lactosamine extension from the putative FA2G1 structure. (see color insert)

The aforementioned MS<sup>n</sup> pathways are diagnostic of the position of each monomer and the linkage position of the terminal galactose of the lactosamine extension, confirming the FA2G1 topology. Unfortunately, under low-energy CID conditions, permethylated glycans exhibit a labile cleavage when HexNAc residues are on the non-reducing side of a glycosidic bond. As such, cross-ring fragments are not present at levels that are indicative of lactosamine antennal placement. To determine lactosamine antennal position, the purified fractions for Peaks 10 and 11 were subjected to exoglycosidase treatment consisting of a combination of hexaminidase and α-(1-3)-Mannosidase and a combination of hexaminidase and α-(1-6)-mannosidase. Exoglycosidase treatment demonstrated that HILIC Peak 10 represents a FA2G1 structure contains a terminal HexNAc residue adorned to the α-(1-3) core mannose arm, whereas Peak 11 contains a terminal HexNAc residue adorned to the α-(1-6) core mannose arm (data not shown).

Time constraints did not allow for the empirical sequencing of the low-abundance structural isomer co-eluting in Peak 11. Although suggested by preliminary data, it has not been conclusively determined that a species having terminal HexNAc residues extending from a unique tetra-mannosyl fucosylated core co-elutes with FA2G1. For convenience, we have named this structure FM4A2. This unique tetra-mannosyl core structure has been previously reported in biotherapeutic monoclonal antibodies by Ashline, et al. (16). The sequencing of this isomer, as well as other unique glycan species, will be the subject of a subsequent manuscript.

## Evaluation of Non-Human Immunogenic Glycan Species

Today, most monoclonal antibodies are produced in mammalian expression systems, such as Chinese Hamster Ovary (CHO) or Murine Myeloma (NS0) cell lines (68). Although the glycosylation pattern of monoclonal antibodies generated from these expression systems are highly similar to human glycosylation, humans do not synthesize Neu5Gc-containing structures nor α-Gal containing structures (69–71). Indeed, humans may produce an immunogenic response directed against these terminal glycan moieties. For example, α-Gal-containing structures in an Fab glycosylation site of cetuximab are thought to be responsible for elevated



hypersensitivity reactions to this drug (72, 73). Due to potential immunogenicity, special attention is given to the detection and quantitation of Neu5Gc and  $\alpha$ -Gal-containing glycans. Process control strategies may be implemented if appreciable levels of these immunogenic species are detected.

Compositional mapping performed above (Table 5) indicated the presence of five Neu5Gc-containing peaks in the 2-AA data. Targeted verification through sialidase treatment followed by HILIC-F-MS/MS detection confirmed five peaks representing Neu5Gc-containing species, shown as a reduction in fluorescence intensity after treatment in Figure 15. The species assignment was based on intact mass composition data, MS/MS fragmentation data, and knowledge of the biosynthetic glycosylation pathways. Relative quantitation could not be accurately determined from the current data due to the presence of multiple neutral species co-eluting with the Neu5Gc-containing glycans.

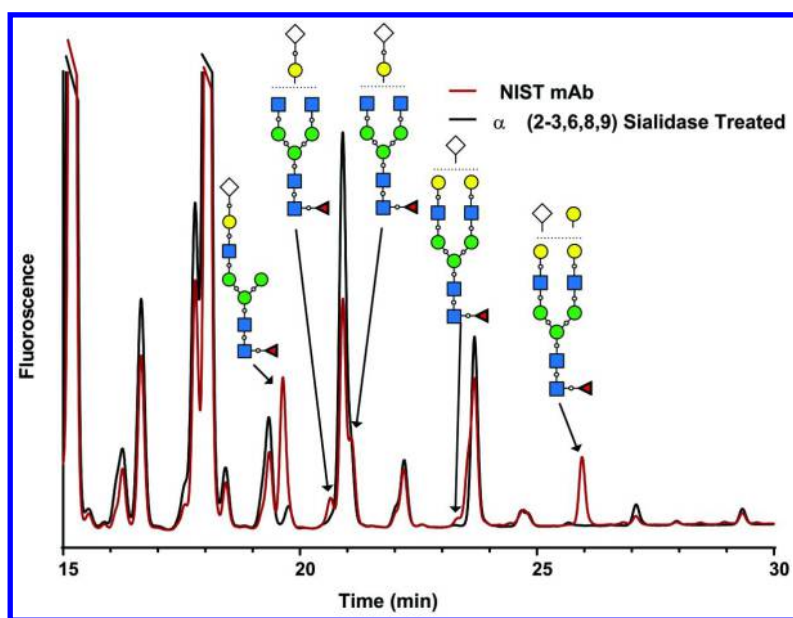


Figure 15. Overlay of zoomed NIST monoclonal antibody (mAb) glycan profiles with and without sialidase treatment demonstrating the presence of *N*-glycolylneuraminic acid (Neu5Gc)-containing species. (see color insert)

More accurate quantitation of the 2-AA-labeled, Neu5Gc-containing species could be achieved by fraction collection and off-line, second-dimension chromatography. As an orthogonal workflow, WAX chromatography on a 2-AB-labeled sample prepared via the high-throughput platform was used for this purpose. Figure 16 depicts the WAX chromatography used to separate and quantitate neutral and charged 2-AB-labeled glycans. Neutral glycans constituted 98.3% of the NISTmAb glycome, whereas glycans with one negative charge (“S1 fraction”) accounted for 1.7% of the glycome.



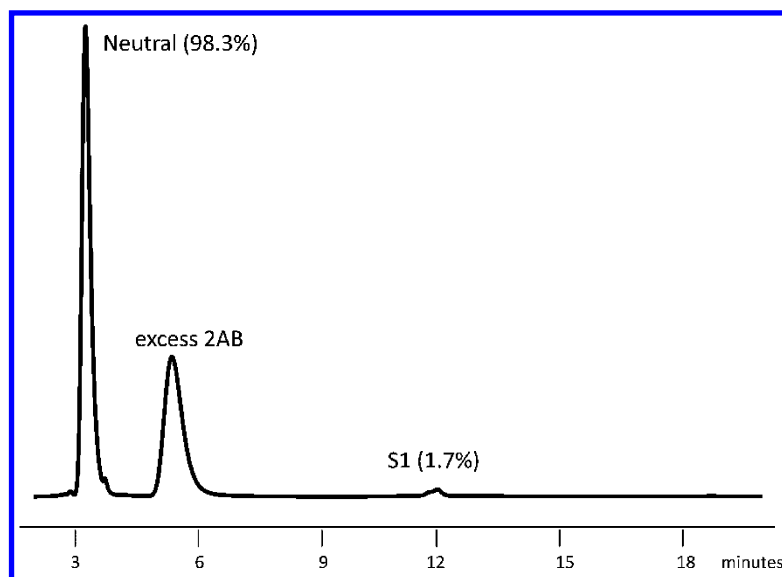


Figure 16. Weak anion exchange (WAX) fractionation of 2-aminobenzamide (2-AB)-labeled NISTmAb glycans.

**Table 8. Identification and Quantitation of N-Glycolylneuraminic acid (Neu5Gc)-Containing Glycan Species Based on Exoglycosidase Treatment of the Weak Anion Exchange (WAX) Chromatography 2-Aminobenzamide-Labeled S1 Fraction**

<i>Proposed Structures<sup>†</sup></i>	<i>Approximate Retention Time</i>	<i>Glucose Units</i>	<i>HILIC % Peak Area of S1 Fraction</i>	<i>% of Total Glycan Population</i>
<b>FA1G1Gc1</b>	8.25	7.93	44.3	0.75
<b>FA2G1Gc1-a</b>	8.651	8.19	5.87	0.10
<b>FA2G1Gc1-b</b>	8.845	8.31	10.53	0.18
<b>FA3G1Gc1</b>	9.42	8.70	3.5	0.06
<b>FA2G2Gc1</b>	9.949	9.07	18.17	0.31
<b>FA3G2Gc1</b>	10.619	9.56	1.74	0.03
<b>FA2G2Ga1Gc1</b>	11.603	9.90	15.89	0.27

HILIC = hydrophilic interaction liquid chromatography. <sup>†</sup> Proposed structures are based on GlycoBase assignment and known glycosylation biosynthetic understanding. Mass spectral analysis was not performed to confirm identification.

To evaluate charged glycans in more detail, the S1 fraction was collected, treated with sialidase, subjected to HILIC separation with fluorescence detection, and glycan species-assigned using GlycoBase (Figure 17). Each of

the Gc-containing peaks identified in Figure 3 were verified utilizing WAX chromatography, including the very low-abundance species FA3G1Gc1 and FA3G2Gc1. In addition, this orthogonal technique allowed resolution of the FA3G1Gc1 peak, which co-eluted with other species in the original HILIC map (Table 1 and Figure 3), allowing specific assignment of percent abundance for this species. Table 8 depicts the seven low-abundance, Neu5Gc-containing species found on the NISTmAb that were confirmed using exoglycosidase digestions in conjunction with GlycoBase data.

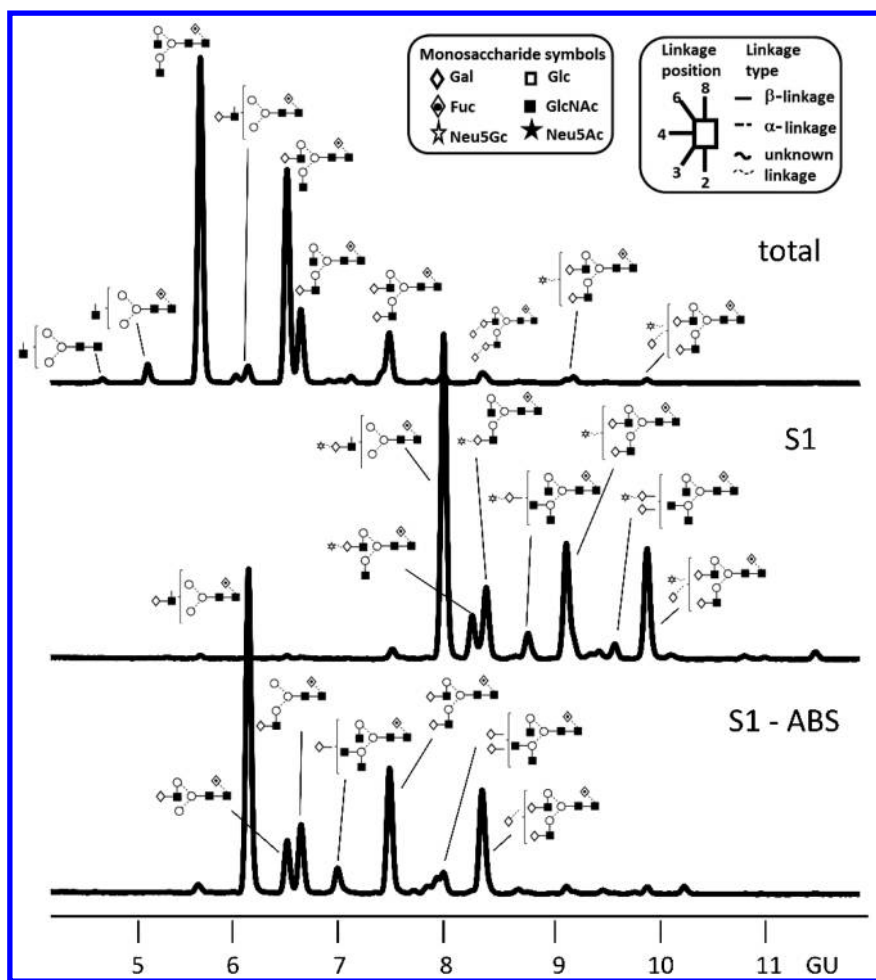


Figure 17. Proposed structural assignments of charged 2-aminobenzamide (2-AB)-labeled glycans from NIST monoclonal antibody (mAb). Top: reference NISTmAb glycome; middle: singly charged glycans (“S1 fraction”); bottom: S1 fraction digested with ABS (*Athrobacter ureafaciens* sialidase).

The NISTmAb 2AA glycans were digested with  $\alpha$ -(1-3,6) Galactosidase to identify any potential glycan structures containing terminal  $\alpha$ -(1-3) galactose. Following digestion with  $\alpha$ -(1-3,6) Galactosidase, the HILIC glycan profile was compared to a control sample. Ten peaks collectively representing 6.96% of the total peak area in the map were detected as glycans with the terminal  $\alpha$ -(1-3) galactose epitope (Figure 18). In this case, no new  $\alpha$ -Gal-containing species were identified, but the data provides increased confidence in the identifications listed in Table 5. Due to the co-elution of neutral, sialylated species and  $\alpha$ -(1-3) galactose-containing species, the percent total peak area (6.96%) is likely an overestimate of  $\alpha$ -(1-3) galactose-containing species. However, orthogonal, second-dimension chromatographic analysis to more accurately quantitate the relative level of  $\alpha$ -(1-3) galactose present on the NISTmAb could be performed as demonstrated for Peak 11 previously.

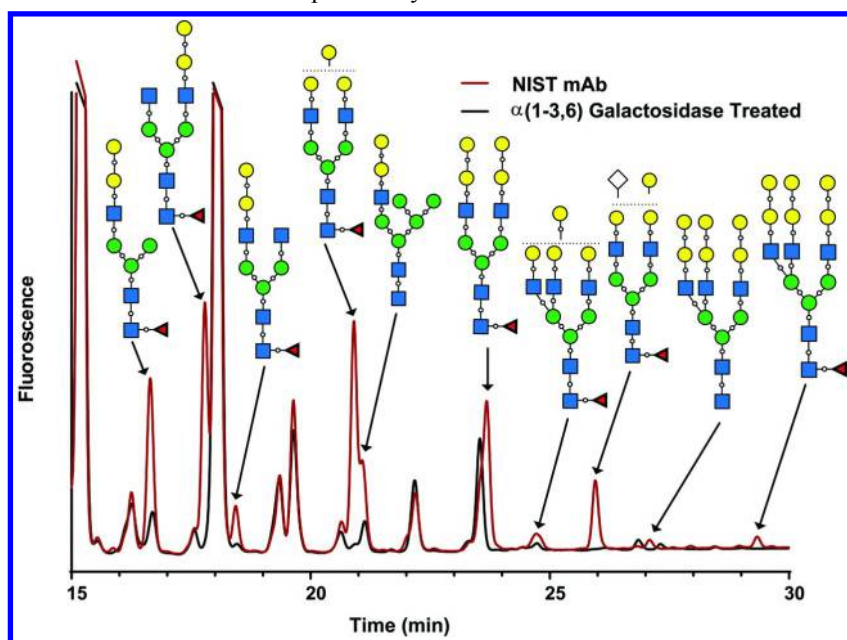


Figure 18. Overlay of NIST 2-aminobenzoic acid (2-AA) glycan profile with and without  $\alpha$ -(1-3,6) Galactosidase treatment. (see color insert)

According to percent relative abundance in the N-glycan profile of NISTmAb, the major species are asialylated, bi-antennary, core-fucosylated species with varying degrees of terminal galactosylation (83.51%), as expected for monoclonal antibodies (Table 9). The second most prevalent type is  $\alpha$ -(1-3)-galactose-capped species (6.96%), followed by asialylated, mono-antennary, core-fucosylated species (4.81%). Modest levels of tri-antennary species (4.37%) and Neu5Gc-containing species (3.63%) are present, and low levels of afucosylated (1.09%) and high mannose species (1.04%) exist. These relative percentage grouping values are subject to variation depending on the chromatographic condition used.

**Table 9. NIST IgG1 Glycan Grouping**

<i>Glycan Classification Groups</i>	<i>% Glycosylation</i>
<b>Asialylated, bi-antennary core fucosylated complex (A2G<sub>x</sub>F, where x = 0, 1, 2)</b>	83.51
<b>Asialylated, mono-antennary core fucosylated hybrid (A1G<sub>x</sub>MyF, where x = 0, 1 and y = 3, 4, 5)</b>	4.81
<b>Tri-antennary glycans</b>	4.37
<b>High mannose (M<sub>x</sub>, where x = 5, 6, 7, 8)</b>	1.04
<b>Afucosylated Complex and hybrid</b>	1.09
<b>Terminal β-(1-4)-galactosylated glycans</b>	52.17
<b>Terminal α-(1-3)-galactosylated glycans</b>	6.96
<b>Neu5Gc-containing glycans</b>	1.70 <sup>†</sup>

<sup>†</sup> Relative quantitation of *N*-glycolylneuraminic acid (Neu5Gc)-containing glycans is based on the relative percentage of the weak anion exchange (WAX) chromatography singly charged acidic fraction.

## LC-FLD Glycosylation Mapping

Glycosylation mapping using fluorescent labeling of *N*-glycans along with a liquid chromatography (LC)-based analytical method was a common workflow component of each participating laboratory. Fluorescent labeling of glycans is a stepwise process involving multiple chemical derivations and purifications, which can result in minor sample preparation artifacts presented as low-abundance peaks in fluorescence and total ion current chromatograms. During the development and commercialization of biotherapeutics, the detection and analysis of minor peaks is an expectation (74). Considering the widespread application of LC-FLD analysis, an additional discussion of minor artifacts and sample preparation perturbations that may influence their appearance is warranted.

A low-abundance sample consistent with dehydrolysis of the labeled glycans was observed as a pre-peak fronting the major species FA2, FA2G1a, and/or FA2G2 in the fluorescence chromatogram. This artifact is denoted as an asterisk in Figure 3 and labeled peak 5 in Figure 9. When using mass spectrometric detection, dehydrolysis artifacts yield a mass value consistent with the loss of water (−18 Da) from the putative species. For example, Peak 5 exhibits an observed compositional mass consistent with the FA2 species less a water, *m/z* 1565.72, compared to the putative FA2 species in Peak 6, exhibiting an observed compositional mass of *m/z* 1583.74 (Table 5). The inclusion or exclusion of these sample artifacts must be considered when quantifying glycan species. In addition, the artifact may co-elute with low-abundance putative glycan species, leading to failure to detect low-abundance species. Further optimization of labeling conditions during method development may help reduce this common preparation artifact. However, all three labs observed some degree of this artifact using their

platform method as demonstrated in Figure 3 (Appendix A methods), Figure 9 (Appendix B methods), and Figures 19 and 20 (Appendix C methods), discussed below.

Sample preparation artifacts may only be detectable using mass spectrometric detection, and special consideration must be taken when mass spectral analysis of labeled glycans is undertaken. Figure 19 depicts the MS base peak chromatogram and fluorescence signals from an LC-MS experiment performed on 2-AB-labeled glycans released from the NISTmAb using the Appendix C protocols. Figure 19a shows the 2-AB MS trace containing a set of three doublet peaks, labeled nG0F/nG0F\*, nG1F/nG1F\*, and nG2F/nG2F\*, which do not correspond to fluorescent peak intensity. Further inspection of the parent mass indicates each of these values—1464.58 Da, 1626.63 Da, and 1788.69 Da, respectively—is exactly one Da greater than expected for unlabeled, enzymatically released FA2, FA2G1, and FA2G2 glycan species. This one-Da shift corresponds to a fully reduced glycamine formation at the reducing terminus. Interestingly, glycamines do not contain an anomeric center, and, therefore, should co-elute as a single peak. This doublet phenomenon can be explained by the base-catalyzed epimerization of the reducing terminal monosaccharide at the C2 position from GlcNAc to ManNAc. Such epimerization has been reported due to slightly basic conditions during PNGase release (75, 76), however, injection of unlabeled reduced glycans from the same preparation showed only free glycans (data not shown). The data therefore indicate the glycamine species are formed as a byproduct of the labeling reaction (43). Interestingly, resolution of epimers is not observed for the labeled peaks, although, undoubtedly, terminal ManNAc and GlcNAc are labeled if they are being formed. It is likely that the label dominates chromatographic selectivity after labeling and causes co-elution of these isobaric species under the HILIC conditions used for this experiment.

The identification of unlabeled and glycamine forms such as those observed in the current experiment would go unnoticed without the addition of MS. It should be noted that 2-AB labeling at 37°C—as opposed to 65°C—still showed the doublet of the unlabeled glycamine peaks. The 2-AA labeling protocol also yielded unlabeled glycamine products. However, in this case, the earlier eluting isomer greatly predominates as opposed to a roughly equimolar content. Both 2-AB and 2-AA samples underwent the same protocol prior to labeling, indicating the use of methanol with boric acid/sodium acetate in the 2-AA protocol limits epimerization (77). Further experiments will be required to fully evaluate this artifact.

The pursuit of labeling conditions that maximize incorporation of label (reaction efficiency) and/or extraction protocols to remove unlabeled materials have been the topic of many studies (78–82). Ultimately, these residual species represent a potential difficulty with identifying low-abundance, labeled glycans via mass spectrometry as they co-elute with low-abundance fluorescence peaks and may suppress ionization. C18 solid phase extraction was added after the HILIC cleanup in the current report to remove unlabeled glycan species (Figure 19B). These plots clearly demonstrate efficient removal of unlabeled glycamines, and resulted in higher incidence of quality MS<sup>2</sup> spectra for low-abundance species for compositional assignment.

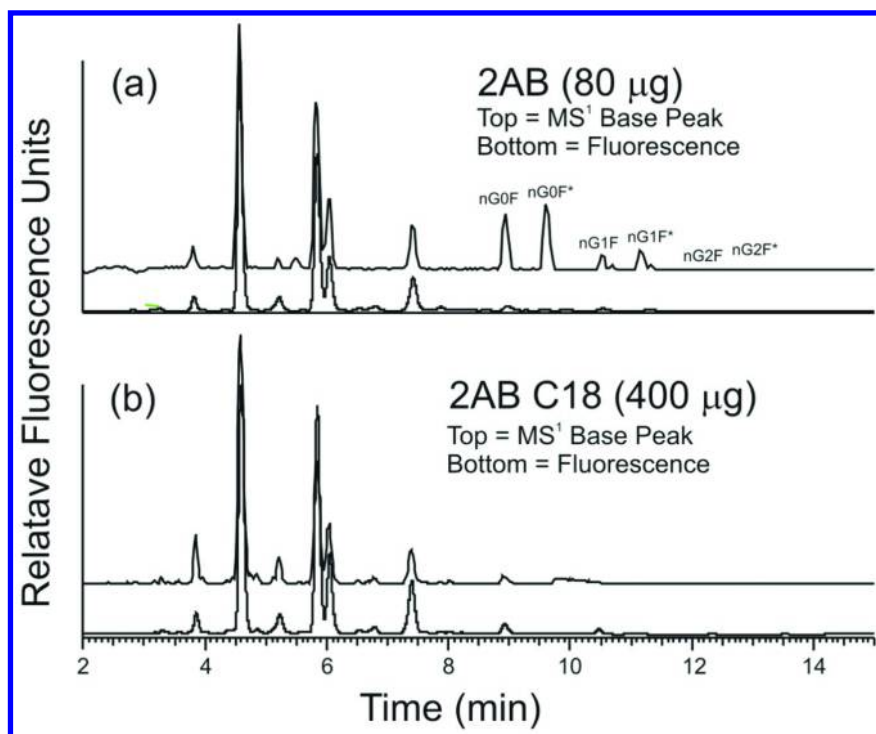


Figure 19. Overlay of fluorescence and mass spectrometry (MS) signals for NIST monoclonal antibody (mAb) glycoanalysis using 2-aminobenzamide (2-AB) label (Appendix C) with (a) and without (b) post-labeling C18 solid phase extraction.

As mentioned previously, numerous labels, labeling conditions, and cleanup protocols were examined in an effort to demonstrate a variety of factors that can contribute to the glycoanalytical profiles. Figure 20 depicts the fluorescence trace from representative LC-FLD-MS/MS analysis under a number of these conditions. All labeling conditions demonstrated some level of unlabeled glycamine species. A comparison of Figures 20a and 20c demonstrates the difference in glycoprofile obtained under the current conditions with and without the orthogonal C18 cleanup stage. While unlabeled glycamine species were removed via C18, the apparent relative abundance of the glycoprofile is also altered. A significant decrease in the apparent percent relative abundance of Neu5Gc-containing peaks was noted, as demonstrated by a reduced apparent relative abundance of peaks FA1G1Gc1, FA2G2Gc1, and FA2G2Ga1Gc1. Each of the affected peaks are Neu5Gc-containing peaks, indicating that the additional carboxylic acid resulted in weaker retention of sialylated glycans. The same trend can also be seen by comparing profiles for 2-AA-labeled glycoprofiles (data not shown) with and without C18 cleanup (Neu5Gc-peaks FA2G1Gc1-a, FA2G1Gc1-b, and M4A1G1Gc1 also visibly altered in 2-AA data). Therefore, although additional C18 offers benefits for compositional identification via mass spectrometry, one must assess and/or optimize the protocol and any effect it may

have on the apparent glycoprofile. In the current example, C18 cleanup along with larger injection volumes was used for confident identification via MS/MS spectra. Further analysis of the relative abundance of the glycoprofile was then made based on profiles without the C18 step in order to minimize sample handling artifacts.

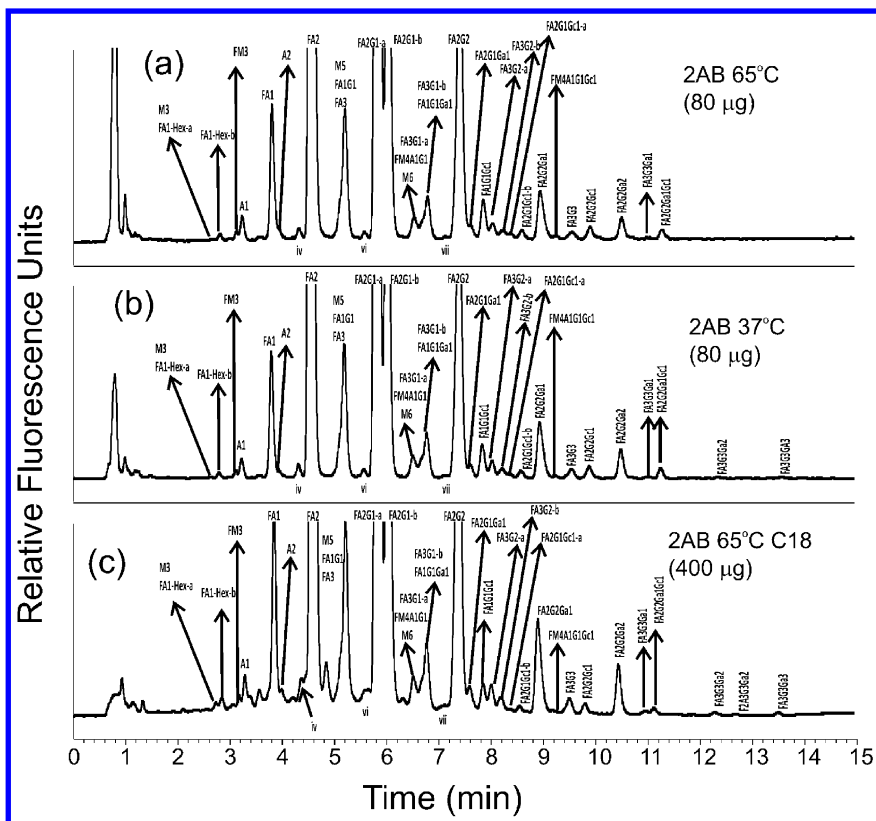


Figure 20. Liquid chromatography-mass spectrometry (LC-MS) analysis using the Appendix C labeling and analysis platform (a) at 65°C; (b) at 37°C; and (c) at 65°C with an additional C18 extraction step. Expanded fluorescence signal is shown to visualize low-abundance species, with glycan identification based on MS and tandem mass spectrometry (MS/MS). Roman numerals represent dehydrolysis artifacts discussed above.

The profiles in Figure 20a and 20b show labeling with 2-AB under different temperatures. The profiles are quite comparable, with consistent retention times and relative abundance of the individual species. Labeling temperature is a critical factor to consider for throughput and potential loss of labile monosaccharides (e.g., sialic acids) (82). In the current sample, loss of sialic acids due to elevated labeling temperature does not seem to be a factor. All Neu5Gc-containing peaks retain consistent relative abundances at each temperature. Therefore, 65°C

labeling was chosen for further optimization due to the increased throughput. It should be noted that the 2-AB, 65°C chromatogram does not have peaks for FA3G3Ga2 and FA3G3Ga3 labeled. These peaks are very low abundance and were readily detected with slightly larger injection volumes. Slight variability in sample preparation likely results in their detectability in only one of the two samples at the lower injection level (from 80 µg of mAb). Using the current gradient and column, it was demonstrated that glycans released from up to 400 µg of mAb could be injected without significantly affecting resolution.

As described above, the use of LC-FLD alone is a powerful technique to match glycan composition with relative elution times based on the GU unit system. Recently, the development of fluorescence detector flow cells capable of higher pressures has allowed the coupling of this technique directly with orthogonal detection via mass spectrometry. Mass accurate  $m/z$  measurement provides an orthogonal measurand for glycan identification; however, careful optimization of mass spectrometer source conditions and data interpretation must be considered as in-source fragmentation can commonly occur with released glycans and their derivatives. In addition, glycans can ionize as a variety of adducts and neutral exchanges (e.g.,  $H^+$ ,  $Na^+$ ,  $K^+$ ,  $NH_4^+$ ). For example, in the current experiment, most glycans ionized predominantly as  $[M+H]^+$  or  $[M+2H]^{2+}$  species. However, larger glycans showed increased ionization with  $NH_4^+$  adducts. In fact, the tri-antennary species FA3G3Ga2, F2A3G3Ga2, and FA3G3Ga3 in Figure 20 ionized predominantly as  $[M+NH_4+H]^{2+}$ . This shift in ionized species is thought to be due to a combination of glycan size (and thus more available  $-OH$  for neutral exchange) as well as the higher concentration of ammonium formate in the late stages of gradient elution (ammonium formate in solvent A only). The use of glycan fragmentation via MS/MS also affords increased confidence in glycan compositional identification. For example, high abundance B3 ions can be useful in confirming retention-based identification of isobaric species such as FA2G2 versus FA2G1Ga1 or FA2G1Gc1 versus A2G2Ac1.

Selection of the label for compositional glycoprofiling (e.g., 2-AB, 2-AA) is also a consideration when implementing glycoprofile mapping. The Appendix C analysis methods were used to compare the NISTmAb glycan profile using 2-AA and 2-AB. The 2-AA-labeled glycans each showed a shift to slightly earlier retention time but retained the same general elution order, as expected. Under the Appendix C gradient conditions, 2-AA seemed to offer additional selectivity for the resolution of peaks eluting at approximately 5 minutes. This is a potentially important advantage as high-mannose glycans (e.g., M5) have been associated with reduced ability to elicit complement-dependent cytotoxicity effector functions, and may represent a CQA of interest during process development and product quality testing. Ultimately, a variety of factors will be considered in selecting a label for a given product, including ability to tailor chromatographic resolution of critical glycans, MS compatibility, polarity-specific ionization efficiency, and amenability to downstream platforms (44, 51, 78, 80, 82). For example, 2-AA may be chosen if linkage analysis will be performed via permethylation and mass spectrometry; 2-AB may be preferable, however, for downstream WAX methods due to the additional charge on 2-AA. As discussed in more detail below, it was found that a variety of LC methods, mass spectrometers,



and labels (2-AA or 2-AB) can be utilized along with appropriate interpretation to achieve consistent glycoprofiling results that are adequate for monitoring glycosylation. The sample-specific method evaluation above highlights a select few of the potential variables to consider when developing and implementing glycan mapping methodologies to monitor glycosylation for process support and product characterization workstreams. These variables may include label identity, labeling reaction conditions, solid phase extraction steps, and chromatography and mass spectrometry instrument setup.

## Interlaboratory Comparison

In the previous sections, related methodology with variations in sample preparation, analytical methodology, and interpretation approach was utilized by three laboratories to evaluate the compositional glycoprofile of the NISTmAb. Although LC-FLD and/or LC-FLD-MS/MS cannot alone delineate every aspect of glycan structural heterogeneity (as demonstrated through the previous discussion), it is one of the most common techniques (in addition to CE-laser-induced fluorescence [CE-LIF] glycoanalysis) likely to be presented in a regulatory filing. Differences between each approach in the current chapter included injection quantity, reducing terminus label, and mobile phase composition and gradient, although each laboratory used the same column chemistry. The differential selectivity of the mobile phases resulted in slight variations in the isomeric species identified in each laboratory. This can be clearly shown by comparison of peaks 16, 17, and 18 in Figure 9 compared to the FA2G2 peak in Figure 3. The 2-AA method (Figure 9) resulted in resolution of three isobaric species, whereas the 2-AB method resulted in only one species identified. Lab 3 (Figure 21) utilized the same gradient as that which produced the chromatography shown in Figure 3, except that Lab 3's mobile phase B contained no ammonium formate. This relatively small change in chromatographic conditions resulted in resolution of the FA2G2 and one FA2G1Gal isomer. On the other hand, selectivity with ammonium formate in mobile phase B (Figure 3) resulted in clear resolution of FM5A1 from M6 and FA3G1, which co-eluted in the Lab 3 data. Collectively, this shows the effect of chromatographic selectivity with regard to glycan separations and the need to tailor quality control methods for the detection and analysis of specific glycans considered to be CQAs.

Despite minute differences in selectivity between analytical methods utilized, a qualitative comparison of the various LC-FLD profiling methods can be made. The Venn diagram in Figure 21 compares the putative structures resulting from each laboratory's qualitative analysis.

The sum of unique glycans identified by each laboratory represent less than 1% of the total glycan-containing peak fraction, and each individual unique glycan represents at most 0.2% of the total relative abundance. These differentially identified glycans probably represent an artifact of co-elution of species due to differential selectivity of the chromatographic methods and analysis approach utilized by each laboratory. The qualitative results indicate that the even in the absence of harmonized sample preparation or analytical method, very highly

comparable results can be obtained with LC-FLD-MS analysis of released glycan fractions, indicating such a method is a robust platform for glycan analysis. Further harmonization of analytics would likely result in near complete agreement (both qualitative and quantitative), and will be the topic of a future round-robin characterization of the NISTmAb. It is expected that the NISTmAb will serve as a representative reference material, useful for evaluating current and emerging glycoanalytical methods (21, 22).

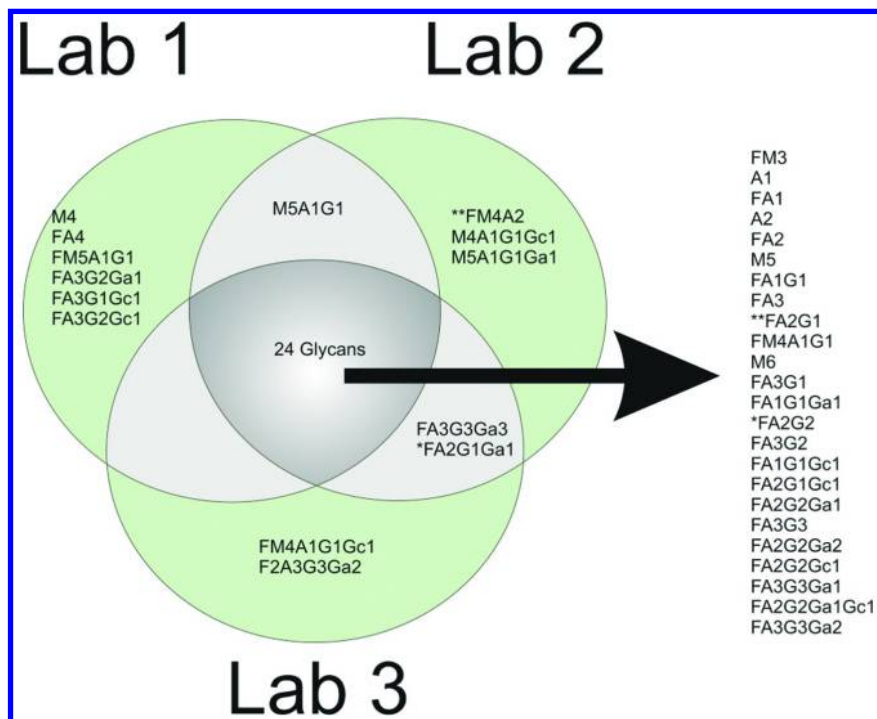


Figure 21. Qualitative comparison of compositional identifications made during glycoprofiling. For clarity of presentation, the *-a* and *-b* isomers corresponding to linkages to differential core antennae were summed as one composition, while other isomeric pairs are labeled with an asterisk. (see color insert)

## Conclusions

Using the NISTmAb reference material as a surrogate biotherapeutic, this chapter described high-throughput glycan profiling using automated sample preparation and HILIC with fluorescence detection to monitor changes in N-glycosylation for process support. For recombinant mAb biotherapeutics, the collection of N-linked glycan species is relatively constant, albeit the relative proportions of the glycan species may differ greatly. This consistent population of glycan species justifies the development of a single-platform, glycan-profiling assay to support a mAb biotherapeutic pipeline.

With advancements in column resin chemistries, UPLC systems, highly sensitive fluorescence detectors, and high resolution/accurate mass spectrometers, glycan species can be readily sequenced. A comprehensive, N-linked glycan characterization workflow purposed for biological license application of mAbs was outlined in this chapter. Although additional characterization is needed to fully identify and sequence the entire NISTmAb glycome, the example highlighted in this chapter outlines a workstream that easily sequences glycans present on monoclonal antibodies, including low-abundance, atypical glycan species.

Glycoanalysis is a unique analytical challenge due to the heterogeneity and non-linearity of N-glycan structures, as well as the extensive sample handling required for analysis. Glycoanalytical standards are of critical importance in evaluating method comparability and commutability. Released and/or labeled glycan standards are quite useful for assessing instrument robustness and reproducibility, however, the complete system suitability includes all operations of the analytical system, including sample preparation and handling. For this reason, the intact NISTmAb reference material described in this chapter affords an additional supplement to in-house reference standards for such system suitability, operator training, and analytical method evaluation for glycoanalytical profiling.

## Acknowledgments

The authors would like to acknowledge the support of the EU FP7 program HighGlycan, Grant No. 278535 and Science Foundation Ireland (08/SRC/B1393, Alimentary Glycoscience Research Cluster [AGRC]) for funding this work.

## NIST Disclaimer

Commercial equipment, instruments, and materials are identified in this paper to adequately exemplify the discussion and experimental procedure. Such identification does not imply recommendations or endorsements by NIST nor does it imply that the equipment, instruments, or materials are necessarily the best available for the purpose.

## Appendix A: Methods for High-Throughput N-Linked Glycan Profiling for Process Support

NISTmAb glycans were released, fluorescently labeled with either 2-aminobenzoic acid (2-AA) or 2-aminobenzamide (2-AB), and purified on a liquid handling workstation, as described earlier (26). Briefly, NIST monoclonal antibody (mAb) was reduced and alkylated, and excess reagents were removed by ultrafiltration on a high-throughput ultrafiltration plate. N-linked glycans were released using peptide N-glycosidase F (PNGase F) and recovered from the plate by centrifuge ultrafiltration. Next, glycans were subjected to hydrazide-mediated cleanup, labeled with 2-aminobenzamide, and purified by automated solid phase extraction.

## Appendix B: Methods for Detailed Glycan Characterization Workflow for Marketing Application

### Materials

Peptidase N-glycosidase F (PNGase F), sialidase A,  $\beta$ -(1-4)-galactosidase, and  $\beta$ -N-hexosaminidase was purchased from Prozyme (Hayworth, CA).  $\alpha$ -(1-3,6)-galactosidase was purchased from QA-Bio. Proteomics-grade trypsin was purchased from Roche (Basel, Switzerland). The 2-aminobenzamide (2-AB), 2-aminobenzoic acid (2-AA), ammonium hydroxide, anhydrous dimethylsulfoxide (DMSO), dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA), small-bead sodium hydroxide, iodomethane, liquid chromatography-mass spectrometry (LC-MS)-grade methanol, and sodium cyanoborohydride were purchased from Sigma (St. Louis, MO). Guanidine hydrochloride was purchased from VWR (West Chester, PA), 1 mol/L Tris buffer at pH 7.5 from TEKnova (Hollister, CA), trifluoroacetic acid (TFA) from Pierce Thermo Scientific (Rockford, IL), and formic acid from EMD Millipore (Billerica, MA). Boric acid was purchased from Alfa Aesar (Ward Hill, MA), and sodium acetate-trihydrate and glacial acetic acid from Avantor (Center Valley, PA). MacroSpin columns were purchased from Harvard Apparatus (Holliston, MA). Normal phase PhyTips were purchased from PhyNexus (San Jose, CA). NAP-5 desalting columns were purchased from GE Healthcare (Pittsburgh, PA). Burdick and Jackson LC-MS-grade water and acetonitrile were obtained from VWR (West Chester, PA). The SepPak tC18 SPE cartridges and Waters Glycan column were obtained from Waters (Milford, MA). All references to water and acetonitrile throughout this chapter refer to LC-MS grade. All other chemicals used were of reagent grade or better.

### Trypsin Digestion Used for N-Site Determination and Occupancy

Deglycosylated NIST monoclonal antibody (mAb) was denatured and reduced for 30 min at 25°C in the presence of 7.5 mol/L guanidine HCl, 0.25 mol/L Tris, 2 mmol/L ethylenediaminetetraacetic acid (EDTA), and 10 mmol/L DTT. The reduced sample was alkylated with 22 mmol/L iodoacetic acid at 25°C for 30 min in the dark. The alkylation reaction was quenched with 0.5 mol/L DTT. Prior to digestion, samples were desalted and buffer exchanged into 0.1 mol/L Tris, pH 7.5, digestion buffer using a NAP-5 gel filtration column (GE Healthcare). The reduced and alkylated sample was incubated with 1:10 (w/w, enzyme:protein) proteomics-grade trypsin (Roche, Basel, Switzerland) at 37°C for 30 min. The digestion was quenched with the addition of TFA.

### Reversed-Phase Chromatography/Mass Spectrometry for N-Site Determination and Occupancy

An Acquity ultra-performance liquid chromatography (UPLC) system comprised of a binary gradient solvent manager, an autosampler capable of maintaining 4°C, a column compartment capable of maintaining 50°C, and a UV detector set at 215 nm was coupled to an Orbitrap Velos Pro Ion Trap Mass Spectrometer (Thermo Scientific; Waltham, MA) with a heated electrospray ionization source (HESI). 2  $\mu$ g of the digested samples were injected onto a

Waters BEH130 C18 column (2.1 × 100mm, 1.7 μm) set at a constant temperature of 50°C (± 2°C). The mobile phases used consisted of solvent A = 0.1% formic acid and solvent B = 0.1% formic acid (v/v) in acetonitrile. Following injection, reagents were eluted at 0.250 ml/min for 5 min at a composition of 1% mobile phase B, then 10% mobile phase B gradually increasing to 40% mobile phase B over 75 min, followed by 80% mobile phase B for 15 min, and then finishing with 1% mobile phase B for 15 min. The ion trap mass spectrometer coupled to the UPLC was set up with data-dependent acquisition enabled.

### **Glycan Release and Fluorescence Labeling for Product Characterization**

100 μg of NISTmAb were added to 6 μL of reaction buffer (250 mmol/L sodium phosphate, pH 7.5) and 3.7 μL of LC-MS-grade water. 6 μL of PNGase F were added to sample mixture. Samples were then incubated for 120 min (± 2 min) at 37°C (± 2°C) in a water bath. 20 μL of a 30 mg/mL solution of 2-aminobenzoic acid prepared in an acetate-borate-buffered solution (4% sodium acetate trihydrate [w/v], 2% boric acid [w/v]) and 30 μL of methanol were added to the sample digest. The sample was incubated at 65°C (± 2°C) for 180 min (± 5 min). Samples were cooled and then centrifuged. Excess dye was removed using the PhyNexus MEA system and Normal Phase PhyTips (83). Samples were desalted on a tC18 SPE cartridge to remove excess salts and unlabeled glycans. The tC18 cartridge was washed with five sequential additions of 1 mL methanol. The cartridge was then re-equilibrated with five sequential additions of 1 mL 0.1% (v/v) TFA in water. The labeled sample was loaded onto the cartridge and washed with 5 sequential 1 mL volumes of 0.1% (v/v) TFA in water. The labeled glycans were eluted with 1.5 mL of 50% (v/v) acetonitrile containing 0.1% (v/v) TFA. Samples were then dried to completion in a centrifugal evaporator and stored at -20°C until analysis.

### **Hydrophilic Interaction Liquid Chromatography/Mass Spectrometry for Product Profiling, Characterization, and Analysis**

An Acquity UPLC comprised of a binary gradient solvent manager, an autosampler capable of maintaining 4°C, a column compartment capable of maintaining 50°C, and a fluorescence detector was coupled to an Advion Triversa Nanomate® set in Fraction Collection/Infusion mode, which directly infused the sample into an LTQ Velos Pro Ion Trap Mass Spectrometer (Thermo Scientific; Waltham, MA). Dried, labeled sample was reconstituted in 15 μL of 50% acetonitrile (v/v) in water and 2 μL of sample were injected. The excitation and emission parameters were 360 nm and 425 nm, respectively. Separation was conducted on a Waters Glycan column (100 mm × 2.1 mm i.d., particle diameter 1.7 μm) at 50°C. The mobile phases used consisted of solvent A = 100 mmol/L ammonium formate, pH 3.0, and solvent B = LC-MS-grade acetonitrile. The gradient was delivered at a constant flow rate of 0.5 ml/min with a starting composition of 22% mobile phase A and increasing to 50% mobile phase A over 50.40 min, followed by 100% mobile phase A for 2.5 min at a 0.4 mL/min flow rate, and then finishing with 22.0% mobile phase A for 7 min at a flow rate of 0.5 mL/min. The linear ion trap mass spectrometer was set with data-dependent acquisition enabled, an isolation width window set at 2 *m/z*, collision-induced dissociation set at 35 and with the activation Q at 0.250 for

100 ms. The advantage of using the Nanomate® in fraction collection/infusion mode is that it is collecting fractions while simultaneously nano-infusing a small portion of the sample for mass detection. Five rounds of fraction collection was undertaken and like fractions pooled and dried. The collected Hydrophilic Interaction Liquid Chromatography (HILIC) fractions were then subjected to off-line reversed-phase/MS.

### **Off-line, Second-Dimension Reversed-Phase/Mass Spectrometry for Product Characterization**

An Acquity H-Class comprised of quaternary pump solvent manager, autosampler capable of maintaining 4°C, column compartment capable of maintaining 50°C, and fluorescence detector was coupled to an LTQ Velos Pro Ion Trap Mass Spectrometer (Thermo Scientific; Waltham, MA). The collected HILIC fractions were reconstituted in 15  $\mu\text{L}$  of 100% LC-MS-grade water and 2  $\mu\text{L}$  of sample injected. The excitation and emission parameters were 360 nm and 425 nm, respectively. Separation was conducted on a Waters BEH130 C18 (100 mm  $\times$  2.1 mm i.d., particle diameter 1.7  $\mu\text{m}$ ) at 50°C. The mobile phases used consisted of solvent A = 0.1% formic acid and solvent B = 0.1% formic acid (v/v) in acetonitrile. Following injection, reagents were eluted (0.3 mL/min) with a 2-min isocratic composition flow at 6% mobile phase B, then a gradient was delivered with a starting composition of 6% mobile phase B and increasing to 14.4% mobile phase B over 28 minutes, followed by 95% mobile phase B for 3.0 minutes, and then finishing with 6% mobile phase B for 8 minutes. The linear ion trap mass spectrometer was set with data-dependent acquisition, an isolation width window set at 2  $m/z$ , collision-induced dissociation set at 35 with the activation Q at 0.250 for 100 ms.

### **Desalting and Permethylation of HILIC Fractions for Product Characterization**

The tC18 cartridge was washed with five sequential additions of 1 mL methanol. The cartridge was then re-equilibrated with five sequential additions of 1 mL 0.1% (v/v) TFA in water. Individual pooled HILIC fractions were loaded onto the cartridge and washed with 5 sequential 1 mL volumes of 0.1% (v/v) TFA in water. The glycan fractions were eluted with 1.5 mL of 50% (v/v) acetonitrile containing 0.1% (v/v) TFA. Glycan fractions were then dried to completion in a centrifugal evaporator and stored at -20°C until analysis. The desalted fractions were permethylated according to a previously reported method (84) that was slightly optimized for 2-AA labeled glycans. Briefly, samples were resuspended in 60% DMSO, 37.2% iodomethane, and 2.8% water. To reduce partial permethylation, 15 sample recycles, as opposed to 8 sample recycles, were used. Dried samples were reconstituted in 75% methanol.

### **Direct Infusion Nanospray Sequential Mass Spectrometry (NSI-MS<sup>n</sup>) of Permethylated HILIC Fractions for Product Characterization**

Sequential mass spectra were obtained from a linear ion trap mass spectrometer (Thermo LTQ Velos Pro) equipped with a Triversa Nanomate® (Advion, Ithaca, NY). Signal averaging was accomplished through adjustment of

the number of scans relative to the ion signal strength. Isolation window width was set at 1  $m/z$ . Collision parameters were left at default values with collision energy set at 35, activation Q set at 0.250, and activation time set at 30 ms.

### **Weak Anion Exchange (WAX) Instrumental Method for Characterization of Glycans According to Charge**

WAX-HPLC was performed on a 2795 Alliance Separation module with a 2475 fluorescence detector (Waters, Milford, MA), equipped with a Prozyme GlycoSep C 7.5 mm  $\times$  75 mm column (Prozyme, Leandro, CA). Solvent A was 20% (v/v) acetonitrile in water, and solvent B was 0.1 M acetic acid adjusted to pH 7.0 with ammonia solution in 20% (v/v) acetonitrile. The following elution gradient conditions were used: 100% A for 5 min, then a linear gradient of 100% to 0% A for 15 min at a flow rate of 0.75 mL/min, followed by 0% A for 2.5 min, returning to 100% A for 1.5 min and then finishing with 100% A for 7 min.

## **Appendix C: Method for LC- Fluorescence Detection (FLD)-MS/MS Method Development Sample Considerations**

### **Materials**

GlykoPrep H and CU solid phase extraction columns were purchased from Prozyme (Hayward, CA). All other materials used are identical to those listed in Appendix B.

### **Glycan Release**

500  $\mu$ g of NIST monoclonal antibody (mAb) (50  $\mu$ L) were mixed with 14  $\mu$ L of 5X reaction buffer and 6  $\mu$ L of peptide N-glycosidase F (PNGase F). Samples were then incubated for 120 min ( $\pm$  2 min) at 37°C ( $\pm$  2°C) in a water bath. GlykoPrep H cartridges were then used to purify released glycans from remaining apoprotein. Cartridges were sequentially washed with 150  $\mu$ L of water, 150  $\mu$ L 50% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid (TFA), and 300  $\mu$ L of water. The released glycan sample was then added to the cartridge and rinsed with an additional 300  $\mu$ L of water. Glycans were eluted using 150  $\mu$ L of 50% (v/v) acetonitrile with 0.1% (v/v) TFA. Samples were then dried to completion in a centrifugal evaporator and stored at  $-20^\circ\text{C}$  until labeling.

### **2-Aminobenzoic Acid Labeling**

2-Aminobenzoic Acid (2-AA) labeling solution consisted of 30 mg/mL 2-AA dissolved in 4% (w/v) sodium acetate and 2% (w/v) boric acid in methanol. Directly before labeling, 6.3 mg of sodium cyanoborohydride ( $\text{NaCNBH}_3$ ) were added to 200  $\mu$ L of the 2-AA labeling solution. Glycans from 500  $\mu$ g of NISTmAb were reconstituted in 20  $\mu$ L of the 2-AA/ $\text{NaCNBH}_3$  solution and incubated for 65 min at 80°C. This solution was directly applied to GlykoPrep CU cartridges for purification of labeled glycan as discussed below.

## 2-Aminobenzamide Labeling

Labeling of each glycan sample (released from 500  $\mu\text{g}$  of NISTmAb) with 2-Aminobenzamide (2-AB) was performed by reconstituting dried glycans in 10  $\mu\text{L}$  of acetic acid/DMSO solution (3/7 v/v) containing 5 mg 2-AB/100  $\mu\text{L}$  and 6 mg  $\text{NaCNBH}_3$ /100  $\mu\text{L}$ . Samples were allowed to incubate at 65°C for 2 hours or 37°C for 17 hours. Each sample was then diluted with 190  $\mu\text{L}$  of acetonitrile and purified using GlykoPrep CU cartridges for purification of labeled glycan as discussed below.

## GlykoPrep CU Cleanup

Cartridges were washed with 150  $\mu\text{L}$  of 97% acetonitrile. 2-AA- or 2-AB-labeled sample was then added to the cartridge and rinsed with 300  $\mu\text{L}$  of 97% (v/v) acetonitrile in water. Labeled glycans were eluted using 150  $\mu\text{L}$  of water, dried to completion in a centrifugal evaporator, and stored at -20°C until analysis. An additional C18 cleanup was performed (only where indicated below) according to the protocol listed in the glycan release and fluorescence labeling for product characterization sections above.

## Instrumental Method

A Dionex UltiMate 3000 ultra-high-performance liquid chromatography (UHPLC) system with fluorescence detection was coupled to an Orbitrap Elite mass spectrometer with heated electrospray ionization source (HESI-II) (Thermo Scientific, Waltham, MA). Dried 2-AB or 2-AA samples were reconstituted in 250  $\mu\text{L}$  of 80% acetonitrile. The samples were injected in volumes corresponding to glycans released from 80  $\mu\text{g}$  or 400  $\mu\text{g}$  of NISTmAb. Separation was conducted on a Waters Glycan column (150 mm  $\times$  2.1 mm i.d., particle diameter 1.7  $\mu\text{m}$ ) at 0.561 mL/min and 60°C. A binary gradient was used for the analytical separation consisting of solvent A = 50 mmol/L ammonium formate (pH 4.4) and solvent B = 0.1% formic acid (v/v) in acetonitrile. The gradient was delivered with a curve factor of 6 and compositions of 70% B at 0 min, 70 % (v/v) B at 1.47 min, 55% (v/v) B at 15 min, 30% (v/v) B at 15.5 min, 30% (v/v) B at 16.25 min, 70% (v/v) B at 16.55 min, and 70% (v/v) B at 18.5 min. Flow rate was set at 0.561 mL/min at all time points except at 16.25 min and 16.55 min when the flow rate was set to 0.300 mL/min. Eluent was monitored at  $\lambda_{\text{excitation}} = 330$  nm (2-AB) or 360 nm (2AA),  $\lambda_{\text{emission}} = 420$  nm (2-AB) or 425 nm (2-AA) and then directed to the Orbitrap Elite. The mass spectrometer was set to collect in a data-dependent mass spectrometry/tandem mass spectrometry ( $\text{MS}^1/\text{MS}^2$ ) collision-induced dissociation mode (CID = 35) with dynamic exclusion enabled. SimGlycan software followed by manual verification was used for tandem mass spectrometry (MS/MS) spectral glycan identification.



## References

1. Wormald, M. R.; Rudd, P. M.; Harvey, D. J.; Chang, S. C.; Scragg, I. G.; Dwek, R. A. *Biochemistry* **1997**, *36*, 1370–1380.
2. *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds.; Cold Springs Harbor: New York, 2009; Vol. 2.
3. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E. *Proteomics* **2009**, *9*, 5398–5399.
4. Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Dwek, R. A.; Rudd, P. M. *Proteomics* **2009**, *9*, 3796–3801.
5. Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; Shitara, K. *J. Biol. Chem.* **2003**, *278*, 3466–3473.
6. 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. *Glycobiol.* **2007**, *17*, 104–118.
7. Jefferis, R.; Lund, J.; Pound, J. D. *Immunol. Rev.* **1998**, *163*, 59–76.
8. Umana, P.; Jean-Mairet, J.; Moudry, R.; Amstutz, H.; Bailey, J. E. *Nat. Biotechnol.* **1999**, *17*, 176–180.
9. Malhotra, R.; Wormald, M. R.; Rudd, P. M.; Fischer, P. B.; Dwek, R. A.; Sim, R. B. *Nat. Med.* **1995**, *1*, 237–243.
10. Goetze, A. M.; Liu, Y. D.; Zhang, Z.; Shah, B.; Lee, E.; Bondarenko, P. V.; Flynn, G. C. *Glycobiology* **2011**, *21*, 949–959.
11. Beck, A.; Reichert, J. M. *mAbs* **2012**, *4*, 419–425.
12. Ivarsson, M.; Villiger, T. K.; Morbidelli, M.; Soos, M. *J. Biotechnol.* **2014**, *188c*, 88–96.
13. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2013**, *85*, 715–736.
14. Gahoual, R.; Busnel, J. M.; Beck, A.; Francois, Y. N.; Leize-Wagner, E. *Anal. Chem.* **2014**, *86*, 9074–9081.
15. Ahn, J.; Bones, J.; Yu, Y. Q.; Rudd, P. M.; Gilar, M. *J. Chromatogr. B* **2010**, *878*, 403–408.
16. Ashline, D. J.; Lapadula, A. J.; Liu, Y. H.; Lin, M.; Grace, M.; Pramanik, B.; Reinhold, V. N. *Anal. Chem.* **2007**, *79*, 3830–3842.
17. Chen, X. Y.; Flynn, G. C. *Anal. Biochem.* **2007**, *370*, 147–161.
18. Higel, F.; Demelbauer, U.; Seidl, A.; Friess, W.; Sorgel, F. *Anal. Bioanal. Chem.* **2013**, *405*, 2481–2493.
19. Houel, S.; Hilliard, M.; Yu, Y. Q.; McLoughlin, N.; Martin, S. M.; Rudd, P. M.; Williams, J. P.; Chen, W. *Anal. Chem.* **2014**, *86*, 576–584.
20. Prien, J. M.; Prater, B. D.; Qin, Q.; Cockrill, S. L. *Anal. Chem.* **2010**, *82*, 1498–1508.
21. Schiel, J. E. *Anal. Bioanal. Chem.* **2012**, *404*, 1141–1149.
22. Schiel, J. E.; Au, J.; He, H. J.; Phinney, K. W. *Anal. Bioanal. Chem.* **2012**, *403*, 2279–2289.

23. Zhuang, Z.; Starkey, J. A.; Mechref, Y.; Novotny, M. V.; Jacobson, S. C. *Anal. Chem.* **2007**, *79*, 7170–7175.
24. Marino, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. *Nat. Chem. Biol.* **2010**, *6*, 713–723.
25. Mittermayr, S.; Bones, J.; Doherty, M.; Guttman, A.; Rudd, P. M. *J Proteome Res.* **2011**, *10*, 3820–3829.
26. Stockmann, H.; Adamczyk, B.; Hayes, J.; Rudd, P. M. *Anal. Chem.* **2013**, *85*, 8841–8849.
27. Tarentino, A. L.; Plummer, T. H., Jr. *Methods Enzymol.* **1987**, *138*, 770–778.
28. Ko, K.; Tekoah, Y.; Rudd, P. M.; Harvey, D. J.; Dwek, R. A.; Spitsin, S.; Hanlon, C. A.; Rupprecht, C.; Dietzschold, B.; Golovkin, M.; Koprowski, H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8013–8018.
29. Sjogren, J.; Struwe, W. B.; Cosgrave, E. F.; Rudd, P. M.; Stervander, M.; Allhorn, M.; Hollands, A.; Nizet, V.; Collin, M. *Biochem. J.* **2013**, *455*, 107–118.
30. Trimble, R. B.; Tarentino, A. L. *J. Biol. Chem.* **1991**, *266*, 1646–1651.
31. Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. *Nat. Rev. Drug Discovery* **2012**, *11*, 527–540.
32. Campbell, M. P.; Royle, L.; Radcliffe, C. M.; Dwek, R. A.; Rudd, P. M. *Bioinformatics* **2008**, *24*, 1214–1216.
33. Royle, L.; Campbell, M. P.; Radcliffe, C. M.; White, D. M.; Harvey, D. J.; Abrahams, J. L.; Kim, Y. G.; Henry, G. W.; Shadick, N. A.; Weinblatt, M. E.; Lee, D. M.; Rudd, P. M.; Dwek, R. A. *Anal Biochem* **2008**, *376*, 1–12.
34. Royle, L.; Dwek, R. A.; Rudd, P. M. Determining the Structure of Oligosaccharides N- and O-Linked to Glycoproteins. In *Current Protocols in Protein Science*; John Wiley & Sons, Inc.: New York, 2001.
35. Validation of Analytical Procedures: Text and Methodology Q2(R1). In *International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use*; ICH: Geneva, Switzerland, 2005.
36. Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H., Jr. *Anal. Biochem.* **1989**, *180*, 195–204.
37. Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977–982.
38. Valliere-Douglass, J. F.; Eakin, C. M.; Wallace, A.; Ketchem, R. R.; Wang, W.; Treuheit, M. J.; Balland, A. *J. Biol. Chem.* **2010**, *285*, 16012–16022.
39. Valliere-Douglass, J. F.; Kodama, P.; Mujacic, M.; Brady, L. J.; Wang, W.; Wallace, A.; Yan, B.; Reddy, P.; Treuheit, M. J.; Balland, A. *J. Biol. Chem.* **2009**, *284*, 32493–32506.
40. Guile, G. R.; Rudd, P. M.; Wing, D. R.; Prime, S. B.; Dwek, R. A. *Anal. Biochem.* **1996**, *240*, 210–226.
41. Melmer, M.; Stangler, T.; Premstaller, A.; Lindner, W. *J. Chromatogr. A* **2011**, *1218*, 118–123.
42. Anumula, K. R.; Dhume, S. T. *Glycobiology* **1998**, *8*, 685–694.
43. Lee, Y. C. *Anal. Biochem.* **1990**, *189*, 151–162.
44. Anumula, K. R. *Anal. Biochem.* **2006**, *350*, 1–23.

45. Brull, L. P.; Kovacic, V.; Thomas-Oates, J. E.; Heerma, W.; Haverkamp, J. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1520–1532.
46. Franz, A. H.; Lebrilla, C. B. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 325–337.
47. Harvey, D. J. *J. Mass Spectrom.* **2005**, *40*, 642–653.
48. Harvey, D. J. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 397–400.
49. Harvey, D. J.; Royle, L.; Radcliffe, C. M.; Rudd, P. M.; Dwek, R. A. *Anal. Biochem.* **2008**, *376*, 44–60.
50. Prater, B. D.; Connelly, H. M.; Qin, Q.; Cockrill, S. L. *Anal. Biochem.* **2009**, *385*, 69–79.
51. Prien, J. M.; Prater, B. D.; Cockrill, S. L. *Glycobiology* **2010**, *20*, 629–647.
52. Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood, J.; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R. *Glycobiology* **2007**, *17*, 677–687.
53. Ashline, D.; Singh, S.; Hanneman, A.; Reinhold, V. *Anal. Chem.* **2005**, *77*, 6250–6262.
54. Costello, C. E.; Contado-Miller, J. M.; Cipollo, J. F. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1799–1812.
55. Harvey, D. J. *Mass Spectrom. Rev.* **1999**, *18*, 349–450.
56. Mechref, Y.; Novotny, M. V.; Krishnan, C. *Anal. Chem.* **2003**, *75*, 4895–4903.
57. Morelle, W.; Slomianny, M. C.; Diemer, H.; Schaeffer, C.; van Dorselaer, A.; Michalski, J. C. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2637–2649.
58. Prien, J. M.; Ashline, D. J.; Lapadula, A. J.; Zhang, H.; Reinhold, V. N. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 539–556.
59. Prien, J. M.; Huysentruyt, L. C.; Ashline, D. J.; Lapadula, A. J.; Seyfried, T. N.; Reinhold, V. N. *Glycobiology* **2008**, *18*, 353–366.
60. Reinhold, V.; Zhang, H.; Hanneman, A.; Ashline, D. *Mol. Cell Proteom.* **2013**, *12*, 866–873.
61. Reinhold, V. N.; Reinhold, B. B.; Costello, C. E. *Anal. Chem.* **1995**, *67*, 1772–1784.
62. Reinhold, V. N.; Sheeley, D. M. *Anal. Biochem.* **1998**, *259*, 28–33.
63. Viseux, N.; de Hoffmann, E.; Domon, B. *Anal. Chem.* **1998**, *70*, 4951–4959.
64. Viseux, N.; deHoffmann, E.; Domon, B. *Anal. Chem.* **1997**, *69*, 3193–3198.
65. Weiskopf, A. S.; Vouros, P.; Harvey, D. J. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1493–1504.
66. Weiskopf, A. S.; Vouros, P.; Harvey, D. J. *Anal. Chem.* **1998**, *70*, 4441–4447.
67. Schiel, J. E.; Smith, N. J.; Phinney, K. W. *J. Mass Spectrom.* **2013**, *48*, 533–538.
68. Ghaderi, D.; Zhang, M.; Hurtado-Ziola, N.; Varki, A. *Biotechnol. Genet. Eng. Rev.* **2012**, *28*, 147–175.
69. Borrebaeck, C. K.; Malmborg, A. C.; Ohlin, M. *Immunol. Today* **1993**, *14*, 477–479.
70. Sheeley, D. M.; Merrill, B. M.; Taylor, L. C. *Anal Biochem* **1997**, *247*, 102–110.
71. Wright, A.; Morrison, S. L. *J. Immunol.* **1998**, *160*, 3393–3402.

72. 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.
73. Chung, C. H.; Mirakhur, B.; Chan, E.; Le, Q. T.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; Slebos, R. J.; Zhou, Q.; Gold, D.; Hatley, T.; Hicklin, D. J.; Platts-Mills, T. A. *N. Engl. J. Med.* **2008**, *358*, 1109–1117.
74. Impurities in New Drug Substances Q3(R2). In *International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use*; ICH: Geneva, Switzerland, 2006.
75. Liu, Y.; Salas-Solano, O.; Gennaro, L. A. *Anal. Chem.* **2009**, *81*, 6823–6829.
76. Toida, T.; Vlahov, I. R.; Smith, A. E.; Hileman, R. E.; Linhardt, R. J. *J. Carbohydr. Chem.* **1996**, *15*, 351–360.
77. Huang, Y.; Mechref, Y.; Novotny, M. V. *Anal. Chem.* **2001**, *73*, 6063–6069.
78. Anumula, K. R. *Anal. Biochem.* **2014**, *457*, 31–37.
79. Bigge, J. C.; Patel, T. P.; Bruce, J. A.; Goulding, P. N.; Charles, S. M.; Parekh, R. B. *Anal. Biochem.* **1995**, *230*, 229–238.
80. Lamari, F. N.; Kuhn, R.; Karamanos, N. K. *J. Chromatogr. B* **2003**, *793*, 15–36.
81. Melmer, M.; Stangler, T.; Schiefermeier, M.; Brunner, W.; Toll, H.; Rupprechter, A.; Lindner, W.; Premstaller, A. *Anal. Bioanal. Chem.* **2010**, *398*, 905–914.
82. Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M. *Anal. Bioanal. Chem.* **2010**, *397*, 3457–3481.
83. Prater, B. D.; Anumula, K. R.; Hutchins, J. T. *Anal. Biochem.* **2007**, *369*, 202–209.
84. Kang, P.; Mechref, Y.; Novotny, M. V. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 721–734.

## Chapter 5

# Separation Methods and Orthogonal Techniques

**David A. Michels,<sup>1</sup> Anna Y. Ip,<sup>2</sup> Thomas M. Dillon,<sup>2</sup>  
Kurt Brorson,<sup>3</sup> Scott Lute,<sup>3</sup> Brittany Chavez,<sup>3</sup>  
Ken M. Prentice,<sup>4</sup> Lowell J. Brady,<sup>4</sup> and Karen J. Miller\*,<sup>5</sup>**

<sup>1</sup>Department of Protein Analytical Chemistry, Genentech,  
South San Francisco, California 94080, United States

<sup>2</sup>Department of Process and Product Development, Amgen Inc.,  
Thousand Oaks, California 91361, United States

<sup>3</sup>Division of Monoclonal Antibodies, Office of Biotechnology Products,  
Center for Drug Evaluation and Research, U.S. Food and Drug  
Administration, Silver Spring, Maryland 20903, United States

<sup>4</sup>Department of Process and Product Development, Amgen Inc.,  
Seattle, Washington 98119, United States

<sup>5</sup>Global Analytical Sciences, Amgen Inc.,  
Thousand Oaks, California 91320, United States

\*E-mail: karen@amgen.com

Chromatographic and electrophoretic separation methods are key tools for the characterization of monoclonal antibodies. The objective of these methods for characterization is to resolve and identify critical quality attributes. Several orthogonal separation methods are often required in order to fully characterize the critical quality attributes of a particular monoclonal antibody preparation. Throughout this chapter, the use of separation methods and orthogonal techniques for monoclonal antibody characterization is demonstrated by the analysis of a monoclonal antibody material available from the NIST. The analysis of the material described in this chapter serves as characterization of the material to enable its use as a standard and exemplifies the

steps that may be taken to characterize a therapeutic monoclonal antibody. Recent advances described in this chapter include the use of high pressure systems in combination with small particle size columns to accomplish rapid separations which save time and reduce solvent consumption. In addition, the use of software tools for the development and optimization of methods is described.

## Introduction

Separation methods have long been the key tools in characterizing proteins such as monoclonal antibodies (mAbs) (1). Gel electrophoresis methods, which have been used for several decades, provide a qualitative assessment of antibody size variants and charge variants (2, 3). However, these methods are difficult to convert to quantitative assessments, and identification of the resolved variants by orthogonal methods is often complicated by the gel matrix. Capillary electrophoresis methods with UV absorbance or fluorescence detection provide superior resolution and quantitative assessment over the former gel-based methods. However, identification of the resolved variants continues to be challenging. Chromatographic methods are commonly used for mAb separations both preparatively, in the large-scale purification of antibodies, as well as analytically for the characterization and routine assessment of antibodies. However, chromatographic methods may not have the resolution of electrophoresis methods. The balance between resolution, quantitation, and opportunity for isolation is ever-present during protein characterization, and a combination of techniques is likely necessary during development.

A significant change in the use of separation methods for the characterization of therapeutic mAbs has been imparted by the application of quality by design (QbD) principles (4). In the past, separation techniques may have been used to generate a chromatographic or electrophoretic profile that could be used as a fingerprint for process consistency. Now, the focus of separation techniques is the resolution and identification of critical quality attributes (CQAs), those attributes that influence the purity, strength, and stability of the product (5). For mAbs, fortunately, many of the CQAs are common to the molecular class, and thus the focus of the characterization is similar from one antibody to another. For example, in the analysis of charge variants, deamidated species may be more critical than C-terminal variants, and therefore the focus of the separation technique may be in assuring that the peaks containing deamidated species are well resolved, even at the expense of resolution of the peaks containing C-terminal variants. Separation techniques for hydrophobic variants, such as reversed-phase and hydrophobic interaction, have been less commonly used for mAbs due to the limited resolution of variant species from the intact antibody. However, these techniques can be even more useful for resolving variant forms when the molecule is reduced to light chain (L) and heavy chain (H) fragments, or cleaved into smaller segments, such as F(ab')<sub>2</sub> and single chain Fc (scFc) fragments.

The major types of chromatographic and electrophoretic techniques used for the analysis of mAbs were applied to the NISTmAb material: cation exchange HPLC (CEX-HPLC), size exclusion HPLC (SEC-HPLC), reversed-phase HPLC (RP-HPLC), hydrophobic interaction chromatography (HIC), reduced and non-reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), reduced and non-reduced capillary electrophoresis with SDS (CE-SDS), microchip electrophoresis with SDS (MCE-SDS), capillary zone electrophoresis (CZE) and capillary isoelectric focusing (cIEF). The use of small particle size columns for rapid resolution separations was also demonstrated where possible. As is typical for therapeutic mAbs, several minor species were resolved. Where possible, orthogonal techniques were used to characterize the minor species, or those techniques that might be employed to identify the minor species are described.

Due to the similarity in structure and size between monoclonal IgG molecules, many of these assays are likely to be appropriate for a broad range of mAbs. Ion exchange is an exception, because the mobile phase pH and gradient must be individualized to a particular mAb because of its unique isoelectric point (pI) and surface charge distribution. HIC and RP-HPLC may also require some optimization based on an individual IgG's hydrophobicity. However, the columns and mobile phase components are likely to be similar across these assays.

## Methods

All analysis was performed using Humanized mAb IgG1 Lot 3F1b, 10 mg/mL in 25 mM histidine from NIST (Gaithersburg, MD) stored at  $-80^{\circ}\text{C}$ . Sample was used neat unless otherwise stated.

### Size Exclusion Chromatography-HPLC

#### *Reagents*

Phosphate-buffered saline (50 mM sodium phosphate with 150 mM sodium chloride [NaCl], pH 7.4) was obtained from HyClone (Salt Lake City, UT).

#### *Sample Analysis*

The analytical system consisted of an Agilent (Palo Alto, CA) 1260 quaternary pump, Agilent 1260 ALS autosampler, and an Agilent 1200 Multiple Wavelength Detector (MWD) absorbance detector. The mobile phase was phosphate-buffered saline (see *Reagents*). A sample amount of 150  $\mu\text{g}$  was injected onto a Tosoh Bioscience (King of Prussia, PA) TSKgel G3000SWxl SEC column (30 cm  $\times$  7.8 mm i.d., particle diameter 5  $\mu\text{m}$ ) with a guard column at a flow rate of 0.400 mL/min for 45 minutes at ambient temperature. Absorbance was detected at 280 nm.

## Size Exclusion Chromatography-Ultrahigh-Pressure Liquid Chromatography

### *Reagents*

Solutions of 0.5 M sodium phosphate monobasic, 0.5 M sodium phosphate dibasic and 1 M NaCl were purchased from Fisher Scientific (Pittsburgh, PA).

### *Sample Analysis*

The size exclusion-ultrahigh-pressure liquid chromatography (SEC-UHPLC) analytical system consisted of a Waters (Milford, MA) H-Class Acquity system with a quaternary pump, a flow through needle (FTN) autosampler, and a tunable ultraviolet (TUV) detector. The mobile phase was mixed to 100 mM sodium phosphate, 250 mM NaCl, pH 6.8 by delivery of a ratio of 7% sodium phosphate monobasic, 13% sodium phosphate dibasic, 25% NaCl, and 55% water using the four channels of the quaternary pump and the solutions in the *Reagents* section as mobile phases. A sample amount of 60  $\mu\text{g}$  was injected onto a Waters BEH SEC column (150 mm  $\times$  4.6 mm i.d, particle diameter 1.7  $\mu\text{m}$ ) at a flow rate of 0.400 mL/min for 6 minutes at ambient temperature. Absorbance was detected at 280 nm.

## Size Exclusion Chromatography with Multi-Angle Light Scattering

### *Reagents*

Solutions of 0.5 M sodium phosphate monobasic, 0.5 M sodium phosphate dibasic, and 1 M NaCl were purchased from Fisher Scientific (Pittsburgh, PA). The solutions were blended to create a mobile phase of 100 mM sodium phosphate, 250 mM NaCl, pH 6.8.

### *Sample Analysis*

The analytical system consisted of an Agilent 1100 binary pump, Agilent 1100 autosampler, Agilent 1100 diode array detector (DAD) absorbance detector, Wyatt (Santa Barbara, CA) Optilab rEX refractive index detector (RID), and a Wyatt DAWN<sup>®</sup> EOS multi-angle light scattering (MALS) detector. The MALS processing software used was Astra v.6.1, also from Wyatt. The mobile phase was 100 mM sodium phosphate, 250 mM NaCl, pH 6.8 (as described in *Reagents*, above). A sample amount of 300  $\mu\text{g}$  was injected onto two TSKgel G3000SWxl SEC columns connected in series with a guard column at a flow rate of 0.500 mL/min for 60 minutes at ambient temperature. The flow path was from the column to the DAD detector, then to the MALS detector, and finally to the RID detector. Throughout the run, absorbance at 280 nm and refractive index measurements were recorded and transmitted to the Wyatt DAWN<sup>®</sup> EOS for integration into the



Astra 6.1 software for analysis. An extinction coefficient of  $1.42 \text{ mL} \cdot \text{mg}^{-1} \cdot \text{cm}^{-1}$  was used for all analyses.

## Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

### *Reagents*

NuPAGE<sup>®</sup> LDS Sample Buffer, NuPAGE<sup>®</sup> Reducing Agent, NuPAGE<sup>®</sup> Antioxidant, Novex<sup>®</sup> Sharp Standard, and MOPS buffer were obtained from Invitrogen (Carlsbad, CA). Brilliant Blue G-250, acetic acid, and 2-propanol were obtained from Fisher Scientific (Fairlawn, NJ).

### *Sample Analysis*

For both the non-reduced and the reduced samples, 6  $\mu\text{g}$  of the sample was mixed with NuPAGE<sup>®</sup> LDS Buffer (4 $\times$ ) and sterile water. For the reduced samples, NuPAGE<sup>®</sup> Reducing Agent was also added during sample preparation. The sample was then incubated at 70  $^{\circ}\text{C}$  for 10 minutes. Three  $\mu\text{g}$  of each sample was loaded on to Novex NuPAGE<sup>®</sup> (4–12%) Mini Gels. Ten  $\mu\text{L}$  of Novex<sup>®</sup> Sharp Standard was used as a reference for molecular weight. MOPS buffer was used to fill the upper and lower chamber, and 500  $\mu\text{L}$  NuPAGE<sup>®</sup> Antioxidant was added to the upper buffer chamber. Gels were run at 200 V for 40 to 50 minutes. Antioxidant was added to ensure that proteins remained in a reduced state during electrophoresis.

Gels were fixed using a 10% acetic acid, 25% 2-propanol solution for at least 30 minutes. The Coomassie staining was performed overnight with a solution of 10% acetic acid and 0.006% Brilliant Blue G-250. De-staining was performed with a 10% solution of acetic acid, replaced 2 times before taking a picture of the gel.

## Capillary Electrophoresis with Sodium Dodecyl Sulfate

### *Reagents*

Potassium cyanide (KCN) and 3-(2-furoyl)quinoline-2-carboxaldehyde (FQ) reagents were purchased from Molecular Probes (Eugene, OR). Monobasic and dibasic sodium phosphate, citrate, dithiothreitol (DTT),  $\beta$ -mercaptoethanol, iodoacetamide (IAM) and *N*-ethylmaleimide (NEM) were purchased from Sigma-Aldrich (St. Louis, MO). SDS, sodium hydroxide (NaOH) and hydrochloric acid (HCl) reagents were purchased from J.T. Baker (Phillipsburg, NJ). Fused-silica capillaries were purchased from Polymicro Technologies (Phoenix, AZ). Replaceable SDS-MW sieving gel was purchased from Beckman Coulter (Fullerton, CA).

## *Sample Analysis*

The platform CE-SDS assay has been described in detail elsewhere (6). In brief, the experiment was carried out on a Beckman-Coulter PA 800 Plus instrument equipped with either a photodiode array (PDA) detector or a laser-induced fluorescence (LIF) detector. The separation was performed with a 31 cm long, 50  $\mu\text{m}$  inner diameter bare fused-silica capillary (10 and 20 cm effective length) filled with the Beckman SDS-MW sieving gel buffer. Samples stored at 20  $^{\circ}\text{C}$  were injected at  $-5$  kV and separated at  $-15$  kV. Between injections, the capillary was rinsed with 0.1 M NaOH, 0.1 M HCl, and purified water at 70 psi for 5, 1, and 1 minutes, respectively. For UV detection, the capillary cartridge was maintained at 20  $^{\circ}\text{C}$ , and protein was monitored at 214 nm absorbance. For LIF detection, the capillary was maintained at 40  $^{\circ}\text{C}$ , and protein excited with the 488 nm solid-state diode laser was monitored using a 600 nm long-pass filter.

### *CE-SDS Sample Preparation for UV Detection*

The NISTmAb was first diluted to 0.5 mg/mL with the CE-SDS sample buffer (1% SDS in 0.04 M citrate-phosphate, pH 6.6) and mixed with either 40 mM IAM for non-reduced or 5% (v/v)  $\beta$ -mercaptoethanol for reduced conditions. Denaturation was carried out at 70  $^{\circ}\text{C}$  for 5 minutes (non-reduced) or 10 minutes (reduced).

### *CE-SDS Sample Preparation for LIF Detection*

The NISTmAb was first prepared by buffer exchanging 300  $\mu\text{g}$  of protein into 0.1 M sodium phosphate, pH 6.7 labeling buffer. An aliquot of the buffer exchanged sample was treated with 0.2% SDS and 16 mM NEM and heated for 5 minutes at 70  $^{\circ}\text{C}$ . Next, 150  $\mu\text{g}$  of the SDS-complexed protein was mixed with fluorogenic FQ at a dye:antibody molar ratio of 25:1 and derivatized for 10 minutes at 50  $^{\circ}\text{C}$  in the presence of 1 mM KCN. The labeling reaction was quenched by diluting the sample with a 1% SDS solution. The reduced sample was prepared by mixing an aliquot of the quenched sample (non-reduced) with 50 mM DTT and heating for 10 minutes at 70  $^{\circ}\text{C}$ .

## **Microchip Electrophoresis with Sodium Dodecyl Sulfate**

### *Reagents*

HT Glycan Protein LabChip kit was purchased from PerkinElmer (Waltham, MA). Monobasic and dibasic sodium phosphate, DMSO, DTT, and IAM reagents were purchased from Sigma-Aldrich (St. Louis, MO). SDS was purchased from J.T. Baker (Phillipsburg, NJ).

## *Sample Analysis*

Analysis was performed with a PerkinElmer LabChip GXII system using the vendor-supplied kit containing the HT High Resolution microchip, sieving gel matrix, lower marker, wash buffer, and lyophilized dye (reconstituted to 1 mg/mL in DMSO before use). The microchip was prepared as recommended by the manufacturer and controlled using the HT Pico Protein Express 200 script. The pre-column labeling procedure was as follows: the NISTmAb was first diluted into 25 mM sodium phosphate pH 6.7 labeling buffer before mixing with a 0.25% SDS solution containing either 40 mM IAM (non-reduced) or 15 mM DTT (reduced), and heated at 70 °C for 5 minutes. Next, 70 µg of the SDS-complexed antibody was mixed with 5 µL of the dye working solution prepared at 0.2 mg/mL; the labeling reaction was heated at 30 °C for 10 minutes. Denatured and labeled samples were introduced to the microchip from the 96-well plate through a sipper vacuum. The microchip was maintained at 30 °C throughout the experiment. Excitation and detection wavelengths were 630 nm and 700 nm, respectively.

## **Reversed-Phase HPLC**

### *Reagents*

All materials were of HPLC grade unless otherwise noted. Trifluoroacetic acid (TFA) and guanidine hydrochloride were purchased from Pierce/Thermo Scientific (Rockford, IL). HPLC-grade water, n-propanol, and DTT were purchased from Sigma-Aldrich (St. Louis, MO). IdeS, a streptococcal cysteine proteinase with unique specificity for IgG molecules, cleaving just below the IgG hinge, which is also known under the name FabRICATOR® IgG Protease, was purchased from Genovis AB/Sigma-Aldrich (St. Louis, MO).

### *Sample Analysis*

An Agilent 1260 HPLC composed of a binary pump (G1312B), auto-sampler with cooling capability (G1329B) set at 5 °C, column heater (G1316C), and a DAD (G4212B) was used. A Zorbax 300SB-C8 rapid resolution column (2.1 × 50 mm, 3.5 µm) from Agilent (Palo Alto, CA) was used.

### *Reduction of Disulfides*

Samples were diluted to 1 mg/mL in a 300 mM Tris buffer at pH 8.0 containing 8.0 M guanidine hydrochloride. DTT was added to the samples to a final concentration of 50 mM followed by incubation at 55 °C for 30 minutes.

## FabRICATOR® IgG Protease (IdeS) Limited Proteolysis

Samples were diluted to 2 mg/mL in a 300 mM Tris buffer at pH 7.5 and digested using the FabRICATOR® manufacturer's protocol.

The NISTmAb was injected neat (10 mg/mL) onto the column at a protein load of 30 µg for intact analysis and at a protein load of 20 µg for analysis of reduced and limited proteolysis samples. The column temperature was 75 °C. The flow rate was held constant at 0.50 mL/min. Mobile phase A (MPA) was composed of 5% n-propanol in water with 0.10% TFA. Mobile phase B (MPB) was composed of 90% n-propanol in water with 0.10% TFA. Samples were injected at a loading condition of 8% MPB. The gradient consisted of a rapid increase from 8 to 19% MPB in 0.5 min, then a gradual gradient to 22% MPB in 7.5 min. A 0.5 min wash step of 95% MPB followed, and then the column was equilibrated at the initial condition. The absorbance was monitored at 215 nm.

### **Hydrophobic Interaction Chromatography**

#### *Reagents*

The HIC method used for the NISTmAb used previously published reagents (7).

#### *Sample Analysis*

The HPLC system used consisted of an Agilent HP 1200 quaternary HPLC equipped with an Agilent MWD detector. A Dionex ProPac HIC-10 column (Dionex, Sunnyvale, CA) was equilibrated in MPA, 1.0 M ammonium sulfate in 20 mM sodium acetate, pH 5.2. One hundred µg of protein diluted to 1 mg/mL with MPA was loaded. The column was held at the initial conditions for 4 minutes followed by a step to 20% MPB, where MPB was 20 mM sodium acetate, pH 5.2, for 5 minutes. Elution was achieved using a gradient to 60% MPB over 26 minutes. The column was operated at 0.5 mL/min, and detection used absorbance at 214 nm.

### **Cation Exchange HPLC**

#### *Reagents*

Mobile phases were prepared from the following reagents: sodium phosphate monobasic monohydrate from Mallinckrodt (St. Louis, MO), NaCl from Mallinckrodt (St. Louis, MO), and 10 N NaOH from J.T. Baker (Phillipsburg, NJ). Carboxypeptidase was obtained from Roche Diagnostics (Indianapolis, IN).

## Sample Analysis

The analytical system consisted of an Agilent 1100 quaternary pump, Agilent 1329A ALS autosampler, Agilent 1314A variable wavelength detector (VWD) absorbance detector, Agilent 1200 series 1160 A 12/13 solvent selector valve, and Agilent 1200 series G1159A 6-column selector. The columns screened were: ProPac® WCX-10 (4.0 × 250 mm, 10 μm), from Dionex (Sunnyvale, CA), BioPro SP-F (4.6 × 100 mm, 5 μm) from YMC Co. (Allentown, PA), Bio MAb NP-5, 4.6 × 250 mm, 5 μm from Agilent, and Antibodix WCX-NP5 (4.6 × 100 mm, 5 μm) from Sepax Tech. (Newark, DE). The final optimized method used a protein load of 10 μg on the BioPro SP-F column. MPB was 20 mM phosphate buffer, pH 6.7, and MPB was 20 mM phosphate buffer, pH 6.7, with 0.5 M NaCl. The column temperature was 35 °C. The column was held at an initial condition of 3% MPB for 3 minutes, followed by a linear gradient to 23% MPB over 30 minutes. The column was held at 23% MPB for 0.1 minute, ramped up to 99% MPB over 0.5 minutes, held for 3 minutes, ramped back down to the initial conditions over 0.5 minutes, and then equilibrated at the initial conditions for 3 minutes. The detection wavelength was 215 nm.

The design of experiments (DOE) software used was Fusion AE-LC Method Development Module from S-Matrix Corp. (Eureka, CA).

## Capillary Zone Electrophoresis

### Reagents

Described in He et al. 2011 (8).

### Sample Analysis

The CZE assay has been described in detail elsewhere (8). In brief, separation was performed on a Beckman PA 800 Plus instrument with a 31 cm long × 50 μm inner diameter bare fused-silica capillary (20 cm effective length) encased in a thermally controlled compartment at 20 °C. The capillary was filled with 0.4 M ε-amino-caproic acid, 0.05% hydroxypropylmethyl cellulose, 2 mM triethylenetetramine (TETA) running buffer, and adjusted to pH 5.7 with acetic acid. Sample was injected at 0.5 psi for 10 seconds, separated at 30 kV, and detected by UV at 214 nm. Between injections, the capillary was rinsed with 0.1 M HCl at 60 psi for 5 minutes before conditioning the capillary with running buffer at 50 psi for 10 minutes. The NISTmAb was diluted with deionized water to 1 mg/mL; an aliquot of the diluted samples was treated with carboxypeptidase B (CpB) at a protein:enzyme ratio of 100:1, and incubated at 37 °C for 20 minutes. Samples were stored at 10 °C in the autosampler.

## Capillary Isoelectric Focusing

### *Reagents*

Described in references (9, 10).

### *Sample Analysis*

The Beckman platform capillary isoelectric focusing (cIEF) assay has been described in detail elsewhere (9, 10). In brief, experiments were performed on a Beckman PA 800 Plus instrument with a 31 cm long  $\times$  50  $\mu$ m inner diameter neutral-coated capillary (20 cm effective length) thermally controlled at 20 °C. The capillary was filled with the protein-ampholyte solution at 25 psi for 99 seconds before immersing the ends into the anolyte and catholyte solutions. The sample was focused at 25 kV, and then mobilized across the detection window by replacing the catholyte with 350 mM acetic acid at 30 kV. Absorbance of protein was monitored at 280 nm. Between injections, the capillary was rinsed at 50 psi with 4.3 M urea and water for 5 and 3 minutes, respectively.

The NISTmAb was diluted to 1.25 mg/mL with deionized water; an aliquot was treated with CpB at a protein:enzyme ratio of 100:1 and incubated at 37 °C for 20 minutes. The ampholyte solution was prepared by mixing Beckman cIEF gel with urea, pH 3–10 Pharmalytes®, anodic and cathodic stabilizers, and two pI markers. The final protein-ampholyte mixture was prepared by combining 40  $\mu$ L of either the native or CpB-treated sample with 160  $\mu$ L of ampholyte solution, and then stored in the autosampler at 10 °C.

## Imaged Capillary Isoelectric Focusing

### *Reagents*

Described in Li et al. (2007) and Susic et al. (2008) (11, 12).

### *Sample Analysis*

### Charge Heterogeneity

The platform imaged capillary isoelectric focusing (ICIEF) assay has been described in detail elsewhere (11, 12). In brief, the experiments were carried out on a ProteinSimple ICE3 analyzer with a fluorocarbon-coated capillary cartridge (100  $\mu$ m inner diameter). The anolyte was 80 mM phosphoric acid, and the catholyte was 100 mM NaOH, both in 0.1% methylcellulose (MC). The capillary was filled with protein-ampholyte mixture, and then focused at 1.5 kV for 1 minute, followed by 3.0 kV for 10 minutes. An image of the focused charge variants was obtained by passing 280 nm UV light through the whole 5 cm capillary and into the lens of

a charge-coupled device (CCD) digital camera. Between injections, the capillary was rinsed at 2000 mBar with 0.5% MC and deionized water for 3 and 1 minute, respectively.

The NISTmAb was diluted to 1.25 mg/mL with deionized water. For the CpB-treated sample, CpB was added at the dilution step at an enzyme-to-substrate ratio of 1:100 followed by incubation at 37 °C for 20 minutes. The ampholyte solution consisted of a mixture of 0.35% MC, 3% total carrier ampholytes (15% pH 3–10 and 85% pH 8–10.5 Pharmalytes), and 0.2% pI markers 7.55 and 9.77 in purified water. The final protein-ampholyte mixture was prepared by combining 40  $\mu$ L of either native or CpB-treated samples with 160  $\mu$ L of ampholyte solution and stored at 10 °C in the PrinCE autosampler.

### Determination of Apparent pI

The experiment was carried out using the same conditions and parameters as the charge heterogeneity assay with the following changes. The NISTmAb was diluted to 0.5 mg/mL with deionized water. The ampholyte solution consisted of a mixture of 0.35% MC, 5% pH 3–10 Pharmalyte®, and 0.2% marker solution containing eight pI markers in purified water. The sample was focused at 1.5 kV for 1 minute, followed by 3.0 kV for 5 minutes. A standard curve was prepared by linear regression analysis of the measure pixel number and the vendor-assigned pI value for each marker.

## Results and Discussion

### Size Variant Analysis

The desired IgG product contains disulfide-linked light chain and N-linked glycosylated heavy chain subunits (H:H:L:L), combining for an approximate mass of 150 kDa. In addition to the desired product, mAb preparations typically contain size variants, which include multimeric forms, such as dimer and higher order aggregates, and smaller molecular weight forms, such as fragments, free light chain, free heavy chain or partial molecules (e.g., half-bodies such as H:L). Process-related impurities (PRIs), such as host cell impurities or cross contamination of products produced in the same facility, may also be detected as size variants in separation methods if present in sufficient quantity. However, PRIs are generally not detectable in separation assays for mAb preparations of acceptable purity. As regulators and health authorities continue to scrutinize product fragments, aggregates, and host cell protein impurities, there is an increased demand for detailed product characterization and control of these as CQAs.

Aggregate forms are considered CQAs of particular concern in therapeutic proteins, due to potential immunogenicity concerns (13). Therefore, the accurate identification and quantitation of aggregates in mAb preparations is a key objective of any characterization effort. Among the orthogonal size-based separation methods commonly used in industry, SEC-HPLC is widely accepted

as the primary tool for characterizing protein monomers and multimeric forms, but is limited in capacity to resolve the various combinations of light chain and heavy chain fragments observed in mAb products. In addition, mAb products characterized by SEC-HPLC may appear as the intact form under the native conditions employed by the assay. However, the native conditions of SEC-HPLC are advantageous because it enables efficient collection of the resolved size variants for characterization of biological activity through the appropriate potency assay.

Lower molecular weight variants may be present in mAb preparations as a result of protease activity or incompletely assembled complexes. Some lower molecular weight size variants can be resolved under the native conditions of size exclusion; denaturing SDS-PAGE or CE-SDS methods, however, are often more useful for the measurement of smaller molecular weight components, because some of these species may exist as non-covalent complexes that co-elute with the intact monomer. These methods can be run on reduced samples (detecting all components in disulfide-bonded complexes) and non-reduced samples (measuring the actual protein complex, including disulfide-linked incomplete assemblies such as H:L or heavy-heavy-light species [H:H:L]). Unfortunately, the further characterization of the variant forms resolved by SDS-PAGE and CE-SDS is difficult due to the use of SDS as a denaturant because it interferes with many assays and is difficult to remove from the sample. Therefore, given the distinct capabilities of the various techniques, orthogonal methods are typically required for characterization of size variants.

### *Size Exclusion Chromatography*

SEC is a method where molecules in solution are separated by size or, more specifically, by hydrodynamic radius. When applied to large molecules or macromolecular complexes such as mAbs, the larger components of the sample mixture, like aggregates, are excluded from the pores in the chromatographic media and elute first, whereas the smaller components (e.g., fragments, free light chain, heavy chain, clips) diffuse further into the pores and elute in subsequent peaks according to size (14, 15).

In modern biotechnology laboratories, SEC is run in HPLC or UHPLC mode (i.e., SEC-HPLC or SEC-UHPLC) under native or denaturing conditions using specialized, high-performance systems. UHPLC, which operates under the same principles as HPLC, is considered to be a step improvement in liquid chromatography based on advances in particle chemistry performance, system optimization, detector design, and data processing.

For antibodies, a SEC-HPLC chromatogram with good separation will feature an antibody monomer elution peak with a symmetrical shape, and allow resolution of higher molecular weight forms (i.e., aggregates) as either a front-shoulder or distinct earlier peak relative to the main monomer peak. The smaller forms of the antibody (if present) may also be apparent as either a trailing shoulder or distinct peak following the main peak, although reduced SDS-PAGE



or CE-SDS (described in sections following) are considered to be more sensitive and informative methods to detect smaller forms.

In the SEC-HPLC analysis of the NISTmAb (shown in Figure 1), the major peak elutes between 24 and 26 minutes, confirming that the product is largely intact monomer. Aggregates, which would elute prior to the main peak, and lower molecular weight species, which would elute after the main peak, are not detected under these assay conditions. The asymmetrical monomer peak and the absence of aggregates suggests that further optimization of the method may be required.

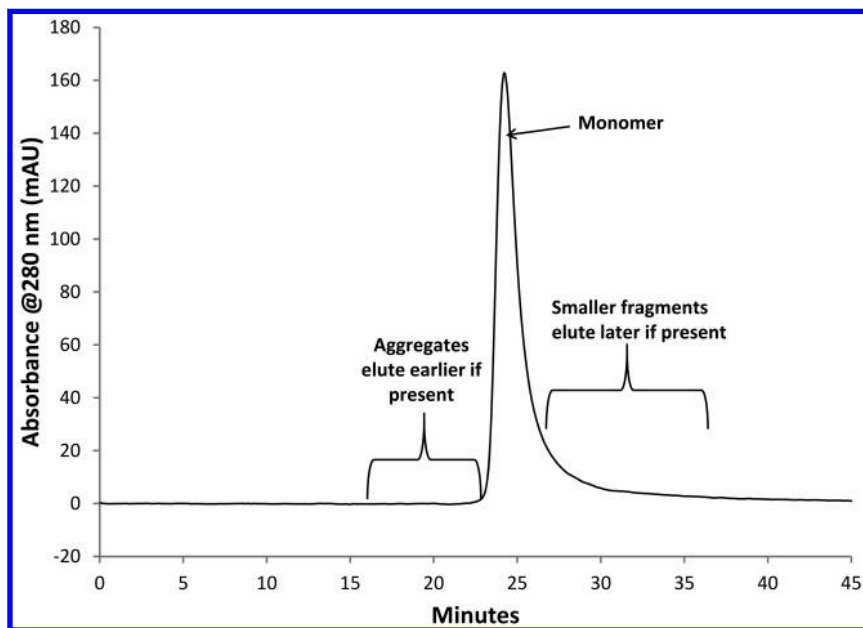


Figure 1. Size exclusion chromatography (SEC)-HPLC analysis of the NISTmAb.

The SEC-UHPLC analysis of the NISTmAb, shown in Figure 2 was performed using a mobile phase with the same constituent components (sodium phosphate and NaCl) but a different pH, as well as higher phosphate and NaCl concentrations. Under these conditions, the main peak is evident at approximately 3 minutes. Aggregates are detected at 1.0% and elute slightly earlier, with 0.3% low molecular weight species eluting later than the main peak. The system peaks usually attributed to the sample matrix elute in the inclusion volume. The resolved species can be characterized by on-line techniques such as MALS (described in the following section) or may be collected for off-line analysis, such as bioassays to determine the potency. Denaturing methods such as reduced and non-reduced CE-SDS (see *Capillary Electrophoresis with Sodium Dodecyl Sulfate*) and RP-HPLC (see *Reversed-Phase HPLC*) can be used to determine relative percentage of covalent and non-covalent multimers in a given SEC-HPLC peak. For smaller molecular weight components, these techniques may be used to determine the specific site of fragmentation, clipping, or the constituent chains of partial molecules.

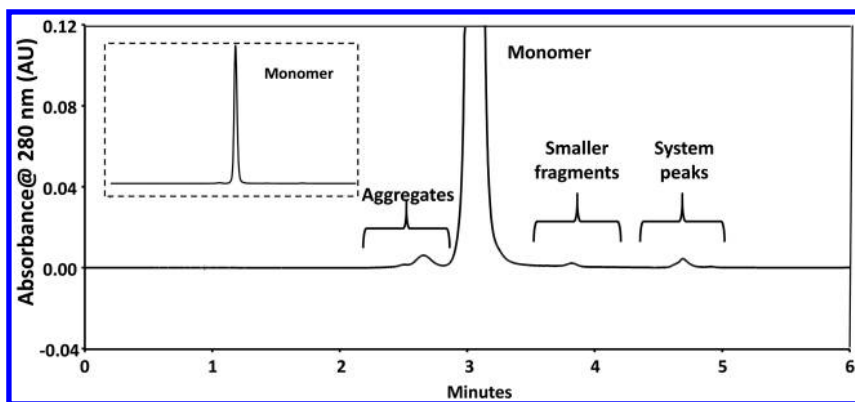


Figure 2. Size exclusion chromatography (SEC)-UHPLC separation of the NISTmAb with UV detection. The inset shows the full-scale profile.

Even relatively straightforward isocratic assays such as the SEC assays shown in Figure 1 and Figure 2 may require optimization to obtain the best possible resolution and recovery of mAb variants while minimizing assay-induced artifacts. By raising the ionic strength or including low amounts of organic modifier in the mobile phase, any non-specific interactions with the column stationary phase may be reduced, thereby improving resolution of the aggregated species and the tailing observed on the monomer peak (16). Additional characterization using an orthogonal technique such as analytical ultracentrifugation (AUC), discussed in the Biophysical chapter/Volume 2, Chapter 6, may be helpful in identifying optimum assay conditions (17).

### *Size Exclusion Chromatography with Multi-Angle Light Scattering*

On-line characterization of SEC-HPLC peaks is typically performed with MALS (18, 19) and, more recently, Quasi-Elastic Light Scattering (QELS) (20). On-line detection enables the characterization of non-covalent species that may dissociate upon collection and off-line analysis. At this time, MALS and QELS are generally run with SEC-HPLC with high protein loads in order to accurately measure the low-level aggregate or low molecular weight species. In MALS and QELS, light scattering from macromolecules undergoing Brownian motion leads to destructive or constructive interference. In QELS, the time-dependent fluctuations in the light associated with the interference are measured by a fast photon counter. Information concerning the molar mass and hydrodynamic radius of a sample can be derived from the intensity and angular dependence of the scattered light (21–23).

As can be seen Figure 3, the material in the NISTmAb monomer elution peak is consistently the same molar mass for the duration of the peak (indicated by a consistently level black line from 32 to 36 minutes). The data derived from MALS calculates the molecular weight of the antibody monomer as 147 kDa, and the QELS calculates the hydrodynamic radius as 6.1 nm. Because the MALS

calculation of monomer molecular weight is dependent on certain assumptions and variables (e.g., solution refractive index), other methods (e.g., mass spectrometry [MS]) should be viewed as more accurate for this purpose. On the other hand, QELS is considered to be probably the most reliable and precise method to measure hydrodynamic radius. The small amount of aggregate in the sample correlates well to dimer at 270 kDa. The measured molecular weight of the smaller fragments was 139 kDa, although this value is likely an average of a mixture of species (see *CE-SDS* and *RP-HPLC* sections for more detailed characterization of these species). The low levels of the non-monomeric forms decreases the MALS accuracy, and a combination of light scattering, refractive index, UV, and QELS data may need to be leveraged for adequate characterization.

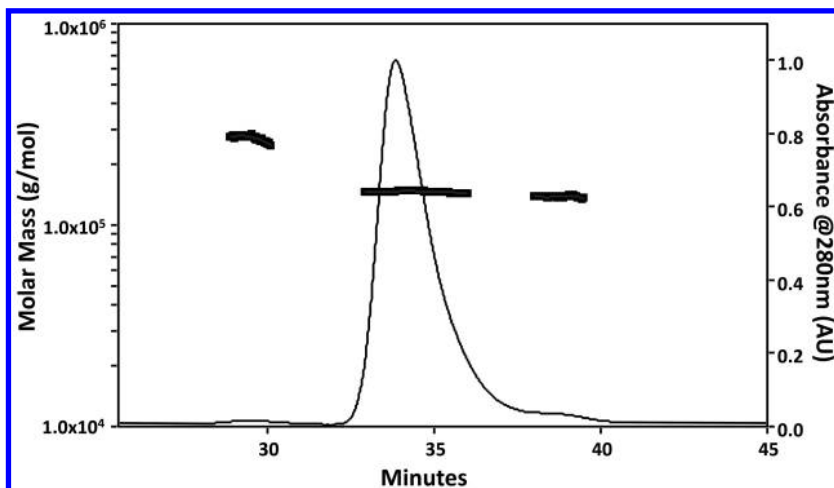


Figure 3. Size exclusion chromatography (SEC)-multi-angle light scattering (MALS) analysis of the NISTmAb main peak. Molar mass at each point of the eluate peak was deconvoluted by the MALS software. A nearly horizontal molar mass signal indicates homogeneity in size distribution in the test article.

### Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

SDS-PAGE has been used for nearly 5 decades for separating size variants under denaturing conditions (2, 3). PAGE separates proteins according to their electrophoretic mobility. Mobility in the gel is a function of the length, conformation, and charge of the molecule. In the case of SDS-PAGE, the negatively charged detergent SDS is added to the sample and then heated prior to electrophoresis. This denatures the protein and allows binding of SDS to the polypeptide chain. This imparts an even distribution of charge per unit mass, thereby resulting in a fractionation by approximate size during electrophoresis. In reduced SDS-PAGE, DTT or  $\beta$ -mercaptoethanol is added to reduce disulfide bonds, breaking down protein complexes like antibodies into polypeptide subunits.

SDS-PAGE of antibodies is typically run in both reduced and non-reduced formats, as shown in Figure 4 for the NISTmAb. The non-reduced format gives a picture of the percentage of intact molecule (i.e., H:H:L:L). SDS-PAGE is especially useful for the detection of incomplete complexes, such as H:L forms, which are particularly problematic for IgG4, or other forms, such as H:H:L species. It can also detect free light chains, as can be seen in the case of NISTmAb in Figure 4, although the detection of such species have also been reported to be artifacts of sample preparation (24). Here again, method optimization may be required. Further characterization of the resolved bands typically requires excision from the gel slab, removal of the stain, and extraction of the protein from the gel matrix (25, 26).

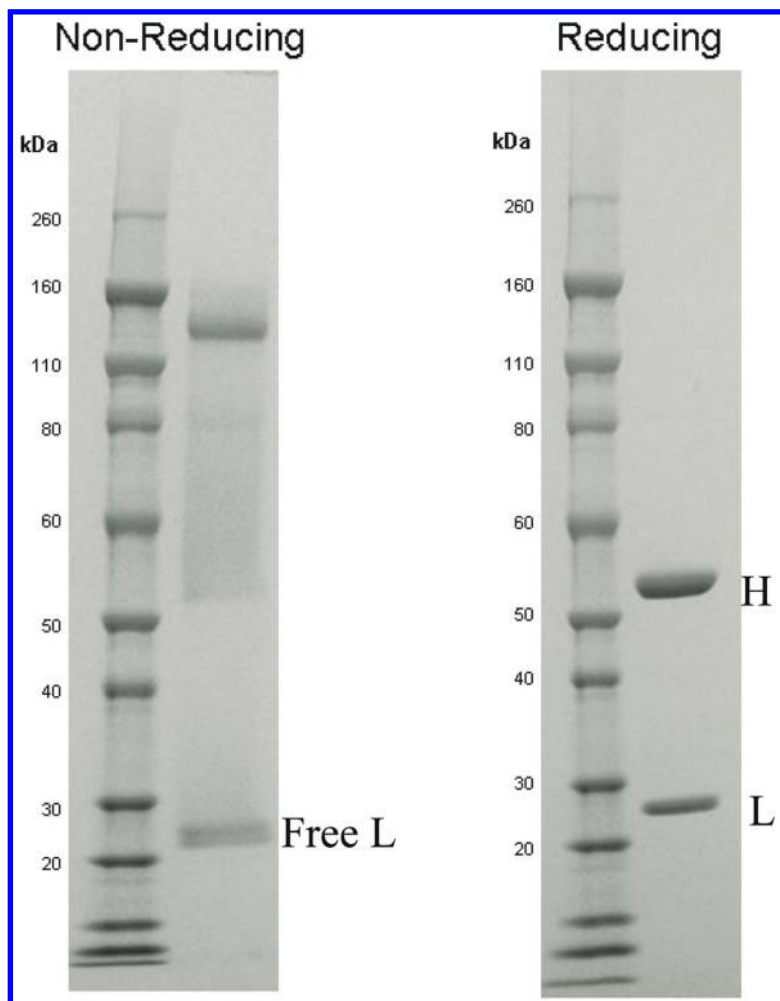


Figure 4. Reduced and non-reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels for the NISTmAb.

Reduced SDS-PAGE can also be used to evaluate the potential presence of antibody-derived fragments or host cell protein components if individual species are present at a sufficiently high level. Because all complexes have been broken down, any species that is not the size of heavy chain or light chain is either an antibody polypeptide fragment or host cell protein (albeit at a higher than usual level). As can be seen in the case of NISTmAb, there are no SDS-PAGE detectable fragments or host cell proteins.

### *Capillary Electrophoresis with Sodium Dodecyl Sulfate*

In recent years, CE-SDS has been shown to be a reliable alternative for the replacement of SDS-PAGE due to its inherent advantages, including the use of a replaceable UV-transparent hydrophilic polymer as the sieving matrix; automated instrumentation; high-throughput efficiencies; enhanced resolution and reproducibility; and in-line detection capabilities with absorbance and fluorescence detectors, allowing better and more accurate quantitation (27). Because of these advantages, CE-SDS has gained tremendous usage from the biopharmaceutical industry for control of both product- and process-related impurities as they relate to size heterogeneity. Therefore, quality control strategies often utilize CE-SDS in good manufacturing practice (GMP) batch release and stability programs. Like SDS-PAGE, the CE-SDS profile, or electropherogram, provides a characteristic fingerprint of the size variants produced during product manufacture, including various combinations of light- and heavy-chain antibody fragments (disulfide-linked subunits), non-specific polypeptide cleavages products, non-dissociable high molecular weight aggregates, proteins produced by the host cell, and cross contamination of products produced in the same facility (28).

Although companies develop efficient manufacturing processes, including several high-capacity purification steps, characterization of drug substance and drug product materials often shows the presence of minor impurities related to various antibody fragments or, in rare cases, proteins derived from the production cell. The fragmentation fingerprint or peak profile of the NISTmAb observed by CE-SDS with UV detection (Figure 5A) is consistent with those that have been well characterized (24). Under non-reducing conditions, Peaks 1 through 5 of the NISTmAb profile correlate with fragment bands observed in SDS-PAGE and are presumably identified as free subunits, such as free light chain and free heavy chain, as well as disulfide-linked subunits, such as H:L, heavy chain dimer (H:H), and heavy-heavy-light (H:H:L) species, respectively. Although sample buffer pH and free-thiol alkylation have been shown to control reduction events and prevent sample preparation artifacts (29), it still remains uncertain as to how much of the observed fragments are induced through the relatively harsh denaturing conditions during sample preparation. Non-reduced CE-SDS also detects non-enzymatic hinge region fragmentation that generate free, ~50 kDa species (e.g., Fab, Fc) and ~100 kDa species (e.g., desFab, F(ab')<sub>2</sub>), but usually migrate with size variants of similar molecular weight, heavy chain (Peak 2) and H:H (Peak 4), respectively. Even smaller low molecular weight clips are

detected in the region prior to the free L fragment (labeled as \*). Peak 6 is presumably the intact IgG minus this small clip species. Under optimal sample preparation conditions, the NISTmAb contains about 2.5% total fragments (i.e., sum of peaks before the main peak), with about 97.5% intact IgG. It should be noted that the UV detector has limited sensitivity with minor forms, such as the high molecular weight aggregate forms that are typically separated in SEC. High order aggregates can exist in covalent forms or forms loosely associated by hydrophobic or electrostatic interactions. The latter generally dissociate at high temperature during SDS-protein complexation in sample preparation. CE-SDS is capable of separating covalent aggregates from the monomeric form, but these aggregates are not observed in the NISTmAb by UV analysis. This may be due to the challenge from the fluctuations observed in the baseline caused by absorption characteristics of the proprietary SDS-MW gel. Thus, quantitation for UV profiles should be reported with valley-to-valley peak integration.

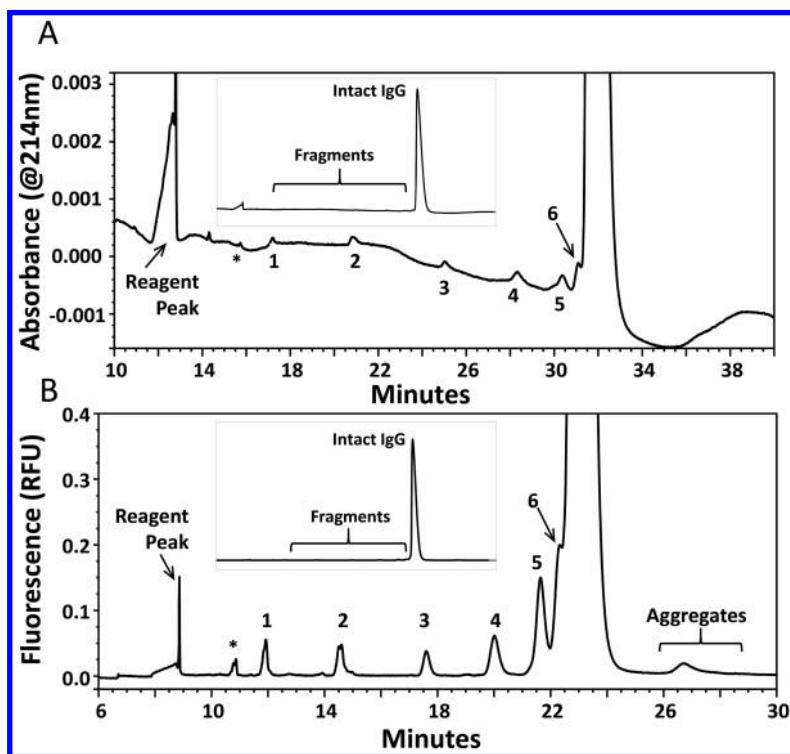


Figure 5. Expanded-view size heterogeneity profiles of the NISTmAb as determined by non-reduced capillary electrophoresis-sodium dodecyl sulfate (CE-SDS) with (A) UV detection, and (B) laser-induced fluorescence (LIF) detection. The capillary cartridge used in A and B was operated at different temperatures, resulting in noticeable migration time differences. Antibody fragments are labeled as peaks 1 through 6 and unknown (\*). The inset is the full-scale profile.

Enhanced sensitivity is achieved with pre-column fluorescent labeling coupled with LIF detection. The use of 5-carboxytetra-methylrhodamine succinimidyl ester (TAMRA.SE) as a fluorophore was first employed by the biopharmaceutical industry due to its simple reaction with the abundant aliphatic epsilon-amines of lysine residues present in recombinant antibodies (27). In addition, the use of TAMRA.SE was also shown to substantially improve detection sensitivity while maintaining a genuine representation of the sample. Despite these advantages, labeling proteins is time-consuming, and may accelerate protein aggregation as hydrophobic moieties are added by each fluorophore molecule. In recent years, fluorogenic reagents like FQ have gained much attention for CE-SDS/LIF applications due to their distinct advantages in chemical structure and fluorescence properties (6, 30). Because the free dye is weakly fluorescent and migrates ahead of protein in the reagent peak, removal of the unreacted dye after protein labeling is not necessary, thereby simplifying the protocol during sample preparation. Furthermore, the use of FQ reagents has been shown to have no impact on antibody aggregation, even at very high dye-to-protein molar ratios, therefore improving the accuracy and method robustness in QC labs.

Advantages of LIF detection are illustrated in the profile of Figure 5B, which shows the separation of FQ-labeled NISTmAb by CE-SDS under non-reducing conditions. Compared to UV detection, the performance and quality of the baseline with LIF detection is far superior and practically allows a single uniform baseline for peak integration starting after the reagent peak until the end of the aggregate peak. The combination of such baseline performance and the enhanced detection sensitivity allowed the LIF assay to detect low-level aggregate forms that were not observable with UV detection. Control of dye-related artifacts was validated by comparing the fragment fingerprint between the UV and LIF detection schemes. As shown in Figure 5, the peak patterns between UV and LIF profiles were comparable, and no atypical or new fragment peaks were observed in the LIF profile. In addition, data analysis of the LIF sample showed that the NISTmAb contains 97.9% intact antibody and 2.0% total fragments, with approximately 0.1% non-dissociable aggregates, which was comparable to the relative peak distribution by UV, demonstrating that the process of dye labeling was sufficiently controlled.

In order to better characterize the product purity, an aliquot of the prepared non-reduced pool was denatured with an excess amount of DTT at high temperature. This treatment effectively reduces both inter- and intra-chain disulfide linkages, converting the product to its most basic forms, mainly free light chain and free heavy chain, and any other low-level variants related to non-specific polypeptide cleavage products. The CE-SDS profile of the reduced NISTmAb is shown in Figure 6. It contains predominantly two main peaks combining for 98.0% of the total corrected peak area, each of which is consistent with the L and H bands observed in by SDS-PAGE (Figure 4). The non-glycosylated heavy chain (NGH) is smaller in hydrodynamic size compared to its glycosylated heavy chain counterpart containing a single *N*-linked carbohydrate site. Reduced CE-SDS is a practical and reliable assay that can accurately determine the percent occupancy of *N*-linked glycosylation on the heavy chain, which is calculated as the ratio of glycosylated heavy chain to the total amount of heavy chain variants based on

percent corrected peak area (i.e.,  $[\%H]/[\%NGH+\%H]$ ). The NISTmAb contains 99.2-99.3% glycosylated heavy chain (see Table 2). Non-reducible thioether cross-links between the light chain and heavy chain subunits of recombinant mAb products have been recently characterized (31). These non-reducible forms are detectable by CE-SDS under reducing conditions. By UV detection, approximately 1.0% of the NISTmAb has a non-reducible form of L-H dimer. Finally, the pre-L (Peaks a and b in Figure 6B) and pre-H (Peaks c and d in Figure 6B) variants observed in the NISTmAb are likely a mixture of product fragments that resulted from non-specific cleavage of the light/heavy chains. Extended characterization, such as excising bands from an SDS-PAGE gel, should be considered to confirm these species are not related to proteins produced by the host cell during manufacture. In combination, these minor forms are present in the NISTmAb at approximately 0.6%.

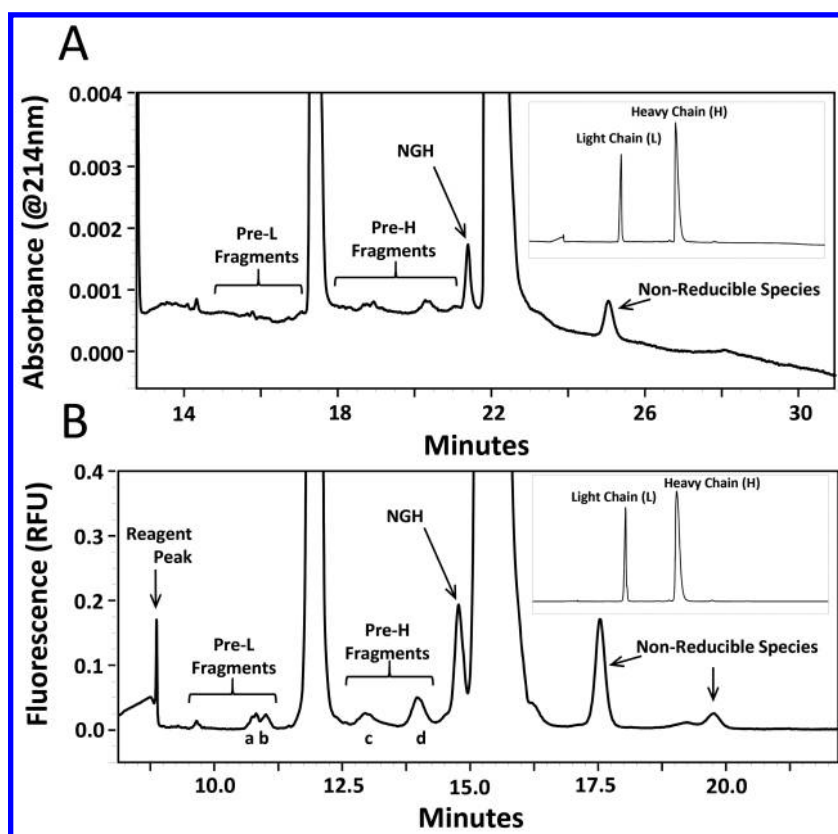


Figure 6. Expanded-view of the NISTmAb size heterogeneity determined by reduced capillary electrophoresis-sodium dodecyl sulfate (CE-SDS) CE-SDS with (A) UV detection, and (B) laser-induced fluorescence (LIF) detection. The capillary cartridge used in A and B was operated at different temperatures, resulting in noticeable migration time differences. Low molecular weight fragments are labeled as peaks a through d. The inset is the full-scale profile.



Product purity determined from LIF detection was comparable to that observed from UV detection and resulted in orthogonal CE-SDS profiles under reducing conditions (Figure 6B). The light chain and heavy chain peaks combine for 97.0% of the total peak area, and approximately 99.2% of the heavy chain was glycosylated.

A shoulder peak is observed on the trailing edge of the light chain, as shown in the expanded view of Figure 7. This shoulder peak of the NISTmAb is consistent with the shoulder observed in mAbs enriched with glycosylated light chain, which is detected as a late-migrating peak in reducing CE-SDS that uses both UV and LIF detection (32). Non-enzymatic glycation of endogenous protein occurs *in vivo* and is commonly observed in therapeutic protein products, often widely distributed among multiple lysine sites in both light and heavy chains (33, 34).

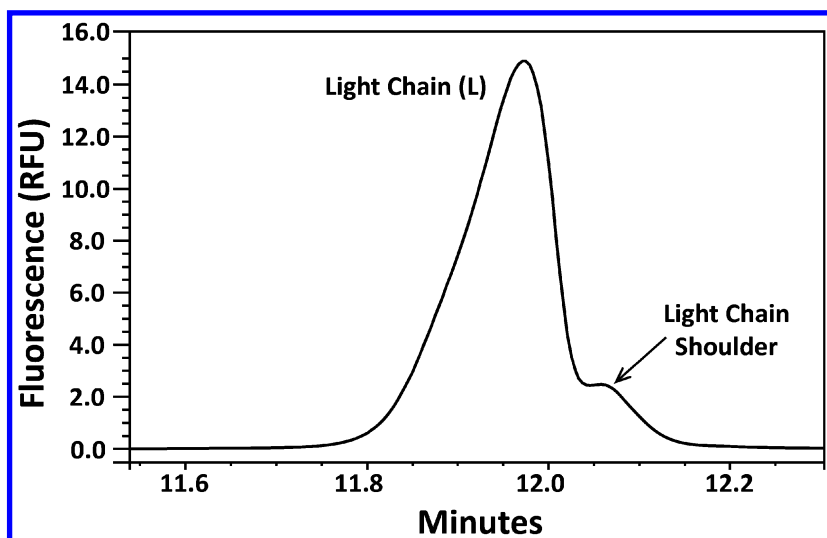


Figure 7. Light chain and glycosylated light chain variant observed in the NISTmAb by capillary electrophoresis-sodium dodecyl sulfate (CE-SDS) with laser-induced fluorescence (LIF) detection.

With the enhanced sensitivity and baseline performance of LIF detection compared to UV detection, the detection of fragments and non-reducible species was significantly improved. For instance, a higher level of the half-antibody covalent dimer (H:H) was clearly detected. The improved signal-to-noise in LIF detection enables users to assign approximate masses to the unknown fragment Peaks a, b, c, and d (Figure 6B). The molecular weight for each fragment peak was extrapolated from a linear fit of migration time and theoretical mass for L, H, NGH and half antibody. Based on the calibration curve ( $r=0.998$ ), estimates of the molecular weights for Peaks a through d were calculated as 12, 14, 32, and 41 kDa, respectively. Additional information from an accelerated degraded panel of the NISTmAb could be useful in the origin of these product related minor form fragments.

The linkage between critical process parameters and CQAs through QbD is now firmly embedded in the biopharmaceutical industry, driving advances in the manufacturing of therapeutic mAbs. Implementation of the QbD paradigm provides significant process and product understanding, which is obtained during process development, process characterization, and continuous process verification (process validation). As a consequence, there is an increased demand for high-throughput technologies capable of supporting the vast number of samples generated from the multivariate designs employed in bioprocess, purification, and formulation studies. In addition, screening studies in early development often utilize scaled down production due to expensive costs of materials that make it impractical to conduct cell culture runs at scale while maintaining good environmental practices, thereby pressuring analytics to improve sensitivity to accommodate much lower mass requirements. Although CE-SDS provides high resolution of size variants, the time scale for capillary-based methods makes it undesirable for high-throughput applications.

At 60 to 70 times faster than the speed of capillary methods, MCE-SDS can meet the testing demands for QbD applications by providing faster separations, lower mass needs, and higher sensitivity (35). In recent years, advances in the manufacture of small diode lasers coupled with commercial instruments such as the 2100 Bioanalyzer (Agilent Technologies) and the LabChip GXII (PerkinElmer) have enabled development of sensitive applications for low concentration samples (36). The advantages of MCE are illustrated for the NISTmAb in Figure 8, showing the separation of size variants in less than 60 seconds on the LabChip system, compared to 30 to 35 minutes in CE-SDS. Furthermore, higher resolution was achieved for Peak 6 with the microchip compared to the separation in CE-SDS. It has been reported that the microchip cross-T injector design facilitates injection of a well-defined narrow sample plug of only a few hundred nanoliters, limiting band-broadening of the sample zones during separation in the very short channel (37). Separation of the NISTmAb under reducing conditions detected similar size variants as shown in CE-SDS, including sufficient separation of the NGH variant. As a means to maximize sensitivity, the NISTmAb was derivatized with the vendor-supplied fluorophore using a custom labeling procedure, which generated signal-to-noise and baseline performance similar to CE-SDS with LIF detection. Purity analysis by MCE-SDS showed that the NISTmAb contained 98.2% intact antibody, 98.4% total light and heavy chain, with 99.2% heavy chain glycosylation.

Results from this section demonstrate that values of the NISTmAb obtained with three different methods gave comparable results for CE-SDS with UV detection, CE-SDS with LIF detection, and MCE-SDS for both non-reducing and reducing conditions.

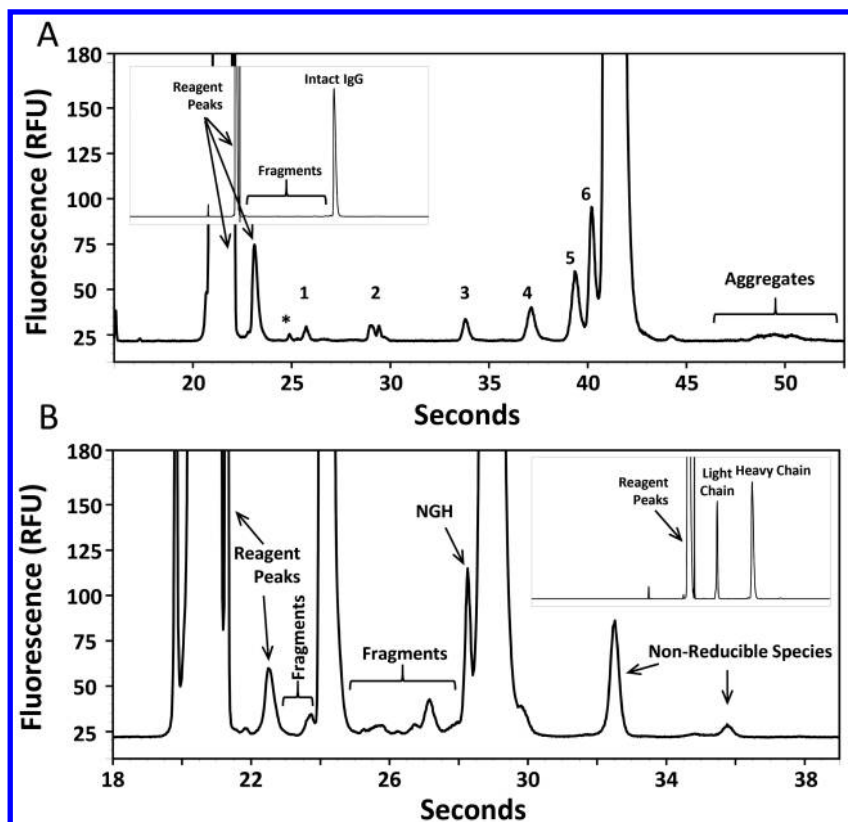


Figure 8. Size heterogeneity of the NISTmAb determined by microchip electrophoresis with SDS (MCE-SDS) under (A) non-reducing, and (B) reducing conditions. Antibody fragments are labeled as peaks 1 through 6 and unknown (\*). Full-scale profiles are shown in the inset.

The quantitative results of the size variant analyses are summarized in Table 1 and Table 2. Higher and lower molecular weight species were detected by SEC-UHPLC at 1.0% and 0.3%, respectively. No minor species were detected by SEC-HPLC when run with a different mobile phase, suggesting that further optimization of the SEC-HPLC may be required. Biophysical analysis of the NIST standard (Biophysical chapter/Volume 2, Chapter 6) confirms that a small amount of higher molecular weight species is present in the NISTmAb; therefore, the higher molecular weight species was likely not recovered under the SEC-HPLC conditions used. CE-SDS and MCE-SDS with LIF detectors detected low levels of aggregates (0.1 and 0.2%, respectively), indicating that a small portion of the dimer detected by SEC-UHPLC is covalent but most is non-covalent. Smaller molecular weight forms detected by CE-SDS with UV and LIF detection were 2.0-2.2% and 1.6% by MCE-SDS, considerably larger than the 0.3% detected

by SEC-UPLC. Some of the smaller molecular weight forms detected by the electrophoresis methods likely exist as non-covalent complexes and thus do not resolve from the main peak by SEC-UHPLC under native conditions. Under reducing conditions, the CE-SDS and MCE-SDS methods detected 0.4% to 0.6% fragments in total and 0.4 to 0.6% non-reducible species. No minor species were visible in the SDS-PAGE gels, indicating that the method is not as sensitive as CE-SDS and MCE-SDS with the sample load and staining method used. The CE-SDS and MCE-SDS methods were able to resolve glycosylated and non-glycosylated heavy chain, thus determining the glycosylation occupancy to be 99.0 to 99.2%. These values are in excellent agreement with that determined by mass spectrometric analysis of glycopeptides (Glycosylation chapter/Volume 2, Chapter 4). The aglycosylated heavy chain was not detected by SDS-PAGE. In addition, the reduced CE-SDS method with LIF detected resolved a light chain shoulder consistent with glycated species. The existence of glycated species in the NIST standard was confirmed by mass spectrometric techniques (PTMs chapter/Volume 2, Chapter 3).

**Table 1. Summary of NIST Monoclonal Antibody Size Variant Analyses (Non-Reducing Conditions)**

<i>Platform</i>	<i>Smaller Molecular Weight Species (%)</i>	<i>Main Peak (%)</i>	<i>Higher Molecular Weight Species (%)</i>
SEC-HPLC	ND	100.0	ND
SEC-UHPLC	0.3	98.7	1.0
SDS-PAGE*	Detected	<100	ND
CE-SDS (UV)	2.2	97.8	ND
CE-SDS (LIF)	2.0	97.9	0.1
MCE-SDS	1.6	98.2	0.2

ND = not detected, SEC-HPLC = size exclusion HPLC, UHPLC = ultrahigh-pressure liquid chromatography, SDS-PAGE = sodium dodecylsulfate-polyacrylamide gel electrophoresis, CE = capillary electrophoresis, UV = UV detector, LIF = laser-induced fluorescence detector, MCE = microchip electrophoresis. \* SDS-PAGE analysis is qualitative.

**Table 2. Corrected Peak Area for Reduced NIST Monoclonal Antibody Standard**

<i>Platform</i>	<i>Sum of Fragments (%)</i>	<i>Light Chain (%)</i>	<i>Heavy Chain (%)</i>	<i>Non-Reducible species (%)</i>	<i>Heavy Chain Glycosylation*</i> (%)
SDS-PAGE**	ND	Qualitative		ND	NA
CE-SDS (UV)	0.6	32.3	66.1	0.4	99.3
CE-SDS (LIF)	0.6	30.4	67.6	0.6	99.2
MCE-SDS	0.4	30.7	67.4	0.6	99.2

ND = not detected, NA = not available, SDS-PAGE = sodium dodecylsulfate-polyacrylamide gel electrophoresis, CE = capillary electrophoresis, UV = UV detector, LIF = laser-induced fluorescence detector, MCE = microchip electrophoresis. \* Calculated as [% Heavy Chain] ÷ [% non-glycosylated Heavy Chain + % Heavy Chain]. \*\* SDS-PAGE analysis is qualitative.

### Hydrophobic Variant Analysis

As a general class, mAbs are considered hydrophobic molecules. Many mAb post-translational modifications cause changes in hydrophobicity, especially changes in surface hydrophobicity, which may make them amenable to resolution by separation assays. These include oxidation and isomerization of aspartic acid (38), which decrease hydrophobicity; and others such as N-terminal pyroglutamic acid (39), C-terminal lysine variants (40), incomplete removal of heavy chain leader sequence (41), reduced disulfide bonds/open loop forms (42), and trisulfide bond (43, 44), which increase hydrophobicity. In addition, many of the size variants discussed earlier (clips, fragments, partial molecules, and aggregates) also result in hydrophobicity differences.

The use of RP-HPLC for large, multi-domain proteins such as mAbs has gradually increased over the past decade, in part due to the need for higher resolution analytical techniques that can be coupled with on-line high resolution liquid chromatography-mass spectrometry (LC-MS) for characterization of the resolved species. For IgG1 mAbs such as the NISTmAb, intact reversed-phase analysis is a powerful technique for monitoring and characterizing fragmentation, structural isoforms, and partial reduction of disulfide bonds. Reduced reversed-phase analysis provides a more detailed view of translational modifications on the individual light and heavy chains, although structural variants related to disulfide bonds are lost. RP-HPLC can also be performed following limited proteolysis. The mAb is digested into scFc and F(ab')<sub>2</sub> fragments, enabling some gains in resolution, as well as assignment of specific variants to a particular region of the IgG and retention of the disulfide structure.

Although RP-HPLC provides excellent resolution and the ability to perform on-line mass spectrometry, the determination of potency for collected fractions is problematic due to the denaturing conditions of the assay. Non-denaturing

orthogonal separation techniques such as HIC can be used to isolate hydrophobic variants, although the resolution capability is usually much lower.

### *Reversed-Phase HPLC*

In reversed-phase chromatography, mAbs are separated based on relative hydrophobicity under denaturing conditions by adsorption on a hydrophobic column followed by a gradient elution with an increasing concentration of an organic solvent. Acetonitrile, methanol, ethanol, and n-propanol are the most commonly used organic solvents.

The successful development of a reversed-phase mAb method is often considered a challenging task due to the molecule's size, the structural complexity inherent to mAbs arising from four separate protein chains, the presence of 16 to 18 disulfide bonds (IgG1, IgG2, and IgG4), and numerous post-translational modifications. Their large size and hydrophobic nature generally result in low recovery and poor resolution in reversed-phase separations unless certain critical parameters are met (45). Acetonitrile is the most widely used solvent for mAb analytics, but low protein recovery and resolution can often arise for more hydrophobic mAbs and certain IgG subclasses (40). Higher elutropic-strength solvents such as n-propanol have shown promise as a more universal solvent for reversed-phase separation of mAbs (45, 46). In addition, low pH (~2) and high column temperature have been shown to be critical parameters for protein recovery and peak shape. In fact, column temperatures of at least 65 to 70 °C should be used to provide optimal protein recovery and resolution (40, 47). TFA is typically used to provide an acidic pH and as an ion-pairing agent, decreasing non-specific protein interactions with the column. Formic acid can be used in place of TFA for LC-MS applications, but it can cause changes in selectivity and lower resolution of certain analytes (48).

Generally, with intact RP-HPLC, smaller and less-hydrophobic protein fragments are eluted earlier in the reversed-phase gradient, at a lower concentration of organic solvent. The intact mAb, as well as other more hydrophobic species, are eluted later in the gradient. The reversed-phase analysis of the NISTmAb IgG1 standard is shown in Figure 9A. The sample was resolved into a single broad main peak (93.5% total relative area), several earlier eluting pre peaks, labeled as Fragments (1.0% total relative area), and two later eluting post peaks, labeled IgG structural variants (5.5% total relative area). In this reversed-phase analysis, the main peak is composed of monomeric IgG, including the combination of all individual glycosylation forms as well as C-terminal and N-terminal variants. The peaks labeled Fragments are generally composed of non-covalently attached products of the IgG, often caused during production and/or during storage. For IgG1 mAbs, common fragments may include partially reduced species L, H, combinations of L and H (49) and hydrolyzed antibody domains (e.g., Fab', Fc, V<sub>H</sub> [variable domain of antibody heavy chain], C<sub>H</sub>3 [constant domain 3 of antibody heavy chain]) (40). Late-eluting peaks related to structural variants have been reported to be disulfide related, often containing unpaired or mis-paired disulfides. More recently, the existence of trisulfide bonds in mAbs has been

reported (43, 44). In order to positively identify each reversed-phase peak, fractions can be collected for subsequent characterization by peptide mapping, or fractions can be directly characterized with in-line high-resolution MS. MS analysis provides direct information about the molecular mass of each eluted species, and in some cases can provide details about the amino acid sequence.

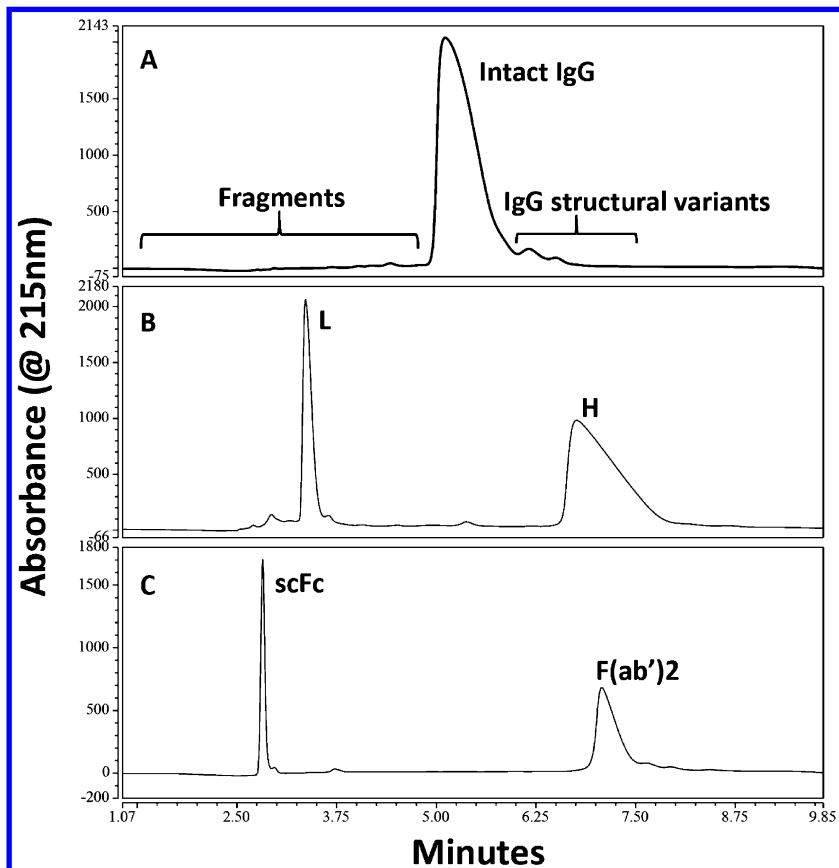


Figure 9. Reversed-phase chromatograms showing a full-scale view for the NISTmAb by intact (A), reduced (B), and (C) FabRICATOR<sup>®</sup> digest.

Although the intact reversed-phase analysis of the IgG1 standard was able to provide a relatively detailed assessment of product quality, for IgG fragments and structural variants, additional information can be gained by analyzing smaller individual domains of the IgG by reversed-phase. Figure 9 shows a comparison of the reversed-phase analysis of the intact NISTmAb (A), after reduction of disulfide bonds (B), and after digestion with the FabRICATOR<sup>®</sup> enzyme (C), a commercially available recombinant modified enzyme with precise proteolytic activity for IgGs. The FabRICATOR<sup>®</sup> enzyme cleaves below the IgG hinge, creating an F(ab')<sub>2</sub> and two scFc fragments (50, 51). The F(ab')<sub>2</sub> portion of the IgG is composed of two L and two H fragments, including the variable heavy (V<sub>H</sub>) and the first constant domain of the heavy chain (C<sub>H1</sub>). The four chains

(2 × L and 2 × H) of the F(ab')<sub>2</sub> are covalently attached through four interchain disulfide bonds (IgG1), with an approximate molecular weight of 100 kDa. The monomeric scFc fragments contain the second and third constant domains of the heavy chain (C<sub>H2</sub> and C<sub>H3</sub>), which include the conserved N-linked glycosylation site. The two scFc fragments, each with an approximate molecular weight of 25 kDa, are not covalently linked through disulfide bonds.

Figure 10 shows the same comparison but with a magnified baseline so that the smaller peaks can be displayed. As shown in Figure 9 and Figure 10, the post peaks observed in the intact analysis (panel A) look quite similar to the post peaks of the F(ab')<sub>2</sub> in the FABRICATOR<sup>®</sup> digest (panel C), indicating that the variants likely occur in the F(ab')<sub>2</sub> region, although without MS confirmation, it is not possible to confirm the identity of the post peaks.

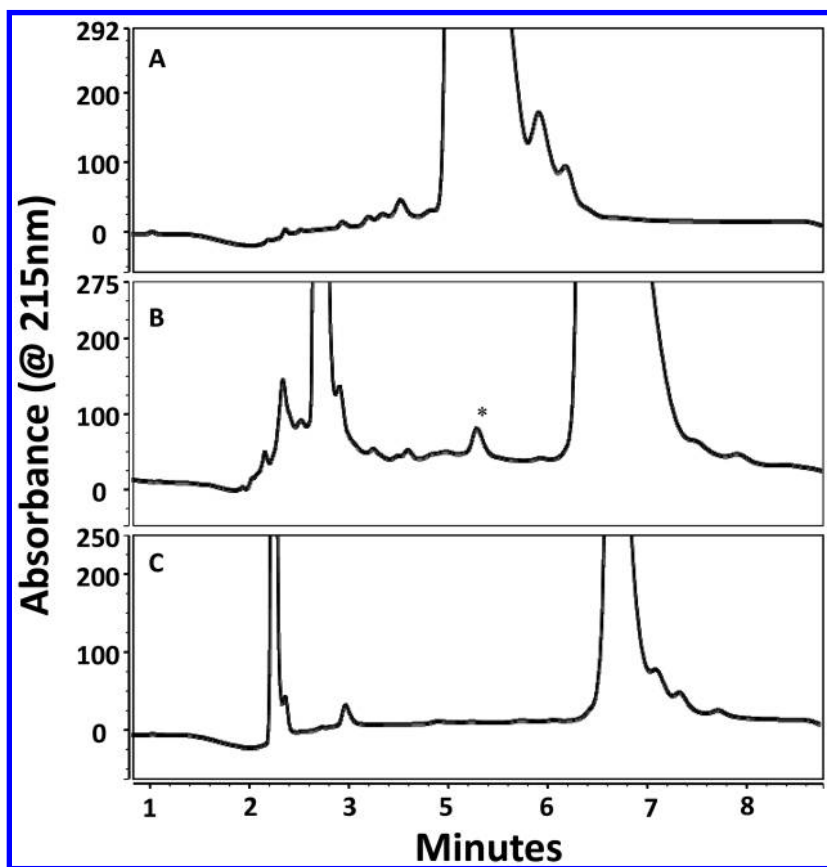


Figure 10. Reversed-phase chromatograms showing a zoomed view for the NISTmAb by (A) intact, (B) reduced, and (C) limited proteolysis. The peak marked with an asterisk in panel B represents non-reducible IgG.



Reduced reversed-phase analysis provides a more detailed view of post-translational modifications on the individual light and heavy chains. Although the reduced reversed-phase analysis can reveal additional modifications (relative to intact analysis) such as oxidation, structural variants related to disulfide bonds are typically not detectable after reduction. Reduced reversed-phase analysis of the NISTmAb is shown in Figure 9B and Figure 10B. Although reduction with DTT will reduce the IgG to its light and heavy chain components, some modified disulfide bonds, such as thioethers, are not reducible (52) and can appear as residual intact IgG or partially reduced species, similar to reduced CE-SDS. The peak labeled with an asterisk in Figure 10B is thought to represent non-reducible IgG, but without MS confirmation, this is a tentative assignment.

### *Hydrophobic Interaction Chromatography*

HIC separates analytes in order of increasing hydrophobicity; in other words, less hydrophobic species typically elute first. A weakly hydrophobic stationary phase is used to bind the analyte in the presence of high concentrations of polar salts. A mobile phase gradient of decreasing salt concentration is applied. As the concentration of salts decreases, analytes desorb into the mobile phase. Selection of a more or less hydrophobic stationary phase, can be used to adjust the selectivity of the separation for the particular analyte or analytes of interest. Compared to other modes of separation, such as CEX- or SEC-HPLC, HIC offers the opportunity to separate molecules based on changes in hydrophobicity that may be imparted by chemical modifications such as oxidation. HIC offers a unique mode of separation that typically does not irreversibly disrupt the structure of protein analytes, allowing different isoforms to be collected and tested by orthogonal assays. The HIC chromatogram for the NISTmAb is shown in Figure 11. The protein elutes as one major peak. Two minor shoulders were observed to elute prior to the main part of the peak. This observation suggests the sample does not have significant heterogeneity with regard to hydrophobic variants, but minor resolution of some low-level hydrophobic variants could be detected. Purification and detailed analysis of these variants would need to be performed to identify these minor species. Possible causes for separation by HIC could include oxidation of the antibody or structural changes to the antibody caused by aspartic acid isomerization but are purely theoretical without further investigation. This method could be used to analyze stressed samples (e.g., heat, light) that might cause an increase in hydrophobic variants. Because this method is performed under conditions that do not irreversibly impact the structure of the antibody, collection of variants and testing by potency- or binding-related biological assays could be conducted to further understand the impact and significance of these species.

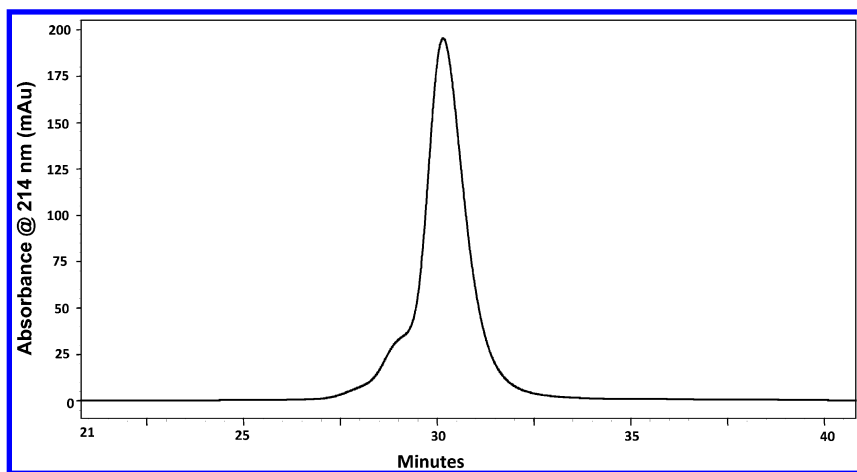


Figure 11. Hydrophobic interaction chromatography (HIC)-HPLC analysis of the NISTmAb.

The results of the hydrophobic variant analyses are summarized in Table 3. Pre peaks, presumably smaller molecular weight fragments, were 1.0% of the total peak area by intact RP-HPLC. Post peaks, thought to be structural variants, were 5.5% of the total peak area. After reduction, more minor species are resolved, with a total of 1.7% eluting before the light chain, 1.7% intermediate species between light and heavy chain (including 1.0% non-reducible species), and 1.3% eluting after the heavy chain. The structural variant peaks, resolved by intact RP-HPLC, are absent, confirming that these are likely disulfide-related. By limited proteolysis, minor species eluting between the scFc fragments and F(ab')<sub>2</sub> totaled 3.5%, including a prominent peak of approximately 1.1% that seems to correspond with free light chain. The post peaks (8.0%) resolved from the F(ab')<sub>2</sub> peak appear to correspond by abundance and relative retention to the same structural variants resolved in the intact RP-HPLC. By HIC-HPLC, two minor shoulders (less hydrophobic variants), eluting prior to the main peak, were resolved, but the resolution was not sufficient for meaningful quantitation. The post peaks resolved by RP-HPLC were not detected by HIC-HPLC, likely due to differences in the resolution between the two assays.

**Table 3. Summary of Hydrophobic Variant Analyses**

<i>Platform</i>	<i>Pre Peaks/Mid Peaks/ Hydrophilic Variants (%)</i>	<i>Main Peak(s) (%)</i>	<i>Post-Peaks Hydrophobic Variants (%)</i>
Intact RP-HPLC	1.0	93.5	5.5
Reduced RP-HPLC	1.7 (pre L) 1.7 (post L)	27.0 (L) 68.3 (H)	1.3 post (H)
Limited Proteolysis	3.5 (post scFc)	29.9 (scFc) 58.6 (F(ab') <sub>2</sub> )	8.0 (post-F(ab') <sub>2</sub> )

RP-HPLC = reversed-phase HPLC, L = light chain, H = heavy chain, scFc = single chain Fc fragment.

### Charge Variant Analysis

Maintaining consistent product quality of mAbs often requires a variety of orthogonal methods to characterize charge heterogeneity. Charge heterogeneity can result from both the manufacturing process and degradation during storage. Heavy chain C-terminal lysine, a common post-translational modification, is believed to be the result of incomplete proteolysis by endogenous mammalian host basic carboxypeptidase(s) during the cell culture operation (53). Glycation, discussed earlier (see *Capillary Electrophoresis with Sodium Dodecyl Sulfate*), also can occur during cell culture and results in charge variants. C-terminal  $\alpha$ -amidated proline is another charge variant that has been reported for mAbs (54) as arising during cell culture. Glycans with negative charges, such as sialylated glycans, are another example of charge variants (55). Deamidation is probably the most common charge variant resulting from degradation of mAbs (56, 57).

Ion-exchange HPLC has historically been the gold standard for determination of charge heterogeneity of proteins. Ion exchange HPLC is typically performed under native conditions, enabling collection of the resolved forms for further characterization, including biological activity assessment with the appropriate potency assay. More recently, electrophoresis has proven to be a useful tool for protein characterization. For instance, slab gel IEF separates protein charge isoforms with high resolution, but its labor-intensive protocols and high assay variability render it less useful for routine application in quality control testing labs. The introduction of capillary-based methods (i.e., cIEF) for determination of protein charge heterogeneity provides key advantages over classical gel formats, including on-line detection, automation, full quantitative analysis, and higher assay precision. This section discusses a variety of orthogonal charge variant assays and their applications for determination of charge distribution (quantitative) and apparent pI value.

## Cation Exchange HPLC

Ion exchange separations are useful tools for the characterization of mAbs in that they typically provide good resolution of many charge variants, especially stability-indicating variants such as deamidation. Ion exchange assays can also complement charge variant electrophoresis assays such as cIEF or CZE, because many of the same charge variants are resolved, and these variants are more readily isolated for further characterization by ion exchange. Charge variants are separated by ion exchange based on the molecule's surface charge. Cation exchange is typically used for mAbs due to the relatively high pI of these molecules. Due to the unique pI and surface charge distribution of mAbs, cation exchange methods usually require optimization specific for each mAb. Such optimization typically requires screening of columns, mobile phase pH, and gradient conditions at a minimum. The CEX-HPLC method for the NIST standard was developed and optimized using a DOE approach (see Appendix 1).

The optimized CEX-HPLC chromatogram for the NIST standard is shown in Figure 12. One major acidic peak with a slight shoulder is observed in the pre-peak region of the chromatogram. One large basic peak and two smaller basic peaks are observed in the post main peak region. Figure 12 also shows the chromatogram following CpB treatment of the NISTmAb. The effects of CpB treatment on the sample are clear. The dominant basic peak and the second smaller basic peak are reduced following CpB treatment, indicating that these two basic peaks contain C-terminal lysine (58, 59) on one heavy chain or on both heavy chains.

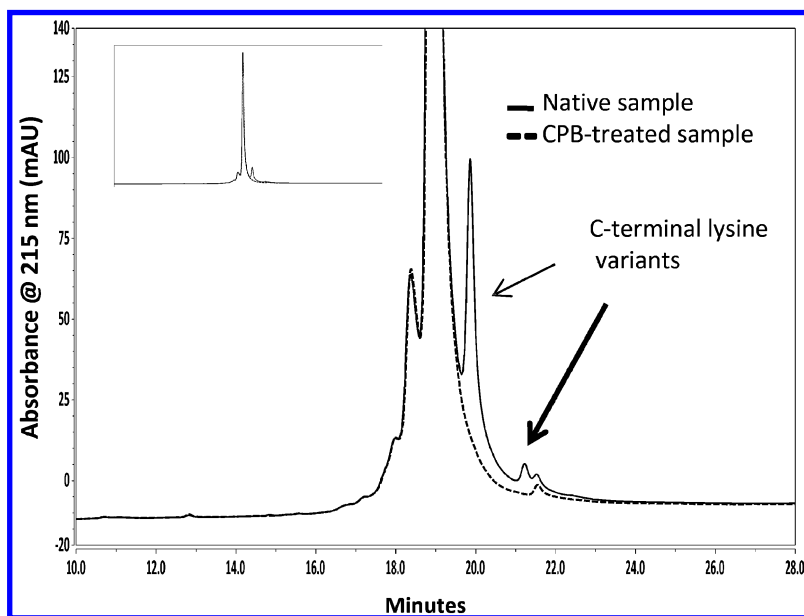


Figure 12. Cation exchange (CEX)-HPLC chromatograms of the NISTmAb before and after carboxypeptidase B treatment.

CEX-HPLC is also useful for resolving deamidated forms as acidic pre peaks. The overlaid chromatograms in Figure 13 show the increasing amount of acidic peak group under conditions intended to induce deamidated species (pH 8.5, incubation at 37 °C). Further characterization is needed to confirm the identity of the components of the CEX-HPLC pre peaks, because several different species may be contained under each peak. However, the incubation studies demonstrate that the method is capable of resolving stability-indicating species.

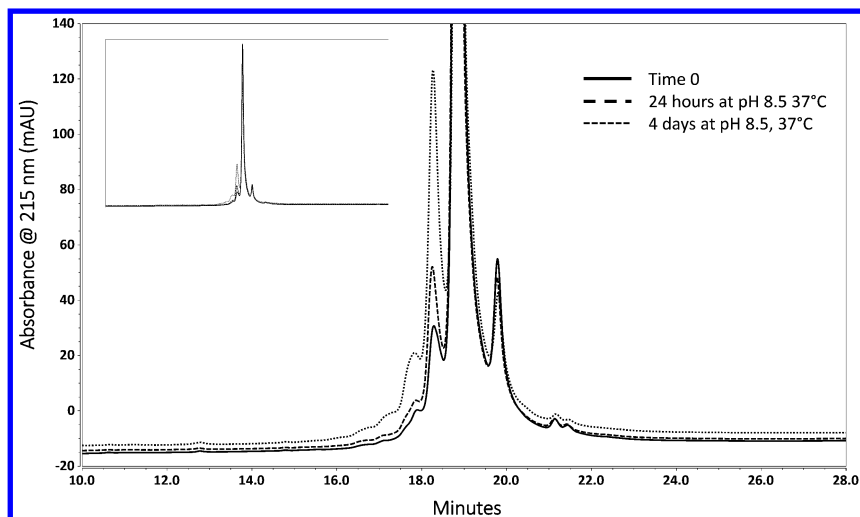


Figure 13. Cation exchange (CEX)-HPLC chromatograms of the NISTmAb following incubation.

### Capillary Zone Electrophoresis

CZE separates proteins in free solution according to electrophoretic mobility, which is largely influenced by buffer viscosity, the protein net charge, and hydrodynamic radius. Because the molecular size distribution of antibody products is typically higher than 95% monomer (measured by SEC), the effect associated with hydrodynamic radius differences is eliminated and allows CZE to effectively separate protein based on charge under native conditions. Figure 14A shows the charge heterogeneity profile of the NISTmAb separated by CZE with the platform pH 5.7 running buffer.

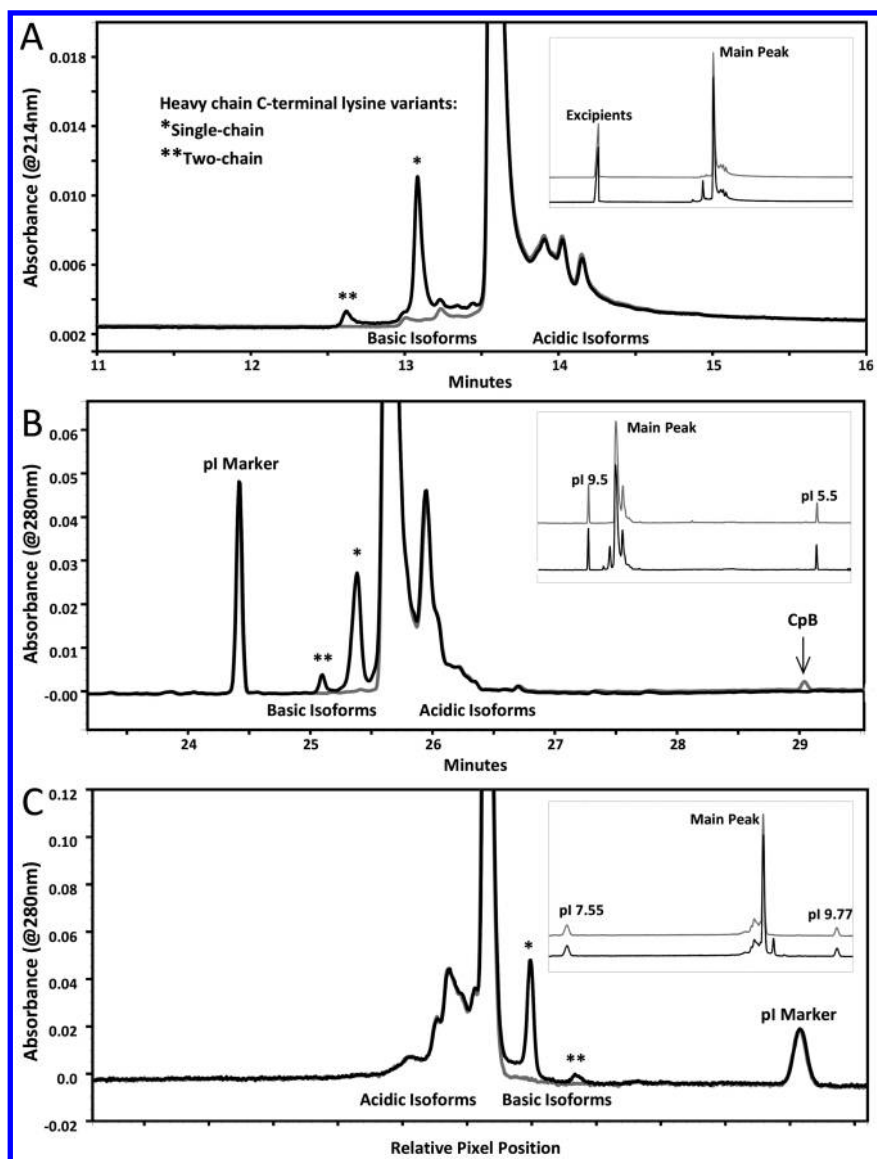


Figure 14. Charge heterogeneity of the NISTmAb determined by (A) capillary zone electrophoresis (CZE), (B) capillary isoelectric focusing (cIEF), and (C) imaged capillary isoelectric focusing (ICIEF). The inset in each figure shows the full-scale profile. Each figure contains an overlay of native (black) and carboxypeptidase B (CpB)-treated (grey) sample.

Although the acidic and basic isoforms observed in the profile are sufficiently resolved from the main peak, additional optimization experiments to enhance resolution may be explored by adjusting the pH of the running buffer or changing various electrophoretic parameters. As a protein characterization tool, CZE can be used to measure variability of heavy chain C-terminal lysine. Peaks associated with C-terminal lysine were determined by comparing a native NISTmAb sample to one that was treated with CpB enzyme. As shown in the overlay, the two major basic peaks affected by CpB treatment suggest that these forms correlate with lysine present on one heavy chain or on both heavy chains. The primary effects observed with CpB treatment is a decrease in basic forms with a corresponding increase in main peak; the acidic variants remained unchanged. Data analysis shows that the C-terminal heavy chain of the NISTmAb is approximately 9% C-terminal lysine-containing, with 24% acidic variants.

### *Capillary Isoelectric Focusing*

cIEF has become widely used as the primary alternative to traditional slab gel IEF for routine protein characterization (17). Protein is separated in cIEF based on differences in pI, the pH at which the analyte possess a net zero charge. cIEF has the resolving capacity of separating protein differences as low as 0.01 pI units. Within the last decade, the biopharmaceutical industry has used this technology in a variety of applications, including the consistency of charge heterogeneity, product identity, and apparent pI determination. The Beckman cIEF platform assay uses fixed-point detection and consists of two steps to complete the experiment. The first step of the assay is focusing the carrier ampholytes, stabilizers, and protein charge isoforms during a process in which a pH gradient forms through the introduction of hydroxyl ions and hydronium ions from the anolyte and catholyte, respectively. With enough basic stabilizer, the pH gradient containing the focused zones is pushed toward one side of the detector window. The second step of the assay is chemical mobilization, which drives mobility of the focused zones, including ampholytes, across the fixed-point detector. Hydrodynamic mobilization is possible but is less desirable due to its effect on peak distortion and resolution.

The charge heterogeneity profile of native and CpB-treated NISTmAb analyzed with the Beckman cIEF platform assay is shown in Figure 14B. The separation profile of the NISTmAb has a striking similarity to the profile observed in the CZE assay, particularly for the two basic peaks attributed to the heavy chain C-terminal lysine variants. Only low levels of basic variants were detected after CpB treatment. Although the acidic variants are well-resolved from the main peak and not influenced by CpB treatment, resolution is quite different compared to the CZE assay but very similar to the CEX-HPLC assay (Figure 12). The profiles for the acidic variants highlight the differences in separation mechanism between cIEF and CZE assays. Given the nature of the CQAs impacting acidic variants, the selection of the final charge method is most informative using degraded materials (as shown in Figure 13 for CEX-HPLC) and peak characterization techniques such as peptide mapping and MS for identification of the peak components.

Although a platform assay was developed and optimized for mAbs, some products may benefit from adjusting critical method parameters for improving resolution. Focusing time is an example of a critical parameter that can affect separation quality: focusing time set too low may lead to incomplete focusing of large proteins into a narrow zone and result in poor method reproducibility; if set too long, hydrophobic proteins may become unstable as they concentrate into a narrow zone and precipitate in the capillary. A study of focusing time is illustrated in Figure 15, which shows the changes in resolution between the main peak and the first acidic peak when the experiment varied between 13 and 20 minutes. Samples focused below 17 minutes resulted in profiles having poor separation of the acidic variants from the main peak; however, at or above 18 minutes resulted in profiles having little to no change in resolution. Basic variants appear to focus much faster than the acidic variants and showed little dependency on focusing time. The selection of carrier ampholytes will also affect the profile. While the pH 3–10 Pharmalyte® provided sufficient separation of charge isoforms, enhanced resolution could be achieved with addition of a narrow-range ampholyte. These additional studies were not performed in this instance.

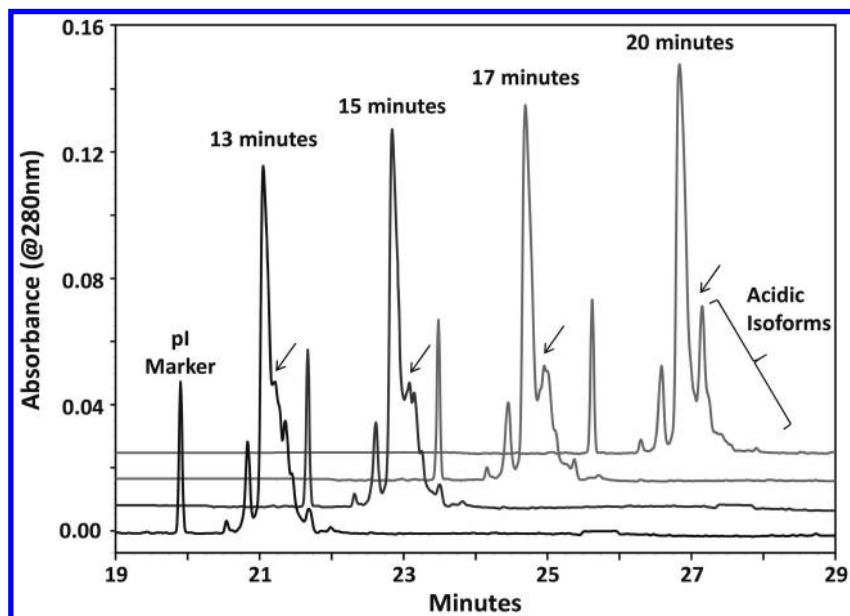


Figure 15. Effect of varied focusing times on the native NISTmAb. The arrow indicates the point of resolution between the main peak and the first acidic isoform.



## Imaged Capillary Isoelectric Focusing

Although the traditional two-step cIEF platform provides reliable characterization of charge isoforms, the mobilization step can introduce problems associated with distortion of the pH gradient after focusing, leading to poor reproducibility and resolution, as well as peak dispersion issues (17, 60). In recent years, whole-column or ICIEF technology was introduced with a configuration that eliminates the chemical mobilization step after focusing by capturing an UV image of the entire contents of the 5 cm separation capillary with a CCD camera. A major benefit to this one-step approach is the significant reduction in the total analysis time, which includes conditioning the 100  $\mu\text{m}$  i.d. capillary with methylcellulose, sample transfer injection, and isoelectric focusing. Approximately 15 minutes is needed to complete a run, improving the throughput by three-fold compared to the traditional two-step format.

Figure 14C shows the absorption image of native and CpB-treated NISTmAb obtained from the ICIEF platform. Charge heterogeneity is plotted as a function of spatial resolution (relative pixel position in the CCD camera) instead of the classic migration time. The ICIEF profile mirrors the profiles obtained by CZE and cIEF separations such that the basic isoforms focus ahead of the main peak instead of the trailing side seen in Figure 14A and Figure 14B. This is a result of the anode and cathode positioning relative to the detector. Maximum resolution was achieved through optimization of narrow-range carrier ampholytes. In this case, a ratio of 85% pH 8–10.5 and 15% pH 3–10 Pharmalyte® (2.5% total ampholytes) is ideal for basic proteins, providing optimal separation of the charge isoforms in 11 minutes. The same species containing single and two-chain C-terminal lysine variants were separated and confirmed by treatment with CpB. Only minor levels of basic isoforms were present after treatment. As with the other charge-based assays, the acidic isoforms were partially resolved from the main peak; higher levels of narrow-range ampholyte or longer focusing time did not improve the separation.

A quantitative summary of the CEX-HPLC and the electrophoresis methods for charge variant analysis are summarized in Table 4 and Table 5. CpB treatment confirmed that the majority of basic forms resolved by all four assays were C-terminal lysine variants. The quantity of basic isoforms with C-terminal lysine (Table 4) is similar for the three electrophoresis assays but higher for the CEX-HPLC assay. The single chain lysine variant form was not as well resolved from the main peak by CEX-HPLC, which may have adversely impacted quantitation. The same species was nearly base line resolved by the three electrophoresis assays. The acidic peaks are likewise similar between the electrophoresis methods; the acidic peaks are considerably lower by the CEX-HPLC methods, however, suggesting that the electrophoresis methods are resolving more species from the main peak than the CEX-HPLC method. Given that CEX-HPLC separates by surface charge and the electrophoresis method separates by overall net charge, some differences are to be expected.

**Table 4. Peak Area for Native NIST Monoclonal Antibody Standard**

<i>Platform</i>	<i>Acidic Isoforms (%)</i>	<i>Main Peak (%)</i>	<i>Basic Isoforms with C-terminal Lysine (%)</i>
CEX-HPLC	14.4	72.6	13.0
CZE	21.6	67.5	10.9
cIEF	24.1	66.6	9.3
ICIEF	24.2	66.6	9.2

CEX-HPLC = cation exchange HPLC, CZE = capillary zone electrophoresis, cIEF = capillary isoelectric focusing, ICIEF = imaged capillary isoelectric focusing.

**Table 5. Peak Area for Carboxypeptidase B-Treated NIST Monoclonal Antibody Standard**

<i>Platform</i>	<i>Acidic Isoforms (%)</i>	<i>Main Peak (%)</i>	<i>Basic Isoforms (%)</i>	<i>C-terminal Lysine* (%)</i>
CEX-HPLC	12.9	82.2	4.9	8.1
CZE	24.2	74.1	1.7	9.2
cIEF	27.2	72.4	0.5	8.8
ICIEF	27.3	72.1	0.5	8.7

CEX-HPLC = cation exchange HPLC, CZE = capillary zone electrophoresis, cIEF = capillary isoelectric focusing, ICIEF = imaged capillary isoelectric focusing. \* Calculated as [Basic Isoforms with C-terminal Lysine] – [Basic Isoforms].

The pI of an unknown protein is an important characteristic in formulation development screening studies. In traditional approaches, the apparent pI of an unknown protein was determined by measuring the spatial resolution between IEF bands of pure proteins with known pI values. Accurate determination of pI value is a challenge when using protein markers given the heterogeneity of producing these markers and their stability over long-term storage. Over the years, however, synthetic low molecular weight amphoteric peptides covering a broad pI range and having strong 280 nm absorbance were introduced as reliable alternatives to replace the protein markers. The assay typically requires three or more pI markers in the sample preparation to calibrate the pH gradient and enable approximation of the apparent pI for a given test sample. An example of the pI determination assay for the NISTmAb is shown in Figure 16. Here, the NIST protein of unknown pI was prepared in a mixture containing eight peptide markers and separated by ICIEF; it should be noted that a similar approach could be applied to the traditional two-step cIEF platform. The pixel position of each marker was plotted as a function of the pI value assigned by the manufacturer. It can be seen from the standard curve that the pH gradient formed under these conditions was linear with a regression coefficient of 0.9995. Using this curve

and the measured pixel position, the apparent pI value of the NISTmAb was determined to be 9.3, slightly higher than the theoretical pI of the NIST IgG1 molecule (pI of 8.9 calculated from the primary sequence).

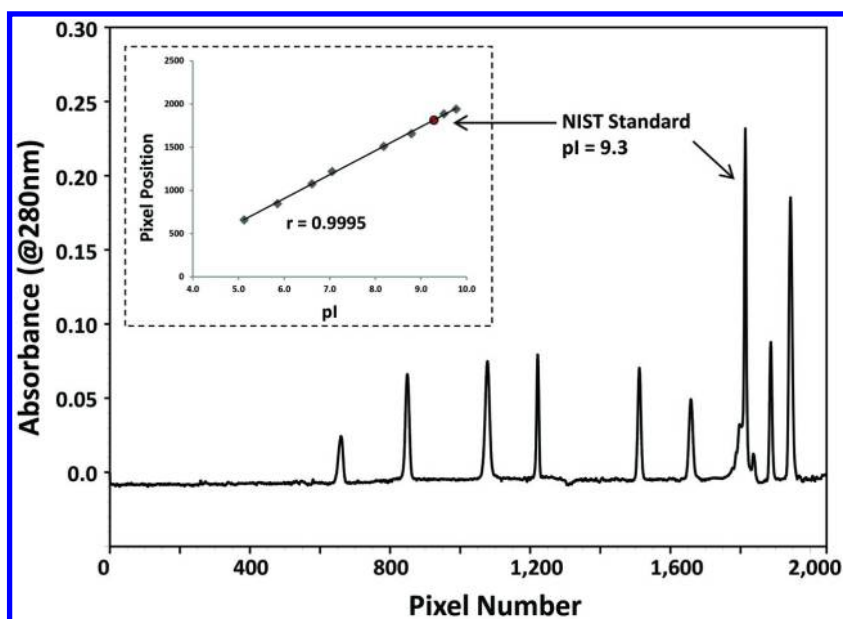


Figure 16. Experimental determination of apparent isoelectric point (pI). The NISTmAb was co-mixed with eight pI markers and separated by imaged capillary isoelectric focusing (ICIEF). A linear pH gradient was illustrated by plotting the pixel number and pI values from each marker and was used to extrapolate the apparent pI of the sample.

## Conclusions

Overall, the work in this chapter demonstrates the value of applying multiple orthogonal separation methods in the characterization of a mAb. Although separation methods are usually major components of a quality control programs for routine lot release and stability assessment, they are also important for the assessment of variants and the determination of those variants that are CQAs. The application of separation methods for characterization was demonstrated using the NISTmAb.

For size variant analysis, SEC is a key tool for the characterization of multimers. The native conditions of SEC enable the collection of fractions for potency determination. In addition, detectors such as MALS or QELS enable on-line characterization of the resolved species. The SEC-UHPLC analysis of the NISTmAb resolved a higher molecular weight species (1%) and a lower molecular weight species (0.3%). SEC-HPLC/MALS analysis determined that the higher molecular weight species was dimer. These results may be used in

conjunction with results from biophysical methods such as AUC to characterize the aggregate content of the NISTmAb.

SDS-PAGE and CE-SDS analyses provide further characterization information for size variants, especially smaller molecular weight components, which may be associated in non-covalent complexes and, therefore, may not be detected under the native conditions of SEC. CE-SDS is more quantitative and, in this work, proved to be more sensitive than SDS-PAGE for the detection of size variants. However, with higher protein loads and other staining techniques, SDS-PAGE may be a useful companion technique to CE-SDS in that the resolved bands may be excised for further characterization.

Both non-reduced and reduced CE-SDS were performed with three different methods: CE-SDS with UV detection, CE-SDS with LIF detection, and MCE-SDS. By non-reduced CE-SDS, the sum of the smaller molecular weight species was 1.6 to 2.2%, considerably higher than that determined by SEC, indicating that these species may exist as non-covalent complexes. However, such fragments can also be the result of reduction events during sample preparation; therefore, further confirmation by other techniques may be necessary. By non-reduced CE-SDS, aggregates were 0.1 to 0.2%, indicating that most of the dimer detected by SEC was non-covalent. In reduced CE-SDS, the intact mAb and any partial molecules are dissociated into the constituent light and heavy chains, thus the fragment species detected (0.4 to 0.6% for the NISTmAb) are the result of non-specific cleavage of the light or heavy chains. In addition, a total of 0.4% to 0.6% non-reducible species was detected, consisting of mostly H:H:L:L, with a small amount of larger molecular weight species. The glycosylation site occupancy was determined by reduced CE-SDS to be 99.2 to 99.3% of total heavy chain peaks, which is in excellent agreement with that determined in the Glycosylation chapter/Volume 2, Chapter 4.

Hydrophobic variant analysis by techniques such as RP-HPLC and HIC may resolve additional post-translational modifications such as oxidation or structural variants (e.g., reduced disulfide bonds, open loop forms). In addition, many of the smaller molecular weight species detected by the electrophoresis methods may also be resolved due to differences in hydrophobicity. RP-HPLC offers the advantage of on-line characterization through MS, whereas HIC analysis may be more valuable as a preparative tool because it can be used to isolate enriched species for determination of biological activity. The reversed-phase analysis of the intact mAb resolved several pre-peak species (typically fragments or reduced species such as free light chain, heavy chain, and so forth) totaling 1%, which is less than that observed by non-reduced CE-SDS, indicating that some of the fragments in the non-reduced CE-SDS analysis may have been artifacts of sample preparation. The intact RP-HPLC analysis also resolved two post peaks consistent with structural variants such as open loop forms.

RP-HPLC analysis was performed on both reduced mAb and after limited proteolysis to enable more detailed resolution of light and heavy chain variants or F(ab')<sub>2</sub> and scFv domains. By reduced RP-HPLC, several minor species were resolved, including a peak consistent with non-reducible H:H:L:L, which was also detected in the reduced CE-SDS analysis. Further work with an on-line mass spectrometer is recommended to definitively identify all of the RP-HPLC resolved

variants. As expected, the HIC provided lower resolution than the RP-HPLC and for the NISTmAb did not resolve variants as separate peaks. Two minor shoulders were detected.

Charge variant analysis was studied by CEX-HPLC, as well as three types of capillary electrophoresis methods (CZE, cIEF and ICIEF). The CEX-HPLC method resolved several pre peaks (acidic variants) totaling 14% and at least 3 post peaks (basic variants) totaling 13%. The two major post peaks were determined to be C-terminal lysine variants through the use of carboxypeptidase. The method was also demonstrated to be stability-indicating by incubating the NISTmAb under conditions intended to increase deamidated species. The expected increase in pre-peak species was observed.

All three capillary methods for charge variant analysis were able to resolve two major basic isoforms (total of 9 to 11%), similar to those observed by CEX-HPLC. The majority of these peaks were identified to be C-terminal lysine variants by use of carboxypeptidase, confirming the CEX-HPLC results. A similar level of acidic variants (22 to 24%) were resolved by all three electrophoresis methods. This level is higher than that observed by CEX-HPLC, indicating that more species were resolved by the capillary electrophoresis methods. In addition, the apparent pI of the NISTmAb was determined to be 9.3.

In this analysis of the NISTmAb material, we have demonstrated the utility of multiple orthogonal separation methods for characterization of a mAb. Due to the conserved properties of mAbs, many of the assays in this chapter may be directly applicable for other mAb molecules.

## Appendix 1. Method Development and Optimization through Design of Experiments

The following appendix demonstrates how a monoclonal antibody (mAb) cation exchange method can be rapidly developed and optimized using design of experiments (DOE) software tools.

In the DOE approach, multiple critical operational parameters (OPs) are tested simultaneously by using statistical models (i.e., bi-level, response surface, mixed level, model-robust) to automatically select the OP ranges to be tested (61, 62). The DOE software tool automatically generates test sequences which assess all of the relevant OP ranges using the fewest number of runs.

The three critical OPs selected for the CEX-HPLC method development for the NISTmAb were column type, buffer pH, and salt gradient. These three OPs were first tested in a general screening experiment covering wide ranges of each (Table A1). The columns to be screened and initial salt gradient conditions were selected based on prior experience with mAb separations. The buffer pH range was 6 to 8, which is roughly 1 to 3 pH units below the theoretical pI of the NIST IgG1 molecule (pI of 8.9 calculated from the primary sequence). Mobile phase B (MPB) did not include the buffering component in the initial screening study to decrease the complexity of the DOE design. The buffering component was added to MPB in subsequent method refinement studies after the buffer pH was selected. The other less critical OPs (flow rate, protein load, detection wavelength, and

column temperature) were held fixed for the screening study and were evaluated at a later time. For this initial screening study, a 31-run DOE sequence was generated varying all three critical OPs.

**Table A1. Critical Operating Parameter Ranges Studied in Design of Experiments General Screen**

<i>Critical Operating Parameter</i>	<i>Range/Type Tested</i>
Column Type	ProPac® WCX-10 BioPro SP-F Bio MAb NP5 Antibodix WCX-NP5
Buffer pH	Mobile phase A: 20 mM phosphate buffer: pH 6.0, 6.5, 7.0, 7.5, and 8.0
Mobile Phase B (0.5 M NaCl) Gradient	Fixed start at 0% B; gradient time fixed at 30 minutes. Varied ending % mobile phase B: 40%, 45%, 50%, 55%, 60%

Flow rate 0.6 mL/min; protein load 50 µg; detection at 215 nm; column temperature 25 °C.

A sample output from the initial screening study (Figure A1) shows the cation exchange (CEX)-HPLC chromatograms from the four different columns, all tested with the same OP settings. For this general screen, there was a clear differentiation among the four columns tested by visual assessment: the BioPro SP-F had superior separation, with some resolution of a pre peak and well-resolved post peaks. This was the case across all of the OP settings studied. Therefore, the BioPro SP-F column was selected for further development. Similarly, the pH range was narrowed from the wide range covering 6.0 to 8.0 to 6.5 to 7.0 for subsequent method development studies based on visual assessment (data not shown).

After the general screen, three more DOE studies were executed for method refinement. The DOE study information and outcome from all four studies are shown in Table A2. For subsequent method refinement studies (Studies 2 to 4 in Table A2), where the chromatographic resolution differences were more subtle, the DOE software's data analysis capabilities were used to objectively quantitate the OPs' effects on peak separation and resolution. The USP resolution factor, which is typically used to quantitate peak separation for small molecules, does not work well for CEX-HPLC chromatography of mAbs because minor variants often show up as shoulders to the main peak. Therefore, other chromatographic peak properties had to be explored to judge peak separation objectively and quantitatively. For this development, the key performance parameters selected from the wide array of chromatographic properties are the peak-to-valley ratios (p/v) for the acidic and basic peaks. It should be noted that the optimal conditions for acidic peak resolution and basic peak resolution may be different; therefore, when the performance parameters for both peaks are given equal weighting, the final conditions are a compromise between the optimal conditions for either acidic or basic peaks. The p/v values used are illustrated in Figure A2. The goal

of these DOE studies was to maximize these p/v values and thus resolution of both acidic and basic peaks for the ion exchange separation.

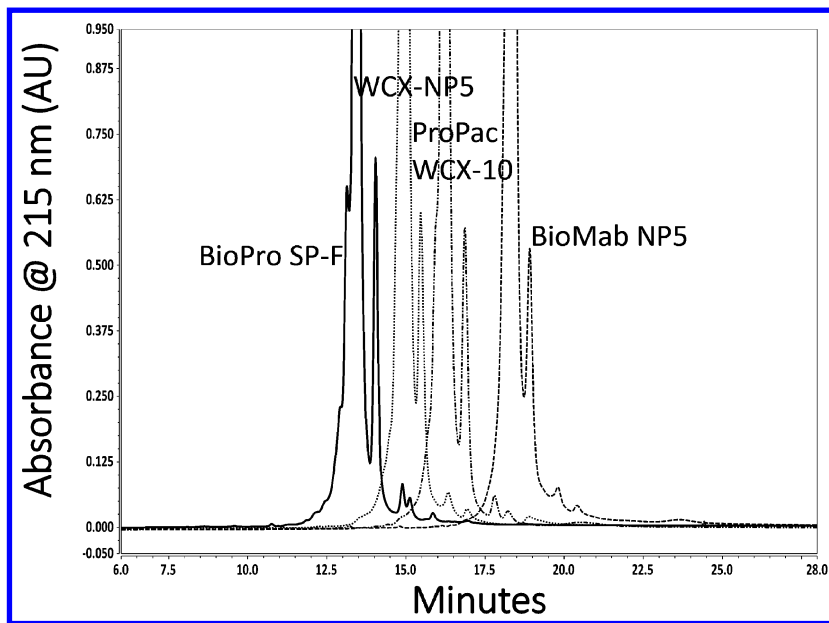


Figure A1. Cation exchange (CEX)-HPLC chromatograms of the NISTmAb from the general screening design of experiment.

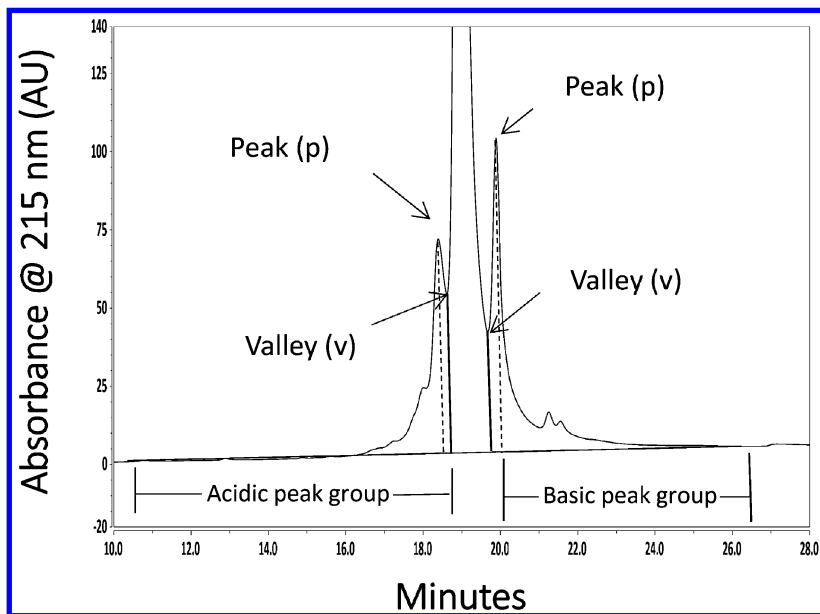


Figure A2. Peak-to-valley (p/v) ratios used to judge peak separation for cation exchange (CEX)-HPLC chromatograms of the NISTmAb.

**Table A2. Design of Experiments Test Sequences Performed for Cation Exchange HPLC Method Development**

<i>Study #</i>	<i>Type of Study</i>	<i>Operating Parameters Studied</i>	<i>Performance Parameters Used</i>	<i>Outcome</i>
1	General Screening	Column type Buffer pH Buffer gradient	Acidic and basic peak resolution (visually assessed)	Column selected pH range narrowed
2	Method Refinement	Buffer pH Column temp. Gradient end (% MPB)	Acidic peak end p/v Basic peak start p/v (DOE software analysis)	pH selected Column temp. selected Gradient end tested
3	Method Refinement	Gradient start (% MPB) Protein load		Gradient start percentage selected Load selected
4	Method Refinement Verification	Test best OP settings recommended by software	Maximize main peak height (DOE software analysis)	Best answer OP settings verified

MPB = mobile phase B, p/v = peak-to-valley ratio, DOE = design of experiment, OP = operation parameter.

With these goals set, the software uses the experimental results imported from the DOE and runs Monte-Carlo simulations to generate a mathematical model that best fits the experiment results. Using this mathematical model, it can predict the best answer (OP settings) to achieve the desired separation. The numerical calculated best answer output from DOE study #2 is shown below in Table A3.

**Table A3. Best Answer Output from Cation Exchange-HPLC Method Development Design of Experiments Study #2**

<i>Operating Parameter</i>	<i>Setting</i>
pH	6.7
Column temperature	35 °C
Gradient end selected	20% MPB

MPB = mobile phase B.

In subsequent DOE studies, the optimal mobile phase pH, column temperature, gradient end, gradient start, and protein load were then determined. Before starting Study 3, the best answer OP settings from Study 2 were slightly modified by increasing the initial percentage MPB from 0% to 3% while keeping the gradient constant in order to reduce the “dead time” before the peaks elute.



This work illustrates how a CEX-HPLC method for an IgG1 molecule can be developed efficiently with a DOE approach. This method was developed using just four studies: one 31-run DOE general screening sequence that explored wide ranges of the critical OPs (column, buffer pH, and buffer gradient), followed by two shorter method-refinement test sequences and a final verification study using the final OPs.

The final refined chromatogram obtained after all four studies is shown in Figure A3. The final conditions are listed in the Materials and Methods section.

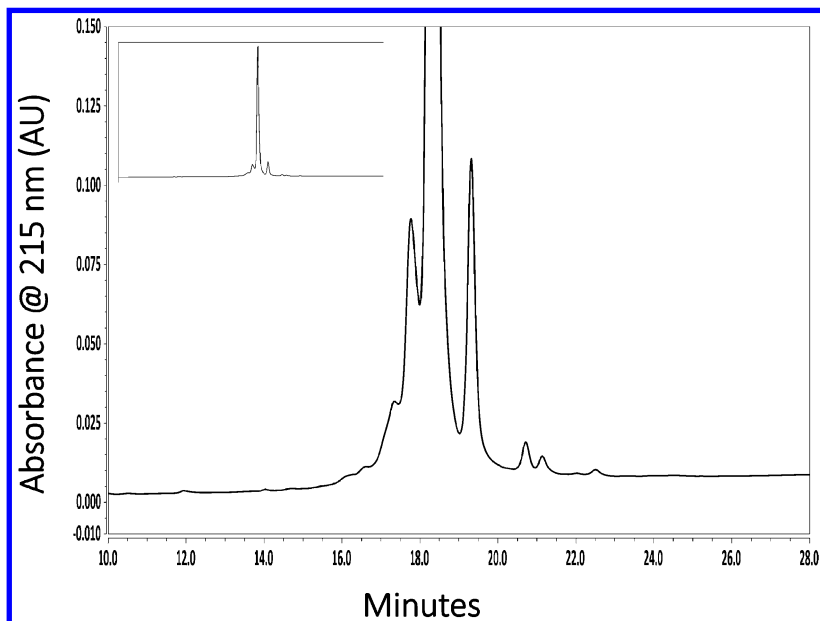


Figure A3. Capillary electrophoresis (CE)-HPLC chromatogram of NISTmAb verification study using best answer operating parameter settings selected by design of experiment.

## References

1. Staub, A.; Guillarme, D.; Schappler, J.; Veuthey, J.; Rudaz, S. *J. Pharm. Biomed. Anal.* **2010**, *55*, 810–822.
2. Weber, K.; Osborn, M. *J. Bio. Chem.* **1969**, *244*, 4406–4412.
3. Reynolds, J. A.; Tanford, C. *J. Bio. Chem.* **1970**, *245*, 5161–5165.
4. Office of Biotechnology Products, U.S. Food and Drug Administration; Docket number FDA-2008-N-03551; 2008.
5. *ICH Q8(R2). Pharmaceutical Development*; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2009.
6. Michels, D. A.; Parker, M.; Salas-Solano, O. *Electrophoresis* **2012**, *33*, 1–12.

7. Valliere-Douglass, J.; Wallace, A.; Baland, A. *J. Chromatogr. A* **2008**, *1214*, 81–89.
8. He, Y.; Isele, C.; Hou, W.; Ruesch, M. *J. Sep. Sci.* **2011**, *34*, 548–555.
9. Mack, S.; Cruzado-Park, I.; Ratnayake, C.; Vigh, G. *Electrophoresis* **2009**, *30*, 4049–4058.
10. Salas-Solano, O.; Babu, K.; Park, S. S.; Zhang, X.; Zhang, L.; Susic, Z.; Boumajny, B.; Zeng, M.; Cheng, K. C.; Reed-Bogan, A.; Cummins-Bitz, S.; Michels, D. A.; Parker, M.; Bonasia, P.; Hong, M.; Cook, S.; Ruesch, M.; Lamb, D.; Bolyan, D.; Kiessig, S.; Allender, D.; Nunnally, B. *Chromatographia* **2011**, *73*, 1137–1144.
11. Li, N.; Kessler, K.; Bass, L.; Zeng, D. *J. Pharm. Biomed. Anal.* **2007**, *43*, 963–972.
12. Susic, Z.; Houde, D.; Blum, A.; Carlage, T.; Lyubarskaya, Y. *Electrophoresis* **2008**, *29*, 4368–4376.
13. Rosenberg, A. S. *AAPS J.* **2006**, *J8*, E501–E508.
14. Hong, P.; Koza, S.; Bouvier, E. S. P. *J. Liq. Chromatogr. Relat. Technol.* **2012**, *35*, 2923–2950.
15. Salinas, B. A.; Sathish, A. U.; Shah, A. U.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2010**, *99*, 2962–2974.
16. Yu, C.; Mun, S.; Wang, N. L. *J. Chromatogr. A* **2006**, *1132*, 99–108.
17. Hjerten, S.; Zhu, M. *J. Chromatogr.* **1985**, *346*, 265–270.
18. Takagi, T. *J. Chromatogr. A* **1990**, *506*, 409–416.
19. Wen, J.; Arakawa, T.; Philo, J. S. *Anal. Biochem.* **1996**, *240*, 155–166.
20. Maity, H.; Lai, Y.; Srivastava, A.; Goldstein, J. *Curr. Pharm. Biotechnol.* **2012**, *13*, 2078–2101.
21. Kendrick, B. S.; Kerwin, B. A.; Chang, B. S.; Philo, J. S. *Anal. Biochem.* **2001**, *299*, 136–146.
22. Ahrer, K.; Buchacher, A.; Iberer, G.; Josic, D.; Jungbauer, A. *J. Chromatogr. A* **2003**, *1009*, 89–96.
23. Hartmann, W. K.; Sapharishi, N.; Yang, X. Y.; Mitra, G.; Soman, G. *Anal. Biochem.* **2004**, *325*, 227–39.
24. Liu, H.; Gaza-Bulseco, G.; Chumsae, C.; Newby-Kew, A. *Biotechnol. Lett.* **2007**, *11*, 1611–22.
25. Cohen, L.; Chait, B. T. *Anal. Biochem.* **1997**, *247*, 257–267.
26. Retamal, C. A.; Thiebaut, P.; Alves, E. W. *Anal. Biochem.* **1999**, *268*, 15–20.
27. Hunt, G.; Nashabeh, W. *Anal. Chem.* **1999**, *71*, 2390–2397.
28. Ma, S.; Nashabeh, W. *Chromatographia* **2001**, *53*, S75–S89.
29. Salas-Solano, O.; Tomlinson, B.; Du, S.; Parker, M.; Strahan, A.; Ma, S. *Anal. Chem.* **2006**, *78*, 6583–6594.
30. Michels, D. A.; Brady, L. J.; Guo, A.; Baland, A. *Anal. Chem.* **2007**, *79*, 5963–5971.
31. 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.
32. Kaschak, T.; Boyd, D.; Yan, B. *Anal. Biochem.* **2011**, *417*, 256–263.
33. Shapiro, R.; McManus, M. J.; Zalut, C.; Bunn, H. F. *J. Biol. Chem.* **1980**, *255*, 3120–3127.

34. Gadgil, H. S.; Bondarenko, P. V.; Treuheit, M. J.; Ren, D. *Anal. Chem.* **2007**, *79*, 5991–5999.
35. Legmann, R.; Schreyer, H. B.; Combs, R. G.; McCormick, E. L.; Russo, A. P.; Rodgers, S. T. *Biotechnol. Bioeng.* **2009**, *104*, 1107–1120.
36. Bousse, L.; Mouradian, S.; Minalla, A.; Yee, H.; Williams, K.; Dubrow, R. *Anal. Chem.* **2001**, *73*, 1207–1212.
37. Manz, A.; Harrison, D. J.; Verpoorte, E.; Widmer, H. M. *Adv. Chromatogr.* **1993**, *33*, 1–60.
38. Rehder, D. S.; Chelius, D.; McAuley, A.; Dillon, T. M.; Xiao, G.; Crouse-Zeineddini, J.; Vardanyan, L.; Perico, N.; Mukku, V.; Brems, D. N.; Matsumaru, M.; Bondarenko, P. V. *Biochemistry* **2008**, *47*, 2518–2530.
39. Rehder, D. S.; Dillon, T. M.; Pipes, G. D.; Bondarenko, P. V. *J. Chromatogr. A* **2006**, *1102*, 164–175.
40. Dillon, T. M.; Bondarenko, P. V.; Ricci, M. S. *J. Chromatogr. A* **2004**, *1053*, 299–305.
41. Ambrogelly, A.; Liu, Y. H.; Li, H.; Mengisen, S.; Yao, B.; Wei, X.; Cannon-Carlson, S. *mAbs* **2012**, *4*, 701–709.
42. Zhang, Tl; Zhang, J.; Hewitt, D.; Tran, B.; Gao, X.; Qiu, Z. J.; Tejada, M.; Gazzano-Santoro, H.; Yung-Hsiang, K. *Anal. Chem.* **2012**, *84*, 7112–7123.
43. Pristatsky, P.; Cohen, S. L.; Krantz, D.; Acevedo, J.; Ionescu, R.; Vlasak, J. *Anal. Chem.* **2009**, *81*, 6148–6155.
44. Gu, S.; Wen, D.; Weinreb, P. H.; Sun, Y.; Zhang, L.; Foley, S. F.; Kshirsagar, R.; Evans, D.; Mi, S.; Meier, W.; Pepinsky, R. B. *Anal. Biochem.* **2010**, *400*, 89–98.
45. Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S. *J. Chromatogr. A* **2006**, *1120*, 112–120.
46. Dillon, T. M.; Speed-Ricci, M.; Vezina, C.; Flynn, G. C.; Liu, Y. D.; Rehder, D. S.; Plant, M.; Henkle, B.; Li, Y.; Deechongkit, S.; Varnum, B.; Wypych, J.; Balland, A.; Bondarenko, P. V. *J. Biol. Chem.* **2008**, *283*, 16206–15.
47. Battersby, J. E.; Snedecor, B.; Chen, C.; Champion, K. M.; Riddle, L.; Vanderlaan, M. *J. Chromatogr. A* **2001**, *927*, 61.
48. McCalley, D. V. *J. Chromatogr. A* **2004**, *1038*, 77–84.
49. Trexler-Schmidt, M.; Sargis, S.; Chiu, J.; Sze-Khoo, S.; Mun, M.; Kao, Y. H.; Laird, M. W. *Biotechnol Bioeng.* **2010**, *106*, 452–61.
50. von Pawel-Rammingen, U.; Johansson, B. P.; Bjorck, L.; IdeS, G. *EMBO J.* **2002**, *21*, 1607–1615.
51. DePeralta, G.; Alvarez, M.; Bechtel, C.; Dong, K.; McDonald, R.; Ling, V. *mAbs* **2013**, *5*, 86–101.
52. Tous, G. I.; Wei, Z.; Feng, J.; Bilbulian, S.; Bowen, S.; Smith, J.; McGeehan, P.; Casas-Finet, J.; Schenerman, M. A. *Anal. Chem.* **2005**, *77*, 2675–2682.
53. Harris, R. J. *J. Chromatogr. A* **2006**, *1133*, 332–339.
54. Johnson, K. A.; Paisley-Flango, K.; Tangarone, B. S.; Porter, T. J.; Rouse, J. C. *Anal. Biochem.* **2007**, *360*, 75–83.
55. Lyubarskaya, Y.; Houde, Dl; Woodard, J.; Murphy, D.; Mhatre, R. *Anal. Biochem* **2006**, *348*, 24–39.

56. Zhang, W.; Czupryn, M. J. *J. Pharm. Biomed. Anal.* **2003**, *30* (5), 1479–1490.
57. Kim, J.; Jones, L.; Taylor, L.; Kannan, G.; Jackson, F.; Lau, H.; Latypov, R. F.; Bailey, B. *J. Chromatogr. B.* **2010**, *878* (22), 1973–1981.
58. Luo, J.; Zhang, J.; Ren, D.; Tsai, W.; Li, F.; Amanullah, A.; Hudson, T. *Biotechnol. Bioeng.* **2012**, *109*, 2306–2315.
59. Dick, L. W., Jr.; Qiu, D.; Mahon, D.; Adamo, M.; Cheng, K. C. *Biotechnol. Bioeng.* **2008**, *100*, 1132–1143.
60. Spanik, I.; Lim, P.; Vigh, G. *J. Chromatogr. A* **2002**, *960*, 241–246.
61. Reid, G. L.; Morgado, J.; Barnett, K.; Harrington, B.; Wang, J.; Harwood, J.; Fortin, D. *Am. Pharm. Rev.* **2013**.
62. Sanford, L.; Shelver, G. *Fusion AE Method Development*; Application Note 002-09; S-Matrix Corp., 2009.

## Chapter 6

# Biophysical Techniques for Characterizing the Higher Order Structure and Interactions of Monoclonal Antibodies

Yatin Gokarn,<sup>\*,1</sup> Sanjeev Agarwal,<sup>1</sup> Kelly Arthur,<sup>2</sup>  
Alexander Bepperling,<sup>3</sup> Eric S. Day,<sup>4</sup> Dana Filoti,<sup>5</sup> Daniel G. Greene,<sup>6</sup>  
David Hayes,<sup>7</sup> Rachel Kroe-Barrett,<sup>7</sup> Thomas Laue,<sup>5</sup> Jasper Lin,<sup>4</sup>  
Brian McGarry,<sup>4</sup> Vladimir Razinkov,<sup>2</sup> Sanjaya Singh,<sup>7</sup>  
Rosalynn Taing,<sup>4</sup> Sathyadevi Venkataramani,<sup>7</sup> William Weiss III,<sup>8</sup>  
Danlin Yang,<sup>7</sup> and Isidro E. Zarraga<sup>4</sup>

<sup>1</sup>Institute of Chemical Technology, Mumbai, India

<sup>2</sup>Amgen Inc., Thousand Oaks, California 91320, United States

<sup>3</sup>Hexal AG, Oberhaching, Germany

<sup>4</sup>Genentech, South San Francisco, California 94080, United States

<sup>5</sup>University of New Hampshire, Durham, New Hampshire 03824,  
United States

<sup>6</sup>University of Delaware, Newark, Delaware 19716, United States

<sup>7</sup>Boehringer Ingelheim, Ridgebury, Connecticut 06877, United States

<sup>8</sup>Eli Lilly, Indianapolis, Indiana 46221, United States

\*E-mail: [yatin\\_gokarn@yahoo.com](mailto:yatin_gokarn@yahoo.com)

One of the major challenges that the biopharmaceutical industry currently faces is maintaining the structural integrity and stability of protein therapeutics through production and shelf life. Changes introduced in antibody production processes or formulations may have major impact on stability, efficacy, and safety. Hence, a robust set of biophysical techniques to characterize and assess the “higher order structure” (HOS) of monoclonal antibodies (mAbs) is an integral part of the drug development process. In a unique collaboration spanning academic/government institutions and the biotechnology industry, we used a set of sophisticated orthogonal biophysical techniques for characterizing the HOS of an IgG1 mAb. Each

section of this chapter focuses on a particular method to probe different molecular properties of the mAb such as solution mass, size, polydispersity, solution charge, thermal stability, aggregation, and secondary and tertiary structural changes in different solutions and at different mAb concentrations. Our collaboration has resulted in the collection of biophysical data rich in the solution structural information of the NISTmAb that we hope will help researchers worldwide as a guide for the biophysical characterization of mAbs and proteins in general.

## Introduction

The pharmaceutical landscape is undergoing a transformation in terms of the medicines available for treatment of disease. Traditional, small molecule therapies are steadily being augmented or replaced with monoclonal antibody- (mAb) and protein-based drugs for treating cancer, cardiovascular, infectious, and neurodegenerative diseases. With more than 30 approved products, mAb-based therapies have emerged as the dominant therapeutic modality over the last decade (1). The increasing therapeutic success of these large glycosylated proteins stems from their inherent drug-like properties and the relative ease and scalability of manufacturing (2–4). mAbs present several desirable pharmacological properties over their small molecule counterparts; mAbs display exquisite target specificity and generally have better safety profiles and longer in vivo half-lives. Advances in recombinant DNA technology, cell culture, and purification processes have enabled mAbs to be on the path to industrialized production (2).

Early biotherapeutic drugs such as growth factors and cytokines necessitated specialized, molecule-specific production processes with commensurate investments in process development. However, mAbs are increasingly being produced using convergent manufacturing processes starting with fed-batch cell culture followed by capture with Protein A chromatography and further polishing and concentration steps of ion-exchange/hydrophobic interaction chromatography and ultra/diafiltration, respectively (2). Platform-based mAb production and purification processes have enabled the rapid evaluation of new scientific disease mechanisms and lowered the barrier for entry to smaller start-ups with novel hypotheses. Presently, there are hundreds of mAbs and mAb-based molecules in worldwide development (1). mAb-based therapies continue to evolve rapidly as is evident from the recent successes of antibody–drug conjugates (ADCs) for oncology indications (5). Simultaneous dual targeting with bispecific antibodies for synergistic biological action (6) and/or targeted delivery (7) is also a significant, promising advance in mAb-based medicine.

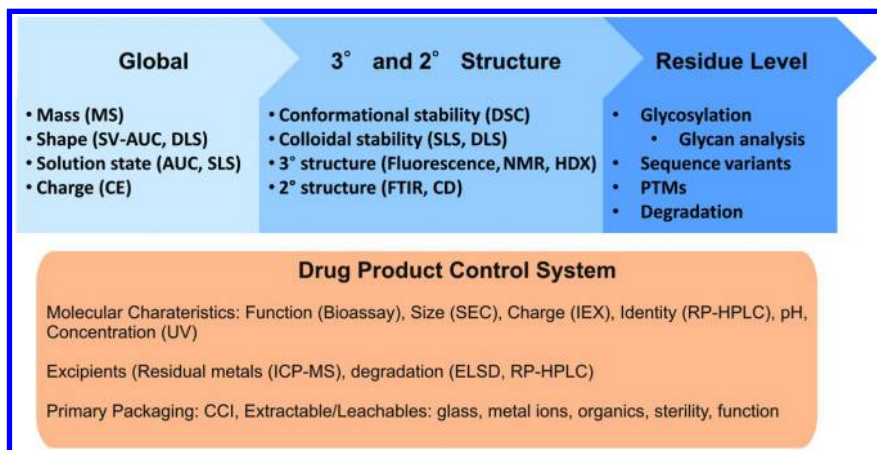
Unlike small molecules, which have well-defined molecular structures and relatively fewer degradation routes, mAb- and protein-based therapeutics present far greater structural complexity and stability issues. The exquisite specificity and safety afforded by protein drugs is hinged upon maintenance of their “higher order structure” (HOS), including secondary, tertiary, and quaternary (when appropriate) structure. In addition to chemical degradation, which can lead

to potency loss in small molecules as well as proteins, the latter are subject to aggregation, a degradation process that can be triggered by subtle changes in the HOS. Protein aggregates are well recognized to induce undesirable immunogenic/antigenic responses in patients (8). Consequently, aggregation needs to be minimized during protein production and through the end of product shelf life. This can be achieved through careful process and formulation design. It should be noted that perturbation of HOS without any accompanying aggregation may be sufficient to cause immunogenicity/antigenicity and/or potency loss.

mAbs and proteins are produced using mammalian cell culture (mAbs, glycosylated proteins) and bacterial or yeast fermentation (Fabs, smaller proteins) and, therefore, are subject to the inherent variability of these living systems. Subtle changes in production processes can lead to changes in HOS. Exposure to harsh process streams, as well as various interfacial and thermal stresses during storage, also can impact HOS. Given its profound effects on product efficacy and safety, it is essential—and a regulatory agency expectation—that the HOS of protein drug candidates be carefully analyzed at each stage of development and through the entire product life cycle. Thorough characterization of the HOS of biosimilars will prove to be essential to minimizing the residual uncertainty in structural differences that may result from using different cell lines, process, and purification conditions.

HOS characterization is an integral component of analytical product characterization. It encompasses a set of orthogonal biophysical/biochemical methods chosen to probe different aspects of solution structure and the interactions of a given molecule in a “solution-space” of interest (Figure 1). For example, the global complementary methods of analytical ultracentrifugation (AUC) and light scattering remain unparalleled in providing first-principles insight about solution mass, state, size, and polydispersity, as well as the thermodynamics of interactions. Electrophoretic methods such as membrane-confined analytical electrophoresis (MCAE) provide fundamental information about the charge of a protein in solution, which can be dramatically different from the calculated sequence-based charge. Differential scanning calorimetry (DSC) and other thermal methods such as differential scanning fluorimetry (DSF) shed light on effects of solution conditions on the thermal unfolding of domains. These methods not only inform us about the molecular conformational stability but also about the interactions of the molecule with its co-solutes in solution. Fourier transform infrared (FTIR) spectroscopy is one of the few spectroscopic techniques that facilitates the study of protein secondary structure, particularly beta-sheet structure, at high concentrations (~100 mg/mL). Also, FTIR can be employed to study protein structure in the solid state in lyophilized dosage forms. Circular dichroism (CD) is a classical technique that can be used to study tertiary and secondary structure in near- and far-UV modes, respectively. Although the information provided by each of the above biophysical methods is focused, these methods when applied in concert enable the construction of a HOS map rich in solution structural information. The HOS map can be employed to ensure consistency in structure with scale-up along the development cycle, during product transfers to new sites, and when production processes are modernized for mature products. Additionally, information from HOS characterization may

be invaluable for evaluating biosimilars, which hold the promise of affordable medicine to large patient populations in developing, as well as developed, countries.



*Figure 1. Characterization methodology informing various product attributes. In this figure, the following abbreviations are used: AUC (analytical ultracentrifugation), CCI (container closure integrity), CD (circular dichroism), CE (capillary electrophoresis), DLS (dynamic light scattering), DSC (differential scanning calorimetry), ELSD (evaporating light scattering detector), FTIR (Fourier transform infrared), HDX (hydrogen-deuterium exchange), ICP-MS (inductively coupled plasma-mass spectrometry), IEX (ion exchange chromatography), MS (mass spectrometry), NMR (nuclear magnetic resonance spectroscopy), PTMs (post-translational modifications), RP-HPLC (reversed-phase-high-performance liquid chromatography), SEC (size-exclusion chromatography), SLS (static light scattering), and SV-AUC (sedimentation velocity-analytical ultracentrifugation). (see color insert)*

In a unique collaboration involving researchers from nine different bio/pharmaceutical company, academic, and government research labs, this chapter aims to illustrate the utility of biophysical methods for characterizing the HOS of an IgG1 mAb. The NISTmAb is under development at the National Institute of Standards and Technology (Gaithersburg, MD) and will be available as a Reference Material in the future. We have focused on experimental aspects of these sophisticated techniques, which may be particularly useful to researchers who embark on the “art” and science of biophysical characterization. We stress that the experimental conditions and the techniques presented in this chapter should not be taken as prescriptive protocols but merely as guides or starting points. The chapter simply aims to help researchers through demonstrative experiments in the HOS characterization of mAbs and other proteins of interest.

The chapter begins with the analysis of global properties of mass, size, solution state, size distribution, and charge followed by a section titled “Signatures of Secondary and Tertiary Structure.” Finally, the chapter concludes with a synopsis of the higher order structural information of the NISTmAb.



## Global Methods: Solution Mass, Size, State, Size-Distribution, and Charge

We employed the complementary methods of AUC, or sedimentation, and light scattering to determine the solution molecular weight, state, size, and size distribution of the NISTmAb. MCAE was used to determine the solution charge of the NISTmAb. Both AUC and light scattering are first-principles-based solution techniques uniquely placed for the solution characterization of macromolecules. Similarly, MCAE is the only first-principles-based method available for the measurement of solution charge. We stress that the solution charge of proteins, and particularly antibodies, can be dramatically different than the assigned sequence-based charge, and hence its measurement is critical to understanding the solution interactions of macromolecules.

### Analytical Ultracentrifugation

An analytical ultracentrifuge used for biophysical measurements is equipped with an optical detection system capable of making absorbance and/or refractive index measurements in real time. As such, the production of a concentration gradient as a result of centrifugal force can be measured. AUC experiments are performed in two modes (i) sedimentation equilibrium, and (ii) sedimentation velocity (9). Sedimentation equilibrium, as the name suggests, is an equilibrium method used to measure solution mass, state, and stoichiometry, as well as the energetics of rapidly associating and reversible macromolecular interactions. Sedimentation velocity, on the other hand, is a transport method that primarily provides hydrodynamic information about molecules based on rates of sedimentation under a centrifugal field. However, recent advances in sedimentation velocity data analysis has enabled the analysis of interactions (10), and the determination of size distributions in protein–mAb mixtures (11).

### Sedimentation Equilibrium-Analytical Ultracentrifugation (SE-AUC)

SE-AUC experiments are conducted at lower rotor speeds such that the protein moves toward the outside of the rotor (centrifugal flux) but not at a speed sufficient to completely sediment the protein. At such speed, there is a diffusive force opposing the concentration gradient imparted by the centrifugal solute flux as described by the Lamm equation:

$$\frac{\partial x}{\partial t} = \frac{1}{r} \times \frac{\partial}{\partial r} \left[ rD \frac{\partial x}{\partial r} - s\omega^2 r^2 x \right] \quad (1)$$

where  $x$  is the concentration distribution of macromolecular species as a function of time ( $t$ ) and radial position ( $r$ ) in the centrifugal field  $\omega^2 r$ .  $D$  and  $s$  are the diffusion coefficient and sedimentation coefficient, respectively. After an appropriate experimental time period, the solute concentration distributions become time invariant, and the system approaches thermodynamic equilibrium. At this point, there is net zero transport of the macromolecule across the

radial concentration distribution, which can be described by a single or a set of exponential equations (Equation 3). This equilibrium concentration gradient is then measured as a function of radial distance using spectroscopic detection. Solute weight-average molecular weight, thermodynamic non-ideality, stoichiometry, and the energetics of reversible self-association can be obtained by applying appropriate models to fit the experimental data.

A typical SE-AUC experiment is conducted with the widest possible solute concentration range and multiple rotor speeds in a step-wise manner. Equilibrium is established sequentially, always from the lowest to the highest speed, and at each, equilibrium concentration profiles are scanned using a suitable optical system. Robust data analysis is possible through global fitting of multiconcentration and multispeed equilibrium data. The appropriateness of a chosen analysis model is determined using parsimony: the simplest model which best describes the data based on analysis of root-mean-square deviation (RMSD) and the randomness of residuals.

Detailed quantitative discussion on SE-AUC can be found elsewhere (12); briefly, however, for a single, ideal, non-interacting solute, molecular weight,  $M$ , is given by:

$$M = \frac{2RT}{(1-\bar{v}\rho)\omega^2} \times \frac{d(\ln c)}{dr^2} \quad (2)$$

and the equilibrium radial concentration gradient is described by Equation 3:

$$c(r) = c_0 \cdot \exp \left[ \frac{M(1-\bar{v}\rho)\omega^2}{RT} \left( \frac{r^2 - r_0^2}{2} \right) \right] = c_0 \cdot \exp \left[ \sigma \left( \frac{r^2 - r_0^2}{2} \right) \right] \quad (3)$$

where  $M$  is the solute molecular weight (g/mol),  $\omega$  is the angular velocity of the rotor (radians/sec),  $\bar{v}$  is partial specific volume of solute (mL/g),  $\rho$  is solvent density (g/mL),  $R$  is the gas constant,  $T$  is absolute temperature,  $c$  is the concentration of solute (g/L) at a radial distance  $r$  from the axis of rotation, and  $c_0$  is the concentration at any arbitrary reference distance  $r_0$ . The term  $M(1 - \bar{v}\rho)\omega^2/RT$  is often referred to as the reduced molecular weight,  $\sigma$ , where  $M(1 - \bar{v}\rho)$  is the buoyant molecular weight of the solute, denoted by  $M_b$ .

If multiple species are present in a sample, the overall equilibrium concentration profile represents the sum of different components of a solute mixture or of a reaction mixture in the case of interacting systems. Deviations from an exponential expected for a single, ideal species indicates the presence of multiple self-interacting/non-interacting species or of thermodynamic non-ideality. High macromolecular charge and excluded volume effects give rise to thermodynamic non-ideality, which manifests as a reduction in the apparent molecular weight ( $M_{app}$ ) measured as a function of solute concentration. In the presence of non-ideality, the true molecular weight is determined using Equation 4:

$$M_{app} = \frac{M}{(1 + c(d(\ln y)/dc))} \approx \frac{M}{(1 + BMc)} = \frac{2RT}{(1 - \bar{v}\rho)\omega^2} \times \frac{d(\ln c)}{dr^2} \quad (4)$$

where  $M$  is the true molecular weight,  $c$  is the total concentration of solute, and  $y$  is the activity coefficient. The activity coefficient,  $y$ , can be expressed as a polynomial in  $c$  as:

$$\ln y = BMc + CMc^2 + \dots \quad (5)$$

By neglecting higher order terms of this polynomial over a limited concentration range, Equation 4 can be simplified in terms of the osmotic second virial coefficient ( $B$ ), a measure of thermodynamic non-ideality.

Here, we conducted sedimentation equilibrium experiments with the NISTmAb at 20 °C in a 25 mM histidine (His)/His-HCl, pH 6.0 buffer using five loading concentrations (0.3, 1, 2, 4, and 8 mg/mL) and at three rotor speeds (12K, 14K, and 16K revolutions per minute [RPM]). Experiments were conducted employing the ProteomeLab XL-I analytical ultracentrifuge (Beckman-Coulter Instruments, Indianapolis, IN) interference optical system given the high loading concentrations of the mAb. A partial specific volume value of 0.725, calculated based on the amino acid composition and carbohydrate content of the NISTmAb using the software program SEDNTERP (kindly made available at <http://sednterp.unh.edu/>), was used. The absorbance optical system is limited to gradients where the highest absorbance is less than 0.8. The sedimentation equilibrium data were analyzed using the software program HeteroAnalysis (kindly made available at <http://www.biotech.uconn.edu/auf/>).

Initially, we performed a global fit to the equilibrium data using the Single, Ideal Species model but observed an RMSD of 0.34 fringes (data not shown), which is much higher than the random optical noise. Further analysis was performed by grouping the data by loading concentration and fitting each of the five groups of data at three different speeds to a model of a single, ideal species. The apparent molecular weight from these five curve fits decreased with the NISTmAb loading concentration (Figure 2A), which demonstrates the presence of non-ideality effects. To account for and estimate the non-ideality in terms of the second virial coefficient, we analyzed the equilibrium data using the single, non-ideal species model. Global analysis of these data yielded a weight average molecular weight of 149.9 kDa and an osmotic second virial coefficient of  $1.67 \times 10^{-4}$  mol mL/g<sup>2</sup>, which is in good agreement with the sequence molecular weight of 148.0 kDa (G0F/G0F glycoform, Primary Structure chapter/Volume 2, Chapter 1) for the NISTmAb. The approximate relative abundance of the five most abundant glycoforms (as measured via intact mass spectrometry in the Primary Structure chapter/Volume 2, Chapter 1) also was used to calculate a population average of the NISTmAb glycoforms (=148.3 kDa), which is also in good agreement. The RMSD reduced from 0.34 fringes (single, ideal species model) to 0.05 fringes upon application of the single, non-ideal species model. Representative fits to the 0.3 and 8.0 mg/mL data are presented in Figure 2B and 2C, respectively. Residual plots for the 0.3 mg/mL sample are relatively random, but some systematic deviation is observed for the 8 mg/mL data. The

U-shaped residuals with sharp increases at higher radial positions is indicative of aggregation (13). However, the extent of aggregation appears to be minimal given the good agreement of  $M_w$  and  $B_2$  with the orthogonal static light scattering (SLS) measurements (Table 1). Measurements with SLS will be discussed in the following section of the chapter.

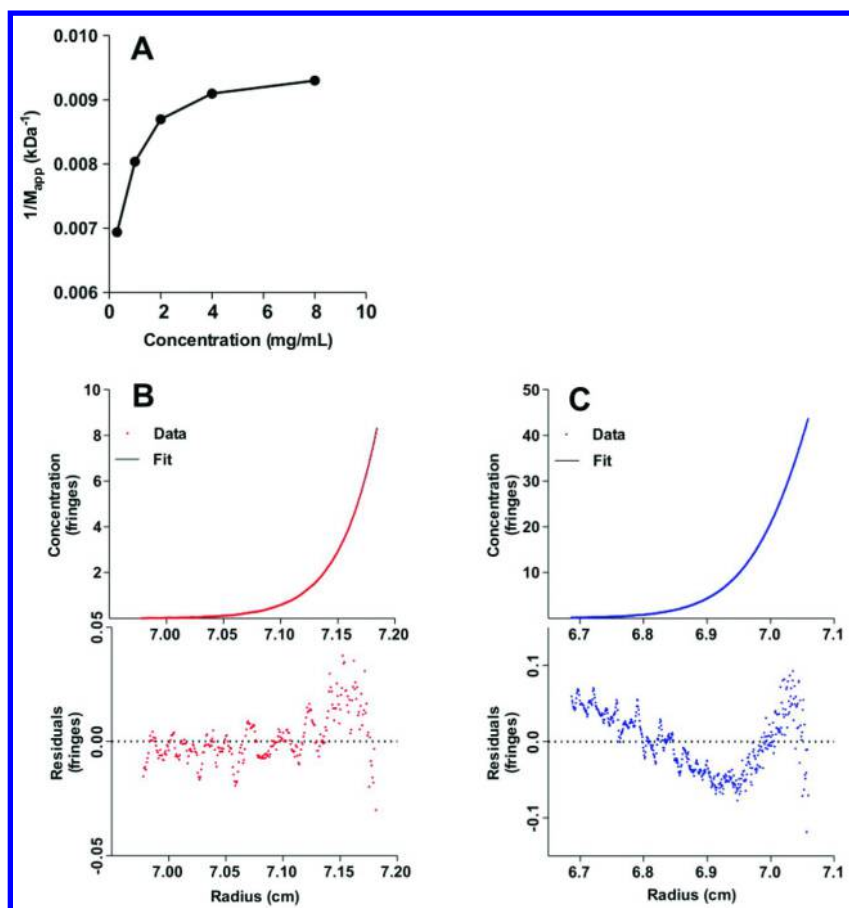


Figure 2. Sedimentation equilibrium results for the NISTmAb demonstrating (A) the change in apparent molecular weight ( $M_{app}$ ) as a function of loading concentration using the single, ideal model fits and representative single, non-ideal model fits to the (B) 0.3 mg/mL and (C) 8.0 mg/mL data.

**Table 1. Comparison of Molecular Weights by Three Orthogonal Techniques**

<i>Technique</i>	<i>Molecular Weight (kDa)</i>	<i>Second Virial Coefficient, B<sub>2</sub> (mol mL/g<sup>2</sup>)</i>
Static light scattering	148 (solution, weight average)	$1.64 \times 10^{-4}$
Sedimentation equilibrium	150 (solution, weight average)	$1.67 \times 10^{-4}$
Mass spectrometry	148.3 (sequence)	NA

### Sedimentation Velocity-Analytical Ultracentrifugation (SV-AUC)

SV-AUC is a transport method useful for the analysis of macromolecular systems on the basis of shape and mass. In a typical SV-AUC experiment, solute particles are centrifuged at higher rotor speeds (compared with SE-AUC). The resulting rapid sedimentation of solutes leads to the formation of a sedimentation boundary representing the sedimentation front. The rate of radial movement of the sedimentation boundary, or the sedimentation rate, can be measured and analyzed to determine molecular shape, size distributions, and even protein–protein interactions. Recent advances in the analysis of sedimentation data have led to resurgence in the use of SV-AUC, particularly in size distribution analysis of therapeutic proteins and in the analysis of interacting systems.

A sedimenting molecule is subjected to three forces: centrifugal force ( $f_c$ ), buoyant force ( $f_b$ ), and frictional drag force ( $f_d$ ). Within a very short time (less than  $10^{-6}$  sec), these forces come into balance (Equation 6):

$$F_{total} = f_c + f_b + f_d = 0 \quad (6)$$

Substituting for the centrifugal, buoyant, and frictional forces, we get:

$$\frac{M}{N} \omega^2 r - \frac{M \bar{v} \rho}{N} \omega^2 r - f u = 0 \quad (7)$$

or:

$$\frac{M(1 - \bar{v} \rho)}{N f} = \frac{u}{\omega^2 r} \equiv s \quad (8)$$

where,  $\omega^2 r$  is the angular acceleration,  $M$  is the molar mass of solute,  $N$  is Avogadro's number,  $\bar{v}$  is the partial specific volume,  $\rho$  is the solvent density,  $f$  is the solute frictional coefficient,  $\omega$  is the angular velocity of the rotor,  $r$  is the radial distance from meniscus, and  $u$  is the observed radial velocity of the solute particle.

The sedimentation coefficient ( $s$ ) is defined as the ratio of the linear velocity to the angular acceleration of a sedimenting molecule and is a constant related to its molecular properties (Equation 8). The sedimentation coefficient is proportional

to the buoyant molecular weight of the solute and inversely proportional to its frictional coefficient, which in turn is dependent on molecular shape and size. The sedimentation coefficient has units of time and is commonly reported in svedberg units (s), which corresponds to  $10^{-13}$  seconds.

Employing the Stokes-Einstein relationship  $D = RT/Nf$  and substituting in Equation 8 gives rise to the Svedberg equation (Equation 9):

$$\frac{s}{D} = \frac{M(1 - \bar{v}\rho)}{RT} = \frac{M_b}{RT} \quad (9)$$

where  $M_b$  is the buoyant molecular mass,  $R$  is the gas constant, and  $T$  is the absolute temperature. In a sedimentation velocity experiment, diffusion causes spreading of the sedimentation boundary, and the translational diffusion coefficient,  $D$ , of a solute can be determined by analyzing the shape of a sedimentation boundary. In sum, the sedimentation coefficient is determined from the rate of boundary movement, whereas the diffusion coefficient can be determined from boundary shape.

There are several model-independent and dependent methods developed by researchers for the analysis of SV-AUC data, detailed descriptions of which can be found elsewhere (11, 14, 15). Here, we employed the program SEDFIT, particularly the c(s) or the continuous size distribution model within SEDFIT, for the analysis of the NISTmAb SV-AUC data. The c(s) analysis is popular and used widely for the determination of size distributions of therapeutic protein mixtures (11). Briefly, the c(s) analysis enables direct modeling of experimental data for computation of diffusion-deconvoluted sedimentation coefficient distribution plots.

SV-AUC experiments were performed with a Beckman XL-I analytical ultracentrifuge (Beckman Coulter, Brea, CA). On the day of each experiment, samples were diluted in respective buffers (Table 2) to a concentration of 0.5 mg/mL. Samples were loaded into the sample sectors of 1.2 cm charcoal-filled Epon centerpieces in analytical cells with sapphire or quartz windows. The sample sector was loaded with 400  $\mu$ L of diluted sample, and 410  $\mu$ L of matched buffer was used in the reference sector of the cell. Experiments were carried out using either a 4-hole titanium An-60 Ti or an 8-hole An-50 Ti rotor with the calibration counterweight placed in hole 4 or hole 8, respectively. Cell alignment to the center of the rotor was verified using an alignment tool (Spin Analytical, Berwick, ME) to provide precision cell alignment with the cell housing. An initial scan at 3000 RPM was performed to verify cell integrity and to perform a radial calibration for the instrument. Samples were allowed to reach temperature equilibrium at 20 °C before initiation of centrifugation at high speed (40,000–50,000 RPM). Cell misalignment or insufficient temperature equilibration, which may create temperature gradients, can result in falsely elevated levels of higher molecular weight species (HMWS) (16). The data—100 scans of each cell, reporting concentration as a function of radial position—are collected by absorbance scanning at 280 nm with a radial step size of 0.003 cm, continuous scanning, 1 flash per channel. Data were analyzed using SEDFIT

version 14.1, producing  $c(s)$  distributions. The parameters used for curve fitting with SEDFIT are detailed in Table 2.

**Table 2. Parameters Used for Curve Fitting with SEDFIT**

<i>Parameter</i>	<i>Value</i>	<i>Source</i>
Partial specific volume	0.725	Protein sequence and carbohydrate content
Buffer density, 25 mM histidine	1.00000 g/mL	Measured
Buffer density, 25 mM Histidine, 150 mM NaCl	1.00615 g/mL	Measured
Buffer viscosity, 25 mM histidine	0.01013 poise	Measured
Buffer viscosity, 25 mM histidine, 150 mM NaCl	0.01025 poise	Measured
Resolution	100, 200, or 451	Fitting setting optimized to have a grid sedimentation coefficient resolution close to the precision of the measurement of the sedimentation coefficient of the main species
Lower $s$ limit	2 svedberg units	Fitting setting based on the molecular weight of expected species (choice confirmed by curve fit, no species detected at or smaller than low limit)
Higher $s$ limit	20 svedberg units	Fitting setting based on the molecular weight of expected species (choice confirmed by curve fit, no species detected at or greater than high limit)
Confidence ratio	0.68 or 0.95	Corresponds to approximately 1 standard deviation confidence
Base of cell	7.2 cm	Geometry of rotor and cell
Meniscus	Fitted	Fitted
Frictional ratio	Fitted	Fitted

$S$ , sedimentation coefficient.

Three different laboratories conducted experiments with the NISTmAb in triplicate. The  $c(s)$  distributions of the NISTmAb in both solvent conditions revealed the presence of a single species sedimenting at 6.3 s (Table 3). The Stokes radius for the NISTmAb calculated using the sedimentation coefficient and solution molecular weight (from SLS and SE-AUC) was 5.5 nm, consistent

with that expected for a monomeric IgG<sub>1</sub> and in good agreement with the average value from dynamic light scattering (DLS) measurements (5.5 nm) as will be discussed in a subsequent section of this chapter. The *c(s)* distribution plots also indicate that the NISTmAb sample in both buffers is monodisperse with >99% monomer and ~0.5% HMWS. These results are consistent with size exclusion chromatography (SEC) analysis of the NISTmAb (Separation chapter/Volume 2, Chapter 5 and studies herein). We also employed SV-AUC experiments coupled with *c(s)* analysis for orthogonal verification of the SEC method developed for the NISTmAb. We analyzed UV light-degraded NISTmAb samples by SV-AUC, as well as SEC, a discussion of which follows.

### **Size Distribution Analysis by Sedimentation Velocity-Analytical Ultracentrifugation: Comparison with Size Exclusion Chromatography**

SEC is the principal method of analysis for monitoring size-related degradation products (SRDPs) in mAbs and other protein-based drug products. The method is simple and robust; its data analysis is model-independent and, consequently, easily amenable for incorporation into quality control laboratories. However, a SEC method has to be carefully developed and verified with orthogonal techniques to ensure accuracy of the resulting size distributions (17). The solution pH and ionic strength of the SEC mobile phases invariably have to be adjusted to optimize the separation of various SRDPs from the main peak (usually monomer). Although separation may be optimized using such techniques, it can lead to significant errors in the quantitation of SRDPs. The chosen mobile phase may cause dissociation of aggregates or precipitation of SRDPs, which in turn have the consequence of the sample displaying erroneously higher purity. A large dilution of the sample upon injection also can cause dissociation of aggregates. Conversely in some cases, the mobile phase solvent conditions can trigger irreversible aggregation, causing an artificially high HMWS readout from the sample. Interaction of the mAb with the gel filtration matrix also can lead to incorrect computation of the relative proportions of various size-related species in solution. It is for these reasons, coupled with the potentially undesirable safety effects of aggregates, that regulatory agencies expect that results from a given SEC method be verified using orthogonal techniques such as DLS, field-flow fractionation (FFF), and/or SV-AUC (17).

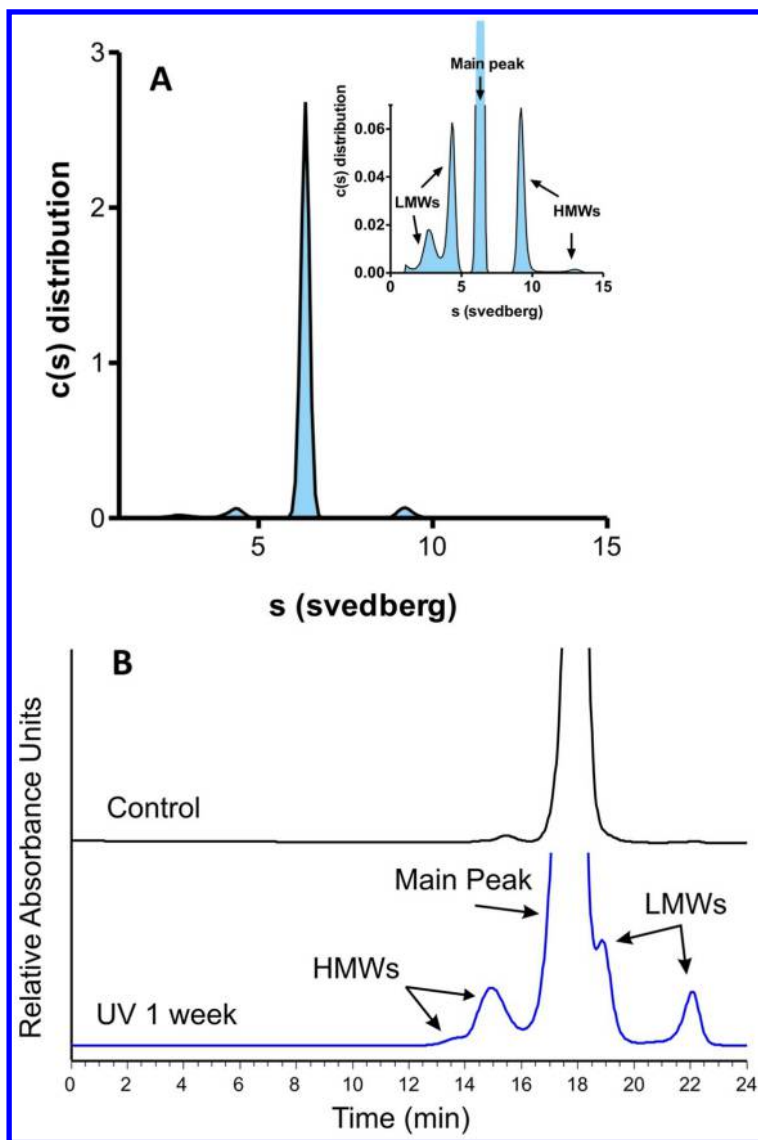
Amongst the orthogonal techniques, SV-AUC has emerged as the dominant method for the quantitation of SRDPs in mAb solutions. In a SV-AUC experiment, a given sample mixture is subjected to sedimentation under a high centrifugal field. The separation of solute components is based on the ratio of their buoyant mass to the frictional coefficient, measured in terms of sedimentation coefficients (*s*). Quantitation of SRDPs in a sedimenting mixture becomes possible upon applying the *c(s)* analysis method assuming solution ideality and when reversible self-association is absent or when its kinetics are slow compared with the time scale of sedimentation. The *c(s)* analysis is a powerful sedimentation velocity data analysis method which enables the construction of high-resolution diffusion-deconvoluted distribution plots of the relative abundance of various sedimenting species as a function of the sedimentation



coefficient (11). The SV-AUC based analysis of SRDPs in mAb solutions presents several advantages over SEC: (i) it is a physical first-principles-based method enabling computation of absolute masses, (ii) does not involve the use of mobile phases and experimentation is often possible in the formulation buffer, (iii) relatively lower sample dilution as compared with SEC, and (iii) minimal matrix effects, as the sector-shaped sedimentation velocity cell has a much lower surface area/volume ratio as compared with gel-filtration matrices.

In the present study, we subjected the NISTmAb to UV-A radiation, which is known to induce aggregation as well as hydrolysis in mAbs, to generate various SRDPs. The exposure duration was optimized to generate ~5% aggregates or HMWS. Exposure of the sample to UV radiation also resulted in the product of ~5% clips or low molecular weight species (LMWs). Three different labs analyzed the light-exposed, degraded NISTmAb mixtures by SV-AUC, and the results were compared with those from SEC for orthogonal verification of the SEC method (Table 3). An SEC method was performed as described in the Separation chapter/Volume 2, Chapter 5.

The cumulative results of the SV-AUC analyses, which accounted for intra-lab and inter-lab variability revealed the presence of a total of ~11% SRDPs, and ~89% monomer in the UV-light degraded the NISTmAb mixture (Figure 3A). A similar distribution was observed by SEC (Figure 3B, Table 3). The SRDPs in turn were found to be composed of ~5% of low molecular weight (LMW) and ~6% of HMWS. We note that although the variation in user-chosen resolution (100–451) during *c(s)* analysis and other variations experimental conditions may have led to differences in HMWS and LMW species peak shapes in the *c(s)* distributions, their relative amounts were very similar, within 20% relative error. Importantly, our analysis shows that SV-AUC is a robust method for the size-distribution analysis in mAb (protein) mixtures with relatively small intra-lab and inter-lab variability (Table 3). The good agreement between SV-AUC and SEC results in the estimated SRDPs shows that the developed SEC method may be suitable for the analysis of SRDPs in the NISTmAb solutions.



*Figure 3. (A) Fitted distribution from one sedimentation velocity-analytical ultracentrifugation (SV-AUC) experiment performed on UV-stressed NISTmAb. The SEDFIT program produced the distribution function based on 100 scans of concentration versus radial position fit with the settings and parameters indicated in Table 2. (B) Chromatogram of the size exclusion chromatography (SEC) analysis of control and UV-stressed NISTmAb. Analysis was performed on a TSKgel G3000SWxL column as described in the Separation chapter/Volume 2, Chapter 5.*

**Table 3. SV-AUC-Based Size-Distribution Analysis of UV Light-Degraded NISTmAb Mixtures<sup>a</sup>**

<i>Technique</i>	<i>Lab</i>	<i>Monomer(%)</i>	<i>LMWS (%)</i>	<i>HMWS (%)</i>	<i>Monomer s (svedberg)</i>	<i>RMSD Fit (AU)</i>
SV-AUC	1	88.63	6.87	4.50	6.343	0.0051
		89.31	5.99	4.70	6.345	0.0049
		89.64	6.02	4.34	6.339	0.0051
	<b>Average (SD)</b>	<b>89.19 (0.51)</b>	<b>6.29 (0.16)</b>	<b>4.51 (0.18)</b>	<b>6.343 (0.003)</b>	<b>NA</b>
	2	88.97	7.29	3.74	6.357	0.0047
		87.19	7.14	4.57	6.360	
		86.87	6.48	6.66	6.361	0.0055
	<b>Average (SD)</b>	<b>87.68 (1.13)</b>	<b>6.97 (0.43)</b>	<b>4.99 (1.50)</b>	<b>6.359 (0.002)</b>	<b>NA</b>
	3	90.22	5.01	4.77	6.254	0.0061
		89.07	5.03	5.91	6.284	0.0061
89.14		4.64	6.24	6.268	0.0060	
<b>Average (SD)</b>	<b>89.48 (0.53)</b>	<b>4.89 (0.18)</b>	<b>5.64 (0.63)</b>	<b>6.269 (0.012)</b>	<b>NA</b>	
<b>SV-AUC overall average (SD)</b>		<b>88.78 (1.66)</b>	<b>6.05 (1.18)</b>	<b>5.04 (1.73)</b>	<b>6.32 (0.05)</b>	
<b>SEC average, n = 3 (SD)</b>		<b>87.6 (0.1)</b>	<b>7.2 (0.1)</b>	<b>5.2 (0.1)</b>	<b>NA</b>	<b>NA</b>

AU, absorbance units; HMWS, higher molecular weight species; LMWS, lower molecular weight species; RMSD, root-mean-square deviation; s, sedimentation coefficient; SD, standard deviation; SEC, size exclusion chromatography; SV-AUC (sedimentation velocity-analytical ultracentrifugation). <sup>a</sup> Experiments were conducted at 20 °C and 40,000 revolutions per minute (RPM) using the absorbance optical system.

## Light Scattering

When light of sufficiently small wavelength (small compared to a solute's size) interacts with a solute, the light scatters in all directions. The intensity, angular dependence, and fluctuation in time of scattered light can provide a wealth of information on the biophysical properties of macromolecules in solution. Light scattering experiments are conducted in two modes: SLS and DLS. SLS is orthogonal to SE-AUC and enables first-principles analysis of absolute averaged-solution molecular weights, interaction energetics, molecular dimensions (radii of gyration when multiple scattering angles are monitored), and solution non-ideality. In DLS, analysis of macromolecular systems is based on measurement of the diffusion coefficients of macromolecules. DLS can be considered to be orthogonal to SV-AUC.

### Static Light Scattering

SLS refers to the time-averaged total scattered intensity of a sample measured at a given angle ( $\theta$ ) relative to an incident (laser) light source. Light scattering arises from the interaction between the oscillating electric field produced by the incident light and the electrons of the molecules in the sample. Detailed treatments of the underlying molecular origins of light scattering with full derivations of the working equations are available elsewhere (18–23). In the interest of brevity, a sufficiently general starting point for the analysis of many systems of interest is the following (24):

$$\frac{Kc}{R_\theta} = \frac{1}{\langle M \rangle_w P(q)} + 2B_2c + O(c^2) \quad (10)$$

where  $R_\theta$  is the excess Rayleigh ratio,  $c$  is the mass concentration,  $K$  is an optical constant,  $\langle M \rangle_w$  is the mass-averaged molecular weight  $\langle M \rangle_w = \sum_i c_i M_i / \sum_i c_i$  (where  $M_i$  and  $c_i$  are the molecular weight and mass concentration of the  $i$ th species, respectively),  $P(q)$  is the form factor, and  $B_2$  is the second virial coefficient.  $R_\theta$  can be determined experimentally as:

$$R_\theta = \left( \frac{I_s - I_0}{I_{ref}} \right) \left( \frac{n_0}{n_{ref}} \right)^2 R_{ref} \quad (11)$$

where  $I_s$  is the scattered intensity of the sample,  $I_0$  is the scattered intensity of the background (solvent, buffer),  $I_{ref}$  is the scattered intensity of a reference solution (commonly toluene or benzene),  $n_{ref}$  is the refractive index of the reference solution, and  $R_{ref}$  is the Rayleigh ratio of the reference solution (25).  $K$  is a function of both the instrument and of the sample of interest and is defined for vertically polarized incident light as follows:

$$K = \frac{4\pi^2 n_0^2 (dn/dc)^2}{N_A \lambda_0^4} \quad (12)$$

where  $n_0$  is the refractive index of the background (solvent, buffer),  $dn/dc$  is the refractive index increment (26),  $N_A$  is Avogadro's number, and  $\lambda_0$  is the wavelength of the laser in vacuo.

Form factor  $P(q)$  captures the decrease in scattered intensity resulting from interference between individual scattering centers within a large molecule and is often written using the Guinier approximation (27) as:

$$P(q) = 1 - q^2 \langle r_g^2 \rangle_z / 3 + O(q^4) \quad (13)$$

where  $q$  is the magnitude of the scattering vector ( $q = 4\pi n_0 \sin(\theta/2)/\lambda_0$ , where  $\theta$  is the scattering angle), and  $\langle r_g^2 \rangle_z$  is the z-average square radius

$\langle r_g^2 \rangle_z = \frac{\sum_i r_{g,i}^2 c_i M_i}{\sum_i c_i M_i}$  where  $r_{g,i}$  is the radius of the  $i$ th species. Combining Equations 10 and 13 yields:

$$\frac{Kc}{R_\theta} = \frac{1}{\langle M \rangle_w \left[ 1 - q^2 \langle r_g^2 \rangle_z / 3 + O(q^4) \right]} + 2A_2c + O(c^2) \quad (14)$$

Inspection of Equation 14 shows that  $\langle M \rangle_w$  is obtained at the limit of infinite dilution ( $c \rightarrow 0$ ) and ( $q \rightarrow 0, P(q) \rightarrow 1$ ) infinitely small scattering angle, and that  $\langle r_g^2 \rangle_z$  and  $B_2$  are obtained from the angular- and concentration-dependence of the scattered intensity, respectively.

A number of useful simplifying approximations can be applied to Equation 14 based on the composition of the sample of interest. If the characteristic length scale of the scatterers is small relative to the wavelength

of the laser ( $\langle r_g^2 \rangle_z^{1/2} \ll \lambda_0/20$ ), then the scattered intensity will be independent of the scattering angle (isotropic), and the contribution of the form factor can be neglected ( $P(q) \rightarrow 1$ ). Likewise, if the system is composed of only one macromolecular scattering species (monodisperse), then the weight- and z-averages are replaced by their single-species equivalents ( $\langle M \rangle_w \rightarrow M, \langle r_g^2 \rangle_z \rightarrow r_g^2$ ). Finally, if the mass concentration of the scatterers is low such that solution non-ideality does not contribute appreciably to the measured scattered intensity, then the term containing the second virial coefficient can be neglected ( $2B_2c \rightarrow 0$ ).

SLS is a versatile and powerful tool for characterizing both reversible and irreversible assembly processes in solution. Size exclusion chromatography with inline multi-angle light scattering (SEC-MALS) is now routinely used to determine molecular weight of monomers and higher order oligomers/aggregates (23, 28). In addition, composition gradient-multi-angle light scattering (CG-MALS) has

been used to characterize both self- and hetero-association (29–31). As one of the few techniques capable of directly interrogating concentrated systems, SLS will continue to play an important role in HOS characterization.

To characterize the NISTmAb, SLS measurements were performed using an ALV CGS-3 Compact Goniometer System with a LSE-5003 Light Scattering Electronics and Multiple Tau Digital Correlator and a 22 mW HeNe laser at 632.8 nm (Langen, Germany). Scattering from all samples was confirmed to be isotropic, and the results from each scattering angle were subsequently averaged. Additional experimental parameters are included in Table 4.

The results of SLS measurements for the NISTmAb in 25 mM His, pH 6.0 buffer with and without 150 mM NaCl are shown in Figure 4. The data are well-described by Equation 10 with  $P(q) \rightarrow 1$  and keeping only the first term in the virial expansion:  $Kc/R_\theta = 1/M_w + 2B_2c$ . The best-fit values and estimated uncertainties for  $M_w$  and  $B_2$  are shown in Table 5. The value of  $B_2$  decreases by approximately an order of magnitude in the presence of 150 mM NaCl, indicating that the increased ionic strength is screening net repulsive electrostatic protein–protein interactions. There is excellent agreement between the SLS estimates for  $M_w$  obtained in the two different buffer systems and with values obtained from SE-AUC (150 kDa) and mass spectrometry (148.3 kDa) analyses.

**Table 4. Experimental Static Light Scattering Parameters Applied for the NISTmAb Analysis**

<i>Parameter</i>	<i>Value</i>
Temperature	Ambient (~23 °C)
Scattering angles	70°, 80°, 90°, 100°, 110°
Averaging time (per angle)	10 sec
Replicates (per angle)	3
Refractive index increment (dn/dc)	0.185 mL/g
Reference solution	Toluene

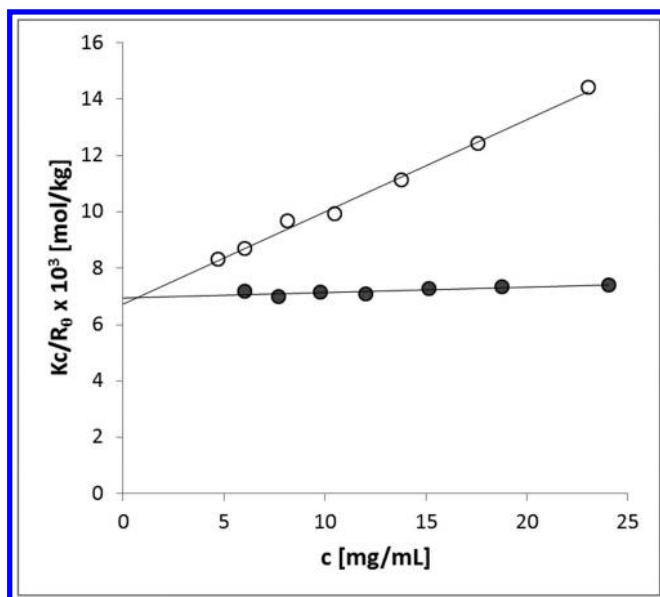


Figure 4. Static light scattering data for NIST IgG1 mAb in 25 mM histidine, pH 6.0 buffer with (filled circles) and without (open circles) 150 mM sodium chloride. Solid lines are best fits to the data from linear least-squares regression.

**Table 5. Fitted Parameters from the Static Light Scattering Experiment on the NISTmAb**

	25 mM Histidine, pH 6.0	25 mM Histidine 150 mM NaCl, pH 6.0
$M_w$ (kDa)	$148 \pm 3$	$144 \pm 2$
$B_2 \times 10^4$ (mol mL/g <sup>2</sup> )	$1.64 \pm 0.06$	$0.10 \pm 0.03$

### Dynamic Light Scattering

DLS, also called quasi-elastic light scattering (QELS) or photon correlation spectroscopy (PCS), uses the time fluctuations in Rayleigh scattering intensity in a colloidal solution exhibiting Brownian motion to determine collective diffusivities and hydrodynamic particle radius (e.g., Stokes radius via the Stokes-Einstein equation for diffusion of spherical particles through a liquid with low Reynolds number). The collective diffusivity relates the flux to the gradient in concentration according to Fick's law. In contrast to SLS and MALS detectors that measure the time-averaged intensity of light scattered, DLS fast photon counter detectors measure the time-dependent fluctuations of scattered light intensity (resulting from interference of light scattered from the mobile colloids or proteins in solution) on the order of 100 ns to 100 ms (32, 33). Dynamic information about

the relative motion of particles is derived from the second order correlation function, given by:

$$g^{(2)}(\tau) = \langle I(t)I(t+\tau) \rangle / \langle I(t) \rangle^2 \quad (15)$$

where  $I(t)$  is the intensity of the scattered light at time  $t$ ,  $\tau$  is a delay time, and the brackets ( $\langle \rangle$ ) indicate averaging over all  $t$ . The exponential decay of this function occurs due to the randomizing tendency of Brownian motion and is mathematically related to the collective or mutual diffusivity coefficient, as follows (32):

$$g^{(2)}(\tau) = B + \beta \exp(-2\Gamma \tau) \quad (16)$$

where  $B$  is the baseline of the correlation function at infinite delay;  $\beta$  is the correlation function amplitude at zero delay; and  $\Gamma$  is the decay rate, which can be obtained using a nonlinear least squares fit through the correlation function. The collective diffusivity is then obtained from  $\Gamma$  using  $D = \Gamma/q^2$ , where  $q$  is the magnitude of the scattering vector and is given by:

$$q = \frac{4\pi n_0}{\lambda_0} \sin\left(\frac{\theta}{2}\right) \quad (17)$$

where  $n_0$  is the solvent index of refraction,  $\lambda_0$  is the wavelength of the incident light, and  $\theta$  is the scattering angle.

DLS is a useful tool at relatively low concentrations, where contributions from multiple scattering are small or negligible. From the collective diffusivity versus concentration plot, the diffusion interaction parameter,  $k_D$ , can be obtained from:

$$D = D_0(1 + k_D c) \quad (18)$$

where  $D_0$  is the self-diffusivity and  $c$  is the concentration. This approximation is typically valid when  $c < 20$  mg/mL (34, 35). Using the self-diffusivity,  $D_0$ , the intercept of the plot of Equation 18, the effective hydrodynamic radius,  $r_h$ , of the protein diffusing in the solvent can be obtained from the Stokes-Einstein equation:

$$r_h = \frac{kT}{6\pi\eta D_0} \quad (19)$$

where  $k$  is Boltzmann's constant,  $T$  is the temperature in Kelvin, and  $\eta$  is the solvent viscosity.

The diffusion interaction parameter,  $k_D$ , measures the deviations of  $D$  from solution ideality as a consequence of protein interactions. In general, a large positive  $k_D$  indicates net repulsive interactions, and a large negative  $k_D$  indicates net attractive interactions. Note that for Brownian hard spheres,  $D = D_0(1 + 1.47\phi)$ , where  $\phi$  is the volume fraction of the spheres (35, 36). In this case, the Brownian hard spheres only interact due to their excluded volume and hydrodynamics, but no other interparticle interaction. An effective hard sphere volume fraction for proteins can be estimated in the dilute regime using:



$$\phi = \frac{4\pi c N_A r_h^3}{3M} \quad (20)$$

where  $c$  is the concentration in g/mL,  $M$  is the protein molecular weight,  $N_A$  is Avogadro's number, and  $r_h$  is the hydrodynamic radius. For mAbs,  $r_h \sim 5.5$  nm and  $M \sim 145$  kDa, thus the effective volume occupied by the model spheres representing the proteins can be obtained from Equation 19 and is  $\phi \sim 2.9 c$ . In this model, if the spheres representing the proteins behave like hard spheres without additional intermolecular interactions, then  $D = D_0 (1 + 4.26 c)$  so that  $k_D \sim 4.26$  mL/g. Therefore a  $k_D$  larger than this indicates substantial repulsive interactions (35, 36). The  $k_D$  and the osmotic virial  $B_2$  coefficient, , are related according to the equation:

$$B_2 = \frac{k_S + k_D + v}{2M} \quad (21)$$

where  $k_S$  is the sedimentation interaction parameter,  $v$  is the partial specific volume, and  $M$  is the molecular mass (34, 37).

To measure the  $k_D$  for the NISTmAb, a concentration series of 20, 15, 10, 5, 2, and 1 mg/mL protein was prepared in 25 mM histidine buffer at pH 6.0. A separate concentration series was prepared in 25 mM histidine buffer at pH 6.0 with 150 mM NaCl. Each of these formulations were filtered through a 0.1 micron filter prior to the measurements using a centrifuge filter that also helped remove air microbubbles. Aliquots (60  $\mu$ L) of the filtered samples were transferred into sterile, 384-well, glass-bottom Greiner Sensoplates (Greiner Bio-One, Monroe, NC), and placed in the DLS instrument, the DynaPro PlateReader Plus (Wyatt, Santa Barbara, CA) with the laser wavelength fixed at 828.9 nm. Wyatt Technology Dynamics software was used to schedule and automate three 20-second acquisitions for each sample. Measurements were performed at 20  $^{\circ}$ C. The collective diffusivity,  $D$ , was determined for each IgG1 solution at the given protein concentrations and in the two buffer systems. Sample replicate ( $n = 3$ ) data for each concentration were averaged to reduce systematic error in the sample preparation and analysis. The diffusivity versus concentration plot is shown in Figure 5.

The significant variability in the diffusivity for the 1 and 2 mg/mL concentrations for the high salt case may be due to a small amount of unintended microbubbles during the liquid transfer process. A weighted least squares method was used to calculate the slope and intercept of the diffusivity versus concentration plot in accordance with Equation 18. The results for the self-diffusivity,  $D_0$ , the diffusion interaction parameter,  $k_D$ , and the hydrodynamic radius,  $r_h$ , for the 0 and 150 mM NaCl cases at 20  $^{\circ}$ C are shown in the Table 6.

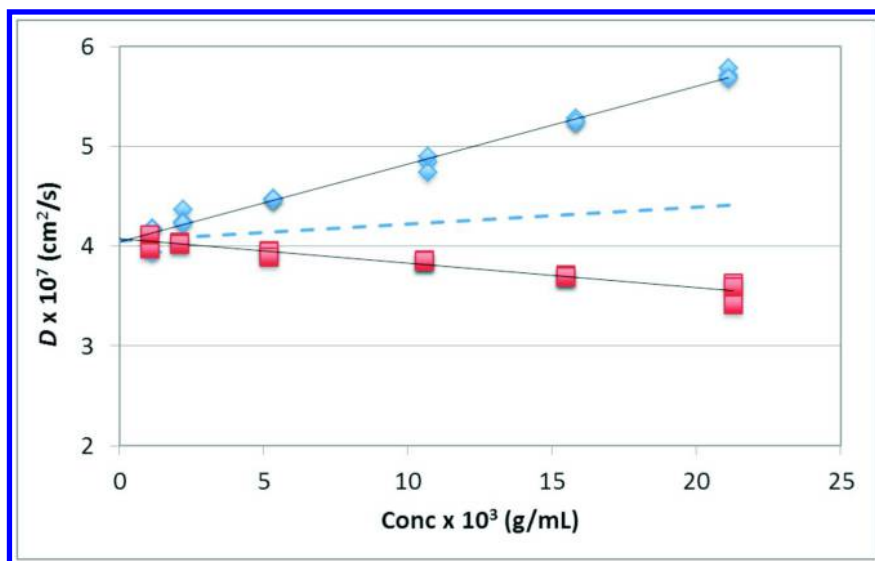


Figure 5. The collective diffusivity,  $D$ , for the NISTmAb is shown as a function of concentration for two different buffer systems. The diamonds represent the  $D$  results in 25 mM histidine buffer, pH 6.0, and the squares represent the  $D$  in 25 mM histidine buffer, pH 6.0, with 150mM NaCl. The dashed line represents the theoretical  $D$  based on the  $k_D$  for colloidal hard spheres. The results for  $k_D$  using a weighted least squares method are  $k_D = 18.7 \pm 1.0$  mL/g and  $-7.1 \pm 1.9$  mL/g for the 0 and 150 mM NaCl cases, respectively (Lab 1 data).

**Table 6. DLS Best Fit Parameters Determined for the NISTmAb by Two Independent Labs**

	Buffer: 25mM His/His HCl with 0 mM NaCl, pH 6.0	Buffer: 25mM His/His HCl with 150 mM NaCl, pH 6.0
<b>Lab 1</b>		
$D_0 \times 10^7$ (cm <sup>2</sup> /s)	4.04 ± 0.04	4.08 ± 0.09
$k_D$ (mL/g)	18.74 ± 1.07	-7.07 ± 1.93
$r_h$ (nm)	5.45 ± 0.05	5.26 ± 0.12
<b>Lab 2</b>		
$D_0 \times 10^7$ (cm <sup>2</sup> /s)	3.89 ± 0.03	4.08 ± 0.02
$k_D$ (mL/g)	13.9 ± 0.5	-4.3 ± 0.3
$r_h$ (nm)	5.57 ± 0.04	5.72 ± 0.02

His, histidine.

The results for  $D$  for the NISTmAb in the buffer with no NaCl showed a positive slope and  $k_D$  of 13.9 to 18.7 mL/g based on two independent lab measurements, indicating repulsive self-interactions at these low concentrations. On the other hand, the collective diffusivity results for the IgG1 in the buffer with 150 mM NaCl showed a negative slope and  $k_D$  of  $-4.3$  to  $-7.1$  mL/g, potentially indicating weak attractive interactions. For reference, a dashed line in Figure 5 shows the theoretical  $D$  based on the  $k_D$  for colloidal hard spheres (36, 37) with the same self-diffusivity and density as the NISTmAb. In the limit as concentration approaches zero,  $D$  becomes the self-diffusivity,  $D_0$ , and the intercept on the vertical axis in Figure 5 shows that it is approximately 4.0 and  $4.1 \times 10^{-7}$  cm<sup>2</sup>/s for the 0 and 150 mM NaCl cases, respectively. The hydrodynamic radius calculated from these are  $r_h = 5.4$  nm and 5.3 nm. For the case without salt, the  $B_2$  obtained for the NISTmAb from the previous section is  $1.6 \times 10^{-4}$  mol mL/g<sup>2</sup>, which indicates repulsive interactions, consistent with the conclusion based on the large positive  $k_D$  value measured here. It is interesting to note that the  $B_2$  corresponding to the NISTmAb in buffer with 150 mM NaCl is small but positive, around  $1.0 \times 10^{-5}$  mol mL/g<sup>2</sup>, whereas its corresponding  $k_D$  is negative. However, closer examination reveals that this value of  $B_2$  is below that predicted from a hard sphere model, which is  $7.1 \times 10^{-5}$  mol mL/g<sup>2</sup> (38), thus pointing to a weak attraction in the presence of 150 mM NaCl in the concentration range studied.

## Membrane-Confined Analytical Electrophoresis

Protein charge is a fundamental property that directly influences its structure, stability, solubility, and ability to interact with other macromolecules. The biopharmaceutical industry is striving to increase therapeutic protein concentrations in order to decrease dosage volume. However, solutions of higher protein concentration may exhibit high viscosity and be prone to phase separation. These unfavorable solution properties are a reflection of the colloidal properties of proteins. Charge–charge repulsion is the only long-range proximity energy that maintains protein solubility and can overcome the attractive forces that lead to high viscosities. Therefore, it is important to have routine ways to measure protein charge.

The net charge is not a property of a protein. Rather, the charge on a protein is a system property that may vary significantly with solvent composition, pH, dielectric constant, and temperature. Therefore, the protein charge, calculated by summing up the charge on each of its ionizable groups, may differ substantially from an experimentally measured value because the calculated charge only takes H<sup>+</sup> binding into account and it is known that proteins may bind other ions, particularly anions, and electrostatic interactions between ionizable sites can alter the charge and lead to shifts in pKa.

Membrane-confined electrophoresis (MCE) is a first-principle method for measuring the electrophoretic mobility,  $\mu$ , effective charge,  $z_{eff}$ , and Debye–Hückel–Henry charge,  $z_{DHH}$ , of macromolecules (39). The electrophoretic mobility is the ratio of the velocity of the macromolecule,  $V$ , to the electric field,

E:  $\mu = V/E$ , and is the fundamental measurement made in free-boundary (also called real-time) electrophoresis. The other two quantities,  $z_{eff}$  and  $z_{DHH}$ , are calculated from  $\mu$  as described below.

The commercial MCE instrument from Spin Analytical (Durham, NH) was used to perform free-boundary (i.e., there is no supporting gel) electrophoretic mobility measurements and steady-state electrophoresis measurements. This instrument is composed of a  $2 \times 2 \times 4$  mm<sup>3</sup> quartz cuvette whose ends are sealed with semipermeable membranes (8000 molecular weight cutoff [MWCO], Spectra/Por Biotech grade) that allow for the flow of water and small ions but effectively trap the macromolecule of interest. An electric field gradient can be applied lengthwise across the cuvette (i.e., membrane to membrane), and absorbance across the cuvette can be measured as a function of distance from one end. Movement of macromolecules under applied electric field can be monitored in a dynamic mode (free-boundary electrophoretic mobility) or after a static equilibrium has been established (steady-state electrophoresis), both of which are described below.

### Free-Boundary Electrophoretic Mobility

This technique requires a 20  $\mu$ L aliquot of a 1 mg/mL protein sample loaded into a  $2 \times 2 \times 4$  mm<sup>3</sup> quartz cuvette whose ends are sealed with semipermeable membranes (8000 MWCO, Spectra/Por Biotech grade) that allow for the flow of water and small ions, but effectively trap the macromolecule of interest. Absorbance across the cuvette can be measured as a function of distance from one end. Throughout the experiment, the volume outside of the cuvette is replenished continuously with fresh solvent at a constant flow of  $\sim 50$   $\mu$ L/min. Initially, the macromolecule is at a uniform concentration in the cuvette. The electric field of 8.0 V/cm was applied longitudinally across the cell (i.e., membrane to membrane) and maintained by a constant current power supply. The charged macromolecules migrate in the electric field from one end of the cuvette to the other, resulting in a moving boundary across the chamber. The intensity of this boundary is measured as a function of distance,  $I(x)$ , across the chamber at multiple time points.

Three independent experiments were performed with fresh 1 mg/mL NISTmAb samples. The 280 nm light intensity scans were obtained for approximately 20 minutes at 10 to 15 second intervals after power was applied. The migrating boundary is presented in Figure 6A, with the starting and ending scans selected for mobility calculations marked in color. Mobility data were analyzed using the time-difference software supplied by Spin Analytical. The software subtracts sequential intensity scans to produce several time difference,  $\Delta I/\Delta t$ , scans. The time-difference scans would present a series of peaks whose midpoint positions would move at the same pace as the boundaries. By dividing the position at each point by the time (an average of the time for the scans used to calculate  $\Delta I/\Delta t$ ), the scan was acquired, the x-axis is converted from distance to electrophoretic mobility, and the  $\Delta I/\Delta t$  curves are superimposed on the peaks in the curve at the electrophoretic mobility of the boundary. The resulting  $\Delta I/\Delta t$  curves (e.g., Figure 6B, showing  $z_{DHH}$ ) provide a visual representation of the distribution. The shape of the peak is partly a consequence of the distribution

of charge isotopes (and their lifetimes) in the sample, but it also is affected by diffusion and conductance changes across the boundary. If there is a second peak in the distribution, it indicates that there is a long-lived charge isomer. At present, it appears that a second peak will be visible if the sample contains ~10% of a component having a twofold charge difference (e.g., dimer or half-mer) from the dominant species.

The experimental quantities useful for charge determination are the electrophoretic mobility (the ratio of the velocity of a molecule, in cm/s, to the electric field, in V/cm,  $\mu = \text{cm}^2/\text{V s}$ ) and the effective valence:

$$z_{eff} = \frac{\mu}{fQ_p} \quad (22)$$

where  $f$  is the translational friction coefficient, determined experimentally from either sedimentation or diffusion measurement, and  $Q_p$  is the proton elementary charge (40). By dividing by  $Q_p$ ,  $z_{eff}$  is actually a valence (e.g.,  $\text{Na}^+$  has a +1 valence and  $\text{SO}_4^{2-}$  has a -2 valence) rather than a charge, which would have the units of coulombs. Both  $\mu$  and  $z_{eff}$  include the effects of solvent ion shielding and the “electrophoretic effect,” which results from the distortion and transport of the ion atmosphere in the vicinity of the protein. Consequently,  $\mu$  and  $z_{eff}$  do not distinguish between bound ions and the Debye-Hückel ion cloud.  $z_{DHH}$  adjusts for the solvent shielding through the Debye-Hückel approximation (Equation 23). By removing the effects of the solvent,  $z_{DHH}$  is a measure of the charge (valence) on the molecule. It is important to recognize that ions may be bound to a protein either at a particular site (e.g., visible in X-ray or nuclear magnetic resonance [NMR] structures) or as territorially bound ions (confined to a region on the molecule that has a high electrical potential such that they cannot dissociate from that region). Both site-bound and territorially bound ions contribute to the charge on the macromolecule (i.e.,  $z_{DHH}$ ). In contrast to bound ions, the Debye-Hückel cloud is the disproportionation of the co-ions and counter-ions in the immediate vicinity of the macromolecule and is the response of the solvent to the presence of the charged macromolecule. The Debye-Hückel cloud does not contribute to  $z_{DHH}$ . Calculation of the Debye-Hückel-Henry valence,  $z_{DHH}$ , adjusts for these effects.  $z_{DHH}$  may be calculated from either  $z_{eff}$  or  $\mu$ :

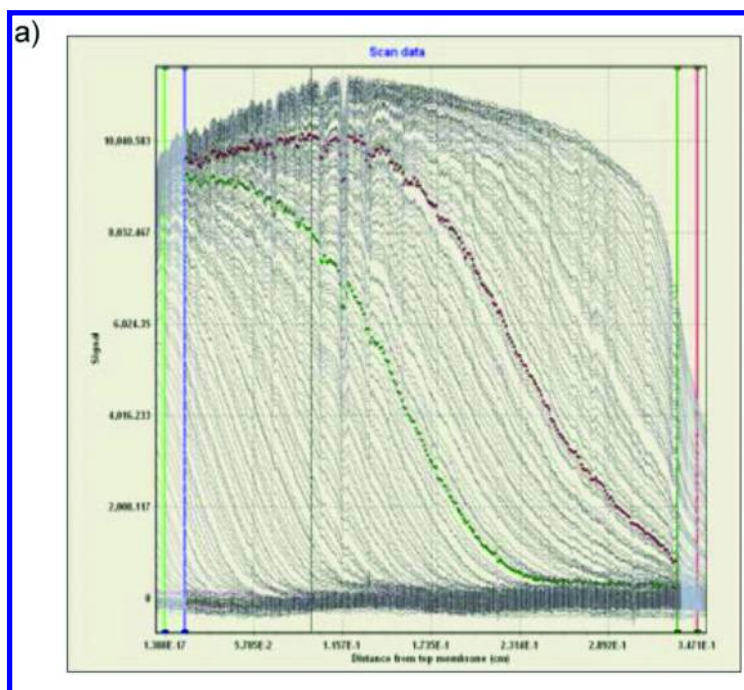
$$z_{DHH} = z_{eff} \frac{(1+\kappa_D a)}{H} = \mu f \frac{(1+\kappa_D a)}{HQ_p} \quad (23)$$

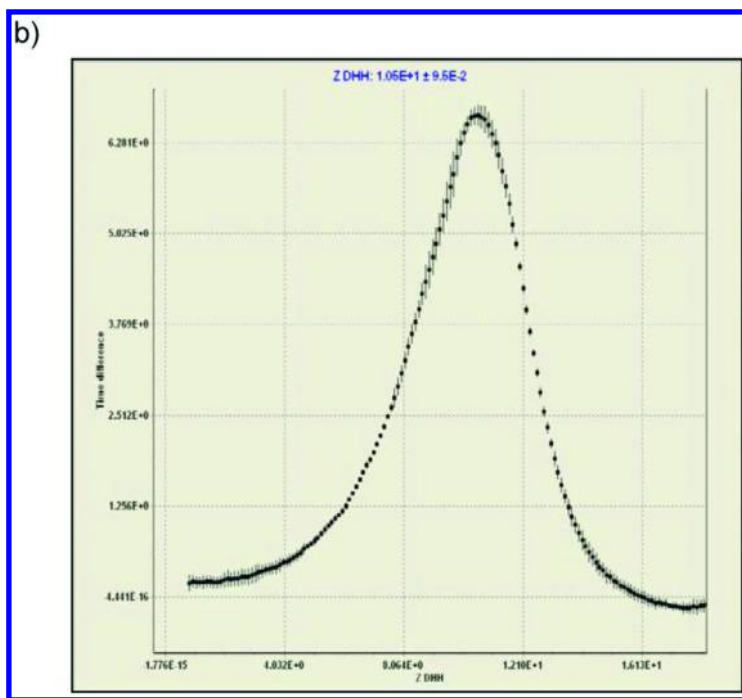
where  $k_D$  is the inverse Debye length (in  $\text{cm}^{-1}$ ), which depends on the temperature and the square root of the ionic strength;  $a$  is the sum of the Stokes radii of the protein and its counterion; and  $H$  is Henry’s function (40).

For the NISTmAb in 25 mM His, 150 mM NaCl, pH 6.0,  $a = 5.68$  nm (the Stokes radius of the NISTmAb is 5.5 nm, and for  $\text{Cl}^-$  is 0.18 nm) and  $k_D = 3.9 \times 10^{-6} \text{ cm}^{-1}$ .

The results presented in Figures 6A and 6B for the NISTmAb sample yield an electrophoretic mobility measurement of  $2.36 \times 10^{-5} \pm 0.02 \text{ cm}^2/\text{V s}$ , resulting in a calculated  $z_{DHH}$  of  $10.8 \pm 0.1$ . The anticipated precision of charge measurements from electrophoretic mobility measurements has been estimated to

be 7% for serum albumin (41), and the accuracy for the charge measurement by electrophoresis has been estimated to be 8% for hen egg white lysozyme (40). It is anticipated that the precision and accuracy for this IgG1 will be comparable. Thus, given this level of uncertainty,  $z_{DHH}$  for the NISTmAb molecule is calculated to be  $11 \pm 1$ . Because charge–charge repulsion contributes to protein solubility and reduced solution viscosity, protein charge is an important contributor to the colloidal properties of a molecule. For a molecule with a radius of  $\sim 5$  nm, the NISTmAb would be predicted to have marginal colloidal stability. The concentration-dependent DLS data (Figure 5) are in accord with this prediction; in low salt, where charge–charge repulsion is greater,  $k_D$  is positive, indicating net repulsive interactions, whereas in 150 mM salt, where charge–charge repulsion is reduced,  $k_D$  is negative, indicating net attractive interactions. These observations are in accord with the proximity energy framework (42).





*Figure 6. Charge ( $z_{DHH}$ ) determination for the NISTmAb in 150 mM NaCl, 25 mM histidine, pH 6.0 by membrane-confined electrophoresis (MCE). (A) Raw MCE intensity scans acquired at 10–15 second intervals during electrophoresis. The data show the light intensity (vertical axis) as a function of the distance (in cm) moved from the membrane (horizontal axis). As the protein boundary passes a position, the light intensity increases. The time difference analysis is done in between the green and red concentration traces. The highlighted scans encompass the data used for analysis, and the vertical lines enclose the region used for subsequent analysis. The first-difference ( $\Delta c/\Delta t$ ) electrophoretic mobility distribution is calculated for successive 2-minute intervals (10 scans) of the selected data in Panel A. The mobility calculated from these data is  $2.36 \times 10^{-5} \pm 0.02 \text{ cm}^2/\text{V s}$ . (B) Charge distribution ( $z_{DHH}$ ) calculated from the mobility distribution using a measured Stokes radius of 5.5 nm for the NISTmAb is  $10.8 \pm 0.1$ . The vertical axis shows the time derivative ( $\Delta I/\Delta t$ ) of the intensity data in panel A as a function of  $z_{DHH}$  (horizontal axis). The peak position is  $z_{DHH}$ . (see color insert)*

### Membrane-Confined Electrophoresis-Steady-State Electrophoresis (MCE-SSE)

MCE-SSE is a steady-state electrophoresis technique that provides a direct measurement of protein's effective charge,  $Z \times e$  (39). For the MCE-SSE measurements, the measurements were conducted in the same instrument as the mobility measurements at  $20.0 \text{ }^\circ\text{C} \pm 0.1 \text{ }^\circ\text{C}$ , using an electric field of 0.4 V/cm. Blank scans were acquired after the NISTmAb had been pushed to the bottom

membrane using a field of 8 V/cm for approximately 10 minutes. The program WinMatch ([www.rasmb.org](http://www.rasmb.org)) was used to establish that steady state had been reached, which required ~18 hours for each field. Concentration profiles from the different fields were analyzed by using the WinNonlin ([www.rasmb.org](http://www.rasmb.org)) program to obtain the  $z_{eff}$  and the ZUtilities program ([www.rasmb.org](http://www.rasmb.org)) to calculate  $z_{DHH}$ . The graph presented in Figure 7 is a representative, time-invariant concentration gradient for the NISTmAb measured by MCE-SSE.

From the exponent,  $\sigma$ , equal to  $21 \pm 2$  (Figure 7),  $z_{eff}$  is determined to be  $1.33 \pm 0.13$ , which results in a calculated value of  $z_{DHH}$  of  $9.7 \pm 0.8$ . Calculations of  $z_{eff}$  and  $z_{DHH}$  were made using ZUtilities, with  $E = 0.4$  V/cm, an ionic strength of 0.172 mM,  $R_s = 5.5$  nm for the protein,  $R_s = 0.18$  nm for  $Cl^-$ , and a dielectric constant of 78. The value of  $z_{DHH}$  from the steady-state measurements is consistent with the value obtained from the electrophoretic mobility measurements.

MCE-SSE is considered the most accurate way to determine protein charge. The accuracy stems from the facts that there are no sharp concentration gradients to affect the electric field and that the macromolecule does not contribute to the current and hence does not contribute to the electric field. However, the long times to reach steady-state (hours) compared to the time needed to measure the mobility (minutes) makes mobility measurements preferable.

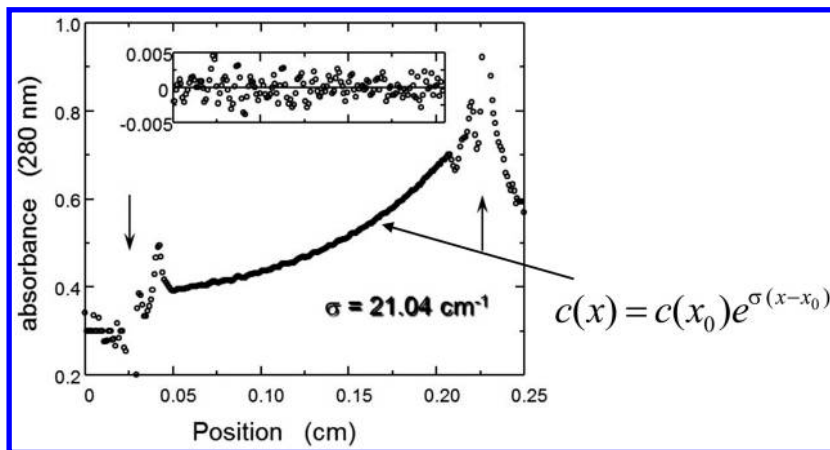


Figure 7. Steady-state electrophoresis results for the time-invariant mAb concentration gradient (measured as  $A_{280}$ ) at 0.4 V/cm. The same MCE instrument is used as for the mobility determination (Figure 6), but at a much lower electric field. At the lower field, the electrophoretic flux is exactly balanced by the diffusive flux at each position in the cuvette. The exponent,  $\sigma$ , is given by  $\sigma = \frac{Q_{eff}E}{Q_P k_B T}$ , where  $z_{eff} = Q_{eff}/Q_P$ ,  $E$  is the electric field,  $k_B$  is the Boltzmann constant, and  $T$  is the absolute temperature. The vertical arrows mark the top and bottom membrane positions. The insert shows the residuals to the nonlinear least-squares fit of the concentration profile to a single exponential.



# Signatures of Secondary and Tertiary Structure

## Differential Scanning Calorimetry

Thermal stability is a significant characterization parameter that is an indicator of overall protein stability and can be important for many aspects of protein expression and functioning. Thermal stability is described by the resistance of a protein to unfolding at elevated temperatures. The assessment of thermostability includes the monitoring of protein unfolding as a function of temperature. This is usually accomplished by following a change in heat capacity, fluorescence, or CD with increasing temperature. The midpoint of the transition from the fully folded state to the fully unfolded one is taken as the protein melting point,  $T_m$ . Low  $T_m$  indicates a low thermal stability and a low energy barrier for conformational changes caused by protein–protein, protein–surface (43), or other protein interactions. The major method for measurement of thermal stability is DSC, which requires a significant amount of protein and the typical sample scan takes about 1–2 hours to complete, depending on the temperature range and the scanning rate. The alternative way to measure protein thermostability is based on the fluorescence of extrinsic probes sensitive to a hydrophobic environment (44). The method is called DSF. The fluorescence intensity is monitored during a temperature scan, and the midpoint of the measured transition is determined as the melting temperature. Originally, this method was applied to the screening of ligand binding to the target protein (45), but it also has been used for the evaluation of crystallization and general stability (46, 47). DSF also has been applied to the screening of antibodies in different formulations and showed a good correlation with the unfolding temperature determined by DSC (48).

DSC finds important application in the characterization of protein stability and structure. Stability is a major challenge associated with protein drug products, and almost every degradation pathway, such as aggregation, deamidation or oxidation, can stem from or lead to the loss of native protein structure. DSC is unparalleled for measuring the thermodynamic stability in terms of the Gibbs free energy of unfolding under solution conditions conducive to ensuring reversibility (49). However, thermodynamic stability is difficult to measure in “real,” pharmaceutically relevant, multidomain proteins such as antibodies. The necessary condition of reversibility for thermodynamic analysis is often difficult to achieve for such large proteins, particularly in solution conditions of relevance to biopharmaceutical formulations. Nonetheless, DSC has found widespread application in screening molecules and solution conditions using domain midpoint transition temperatures ( $T_m$ ) or onset temperatures of unfolding. In many cases, the  $T_m$  serves as a practical measure of protein conformational stability (50). Importantly, the  $T_m$  for a given fold of a protein domain and for a given set of solution conditions can be expected to remain constant. Any changes in  $T_m$  can indicate subtle misfolding, covalent degradation-induced reduction in conformational stability, or a change in solvent composition. This fact can be employed effectively to make a comparative assessment of changes to the HOS of the protein. The effect of solutes on  $T_m$  is routinely employed to select favorable formulation solution compositions.

Here, we measured the thermal unfolding transition using a capillary-VP-DSC (MicroCal, MA) instrument. The experimental parameters chosen are detailed in Table 7.

**Table 7. Experimental DSC Parameters Applied to the NISTmAb Analysis<sup>a</sup>**

<i>Parameter</i>	<i>Value</i>
Approx. protein conc. (mg/mL) <sup>b</sup>	0.5
Starting Temperature (°C)	30
Final Temperature (°C)	110
Scan rate (°C/h)	60
Prescan thermostat (min)	15
Postscan thermostat (min)	0
Filter period (s)	10
Feedback	None
Molar Mass (kDa)	148
Ext. coeff. at 280 nm (mL/mg cm)	1.42

DSC, differential scanning calorimeter. <sup>a</sup> Instrument used was a MicroCal Capillary VP-DSC with autosampler. <sup>b</sup> Actual protein concentration was measured using a UV-visible spectrophotometer. The measured protein concentration was used to normalize the thermograms during data analysis.

Additional experimental details were as follows. Thermograms for the NISTmAb were acquired in both 25 mM His/His-HCl and 25 mM His/His-HCl plus 150 mM NaCl solutions. For each solution condition, two buffer scans were acquired prior to each protein (the NISTmAb) scan to establish thermal history. Between injections, the cells were rinsed with formulation buffer without NaCl (1 syringe, 2 value rinses). No “cleaning” cycle between samples was used. During data analysis, a given buffer scan was subtracted from its corresponding sample scan following which a progressive baseline was determined. Samples were normalized to moles of protein loaded. The  $T_m$  for each unfolding event was determined by integrating each peak. The results of the DSC studies are tabulated in Table 8.

As seen (Figure 8), the thermal unfolding of the NISTmAb is characterized by three endotherms at ~70, 80, and 90 °C. Based on previous calorimetric studies with IgGs, the first transition can be assigned to the melting of the C<sub>H2</sub> domain, followed by the C<sub>H3</sub>, and finally by the Fabs. In this study, the NISTmAb displays peculiar behavior as in most antibodies, the C<sub>H3</sub> domain unfolding transition is usually observed after the Fabs (which display the largest endotherms) (50, 51). However, this reversal of order of unfolding transitions precisely would form

a characteristic HOS imprint for the NISTmAb. Further, the presence of NaCl appears to destabilize the NISTmAb (Figure 8), consistent with the salt effects observed for other antibodies (50). There is a consistent decrease in the  $T_m$ s of all three domains in the presence of NaCl, which is also consistent with previously reported studies on other mAbs (50).

**Table 8. DSC Results for the NISTmAb Analysis**

<i>Sample</i>	$T_{m1}$ (°C)	$T_{m2}$ (°C)	$T_{m3}$ (°C)
<b>Solvent: 25 mM His/His-Cl, pH 6</b>			
Replicate 1	69.179	82.967	93.434
Replicate 2	69.139	83.102	93.391
Replicate 3	69.268	83.239	93.366
mAb no salt average	69.195	83.103	93.406
mAb no salt std. dev.	0.066	0.136	0.025
<b>Solvent: 25 mM His/His-Cl, 150 mM NaCl, pH 6</b>			
Replicate 1	66.973	81.461	92.285
Replicate 2	67.123	81.438	92.265
Replicate 3	66.958	81.443	92.272
mAb salt average	67.018	81.447	92.274
mAb salt std. dev	0.091	0.012	0.010

DSC, differential scanning calorimeter; His, histidine;  $T_m$ , melting point.

We also evaluated the reversibility of unfolding of domains in three independent experiments in which the temperature was ramped to different temperatures, cooled, and rescanned. In the His/His-HCl buffer system, C<sub>H2</sub> domain displays reversibility of unfolding, whereas the C<sub>H3</sub> and Fab transitions appear to be irreversible in this buffer system (Figure 8). The addition of NaCl also appears to reduce the reversibility of the C<sub>H2</sub> domain further, suggesting a reduction in conformational stability of the molecule. Loss of reversibility has been previously correlated to increased aggregation in solution (52). These data are consistent with the results of the complementary DSF studies discussed below.

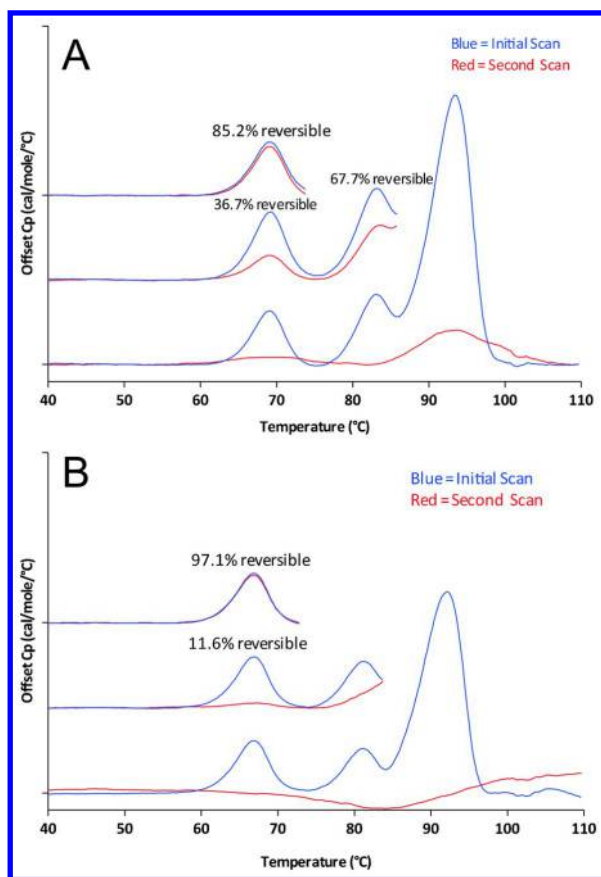


Figure 8. Differential scanning calorimetry (DSC) profiles of the NISTmAb in 25 mM histidine, pH 6.0 buffer without (A) and with (B) 150 mM NaCl at three different maximum temperatures. The NISTmAb was analyzed three times in each buffer system to evaluate the reversibility of the domains. For each analysis, the initial scan is shown in blue and the second rescan to assess percent reversibility is in red. (see color insert)

### Differential Scanning Fluorimetry

As a complimentary technique to DSC, DSF measurements were performed according to the previously developed method (48) with some modifications. NISTmAb solutions at corresponding concentrations were added to a 96-well plastic microplate at 49.0  $\mu$ L per well. SYPRO® Orange dye (Invitrogen Inc. Carlsbad, CA) was diluted (1:100 dilution of stock) in 10 mM acetate, pH 5.0 buffer, and 1  $\mu$ L of the diluted dye was added to each well, making the final dilution 1:5000. DSF measurements were made using a BioRad CFX96 thermal cycler platform instrument (Bio-Rad Laboratories, Inc., Hercules, CA).

The fluorescence of the dye during protein melting was monitored using the fluorescence resonance energy transfer (FRET) mode while the temperature was increased from 20.0 to 95.0 °C by 0.2 °C increments. Samples were equilibrated at each temperature for 5 seconds before taking the fluorescence readings. The fluorescence signal was then plotted as a function of temperature, and the temperature of hydrophobic exposure,  $T_h$ , was determined as the mid-point of the lowest temperature transition from the first derivative plot as determined by the CFX Manager™ software (Bio-Rad Laboratories, Inc., Hercules, CA).

The NISTmAb was characterized by DSF in two buffers (Table 9) at concentrations of 1, 2, 5, and 10 mg/mL. The results are shown in Figure 9. The temperatures of hydrophobic exposure are shown in Table 9. As is shown in Figure 9, the NISTmAb has a DSF profile similar to a typical IgG profile, with a  $T_h$  value between 60 and 70 °C. The slight  $T_h$  decrease in the presence of NaCl also is expected. No significant change of  $T_h$  is observed at different protein concentrations. At pH 6.0, the temperatures of unfolding for the C<sub>H</sub>2 and Fab domains of the IgG molecule are overlaid, and it is difficult to distinguish them on a DSF scan. The method is applicable to a wide range of protein concentrations and formulation conditions. A significant limitation for the measurements could be the presence of detergent in the formulation, which interferes with Sypro® Orange and some other extrinsic probes. New probes insensitive to surfactants have been introduced to avoid this problem (53).

**Table 9. NISTmAb Hydrophobic Exposure Temperatures Using DSF**

	10 mg/mL	5 mg/mL	2 mg/mL	1 mg/mL
25 mM histidine, pH 6.0	66.8	66.2	66.2	66.6
25 mM histidine, 150 mM NaCl, pH 6.0	64.2	64.4	64.4	64.8

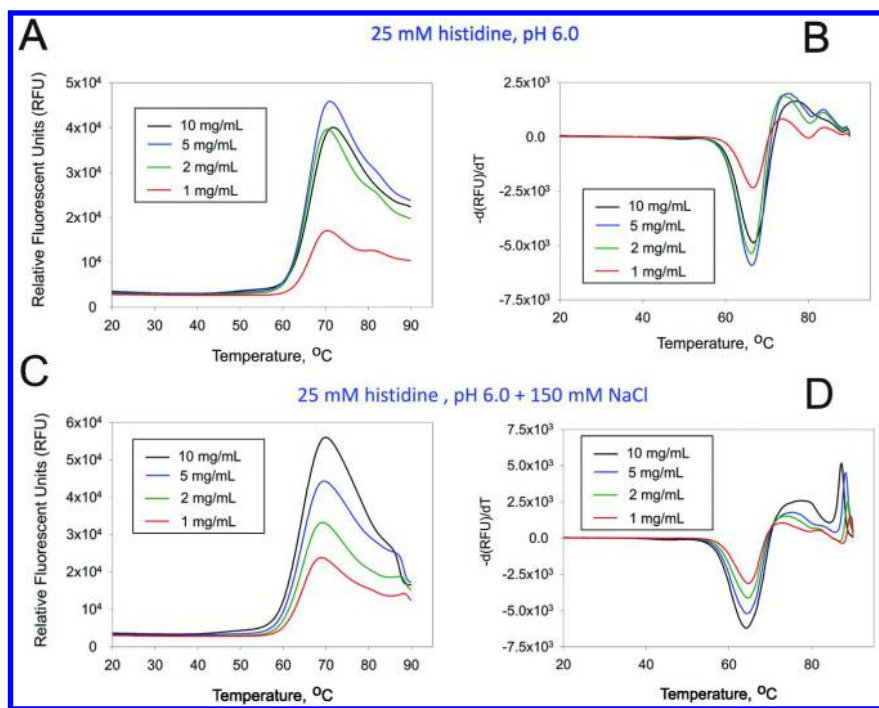


Figure 9. Differential scanning fluorimetry (DSF) profiles of the NISTmAb at different concentrations in 25 mM histidine, pH 6.0 buffer without (A and B) and with (C and D) 150 mM NaCl. The plots for fluorescence versus temperature dependence (A and C) and its first derivative (B and D) are shown. (see color insert)

## Fourier Transform Infrared Spectroscopy

Therapeutic antibodies are typically formulated in liquid at high concentrations, which is particularly important for subcutaneous delivery to avoid high dosage volume (54). When developing high concentration formulations, stress factors such as pH, temperature, and mechanical and oxidative stress, among other stressors play critical roles in determining the physical and chemical stability of biopharmaceuticals. With many biophysical methods, it is difficult to characterize the solution behavior of antibodies at high concentrations. This problem is avoided using FTIR spectroscopy (55). The main advantage of FTIR spectroscopy is the ability to monitor the aggregation that is commonly encountered by proteins during high concentration formulation development (56, 57) both in solid and liquid dosage forms. The technique also is well suited for analyzing  $\beta$ -sheet-rich proteins (e.g., antibodies) because the  $\beta$ -sheet contribution has the highest absorption coefficient (58). More importantly, the absorbance of the  $\beta$ -sheet component of a well-folded antibody can be distinguished from that of an aggregate because of the conversion of the intramolecular  $\beta$ -sheet at  $\sim 1635\text{ cm}^{-1}$  to that of an intermolecular aggregate at  $1620\text{ cm}^{-1}$ . In particular,

attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy is a powerful method for studying protein conformational changes (59) as a function of solution and temperature changes. It is a fast, label-free, non-destructive technique requiring only a few micrograms of sample.

The absorption of infrared radiation excites vibrational transitions in molecules. The infrared region of the electromagnetic spectrum ranges from the visible to the microwave region. By convention, it is expressed as wavenumber, (i.e., as number of waves per centimeter,  $\text{cm}^{-1}$ ) and usually defined for wavenumbers of 10,000 to  $10\text{ cm}^{-1}$ . An IR spectrum of a sample is obtained by measuring the intensity of the IR radiation before and after passage through the sample. The introduction of FTIR spectroscopy removed a main drawback of the classical dispersive IR spectroscopy, which is the slow accumulation speed due to collection of signals from one wavelength after another. In FTIR spectroscopy, an interferogram of two beams produced by a beam splitter and a movable mirror is recorded. It turned out that this interferogram is the Fourier transformation of a spectrum. A second, digital Fourier transformation converts the recorded signal back into a spectrum. For protein analytics, the mid-IR region ( $4000\text{--}1000\text{ cm}^{-1}$ ) is of main interest, as it is dominated by two major types of vibrations: stretching vibrations ( $\nu$ ), which involve changes in the length of a chemical bond, and bending vibrations, which involve changes in the angle of chemical bonds ( $\delta$ —in plane,  $\pi$ —out of plane). The vibration strength thus depends on the strength of the bond (single, double, or triple) and the length of the bond, in other words, the sort of atoms which are connected (C-H, C-O). The FTIR spectrum of a protein is dominated by the vibration of the peptide bond and is sensitive to its chemical environment. Therefore, the hydrogen bonds in the secondary structure influence the wavenumber of the peptide bonds in a protein FTIR spectrum. The vibrational stretching of the C=O bonds of the peptide backbone contribute in the  $1680\text{--}1620\text{ cm}^{-1}$  range and form the amide I band, and the vibrational stretching of the peptide C-N bonds and the vibrational change in the angle of the N-H bonds contribute in the  $1560\text{--}1520\text{ cm}^{-1}$  range and form the amide II band. The complex mix of N-H in-plane bending and C-N stretching vibrations give rise to amide III bands in the  $1330\text{--}1230\text{ cm}^{-1}$  range. In general, amide I ( $1700\text{--}1600\text{ cm}^{-1}$ ), amide II ( $1580\text{--}1480\text{ cm}^{-1}$ ), and amide III ( $1330\text{--}1230\text{ cm}^{-1}$ ) are used to study the conformational changes of proteins in IR. In particular, amide I is the most sensitive, fingerprint region for protein secondary structure.

In this study, the NISTmAb samples were prepared at 10 mg/mL, 25 mg/mL, 50 mg/mL, and 100 mg/mL concentrations in 25 mM histidine, pH 6.0 buffer with and without 150 mM NaCl. The concentrations were measured in triplicate using a Nanodrop UV-Visible spectrophotometer (Nanodrop 8000, Thermo Scientific), which is one of the very few techniques that offers precise measurement of high-concentration protein samples.

Spectra were acquired using a Smart Orbit diamond attenuated total reflection (ATR, attenuated total reflectance) accessory (Thermo Scientific) that was installed in a FTIR spectrophotometer (Thermo Scientific, Nicolet 6700 spectrophotometer). The mercury cadmium telluride (MCT) detector was cooled with liquid nitrogen 30 minutes prior to the analysis. A drop of mAb solution at different concentrations was placed onto the surface of the ATR diamond crystal,

and 128 scans were recorded at 4  $\text{cm}^{-1}$  resolution for background, buffer, and sample collections in the 1800–1400  $\text{cm}^{-1}$  region. The protein spectra (after buffer subtraction) were plotted in OriginPro 8G (OriginLab, Northampton, MA) and smoothed using a 9-point Savitsky-Golay function (four each side). Second derivative plots were then constructed in amide I region (1700–1600  $\text{cm}^{-1}$ ) of the protein spectra that were baseline corrected.

Assignment of amide I peaks of the antibody is shown in Table 10. Absorbance spectra of the antibody at different concentrations and the corresponding second derivative plots in the buffer with no salt (Figure 10A and 10B) and in the buffer with 150 mM salt (Figure 11A and 11B) are shown below.

**Table 10. Assignment of Amide I Frequency Peaks of IgG According to the Literature. Source: (60, 61).**

<i>Frequency (<math>\text{cm}^{-1}</math>)</i>	<i>Assignment</i>
1612	Side chain vibration or intramolecular $\beta$ -sheet
1636, 1689	Intramolecular $\beta$ -sheet
1664	Turns

The spectrum of the native NISTmAb is dominated by the band at 1636  $\text{cm}^{-1}$  in the amide I region, indicative of intramolecular native  $\beta$ -sheets. Second derivative plots (Figure 10B and Figure 11B) showed additional bands. The two bands at 1664  $\text{cm}^{-1}$  and 1689  $\text{cm}^{-1}$  can be assigned to turns and  $\beta$ -sheets, respectively. The mode that is seen at 1612  $\text{cm}^{-1}$  can be assigned to side chain vibration or  $\beta$ -sheets. In the above figures (Figures 10 and 11), the NISTmAb spectra at different concentrations and buffers are overlaid to see if there is any concentration- or salt-induced conformational changes. As clearly seen from the peaks, the spectra of the NISTmAb at all different conditions showed only major components that correspond to native  $\beta$ -sheets and turns. The results imply that the protein neither undergoes secondary structural changes nor shows propensity for aggregation as the concentration increases. Thus, the secondary structure of the NISTmAb can be shown to have native structure at all experimental conditions measured. This observation is a critical factor during high-concentration formulation development.

FTIR also is used widely in monitoring conformational changes in solid state as lyophilization is one of the commonly used methods to enhance stability of protein therapeutics during long-term storage.



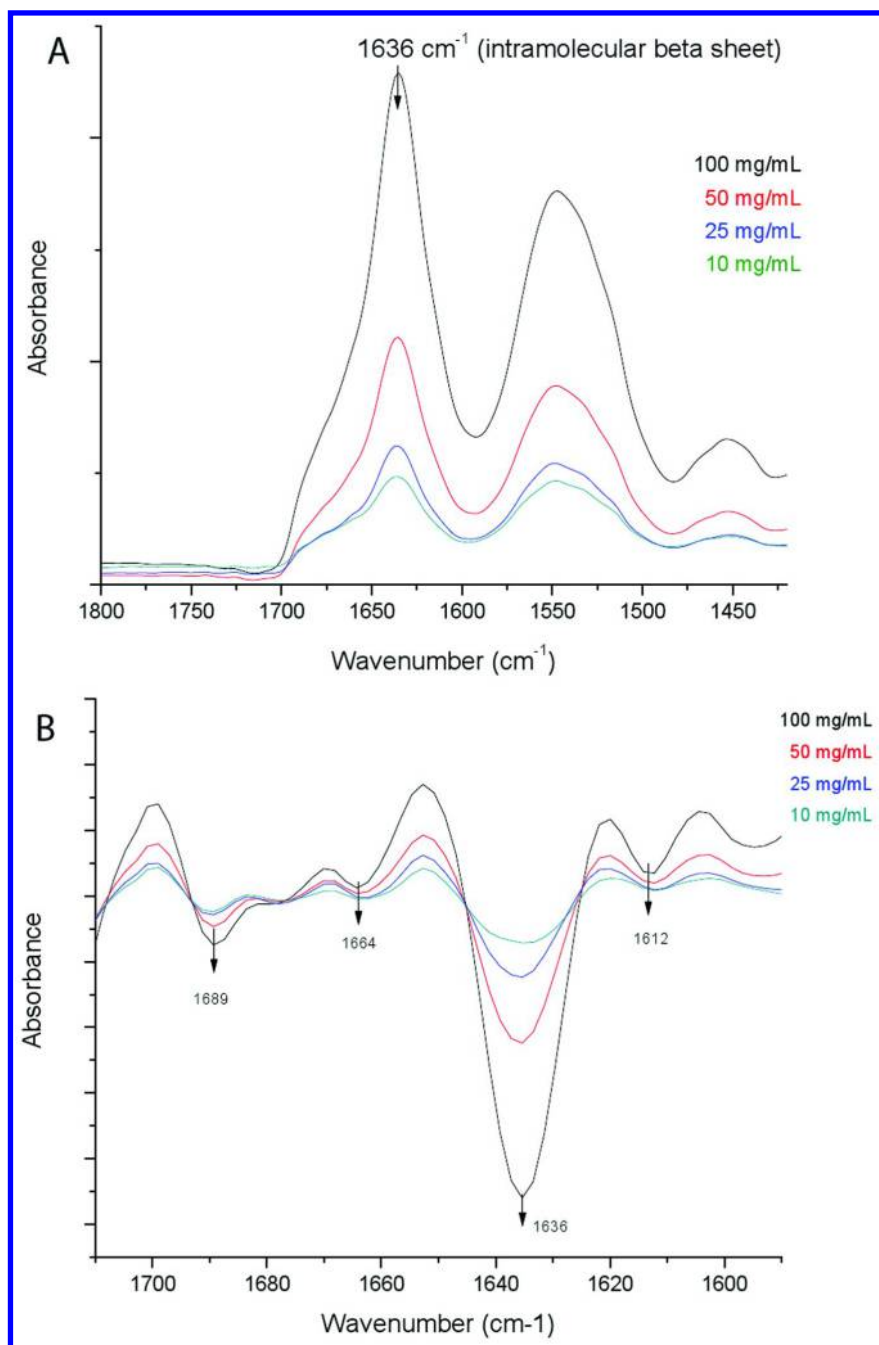


Figure 10. (A) Absorbance spectra of the NISTmAb samples at different concentrations in histidine buffer with no salt. (B) Second derivative plots of the corresponding NISTmAb samples in the amide I region. (see color insert)

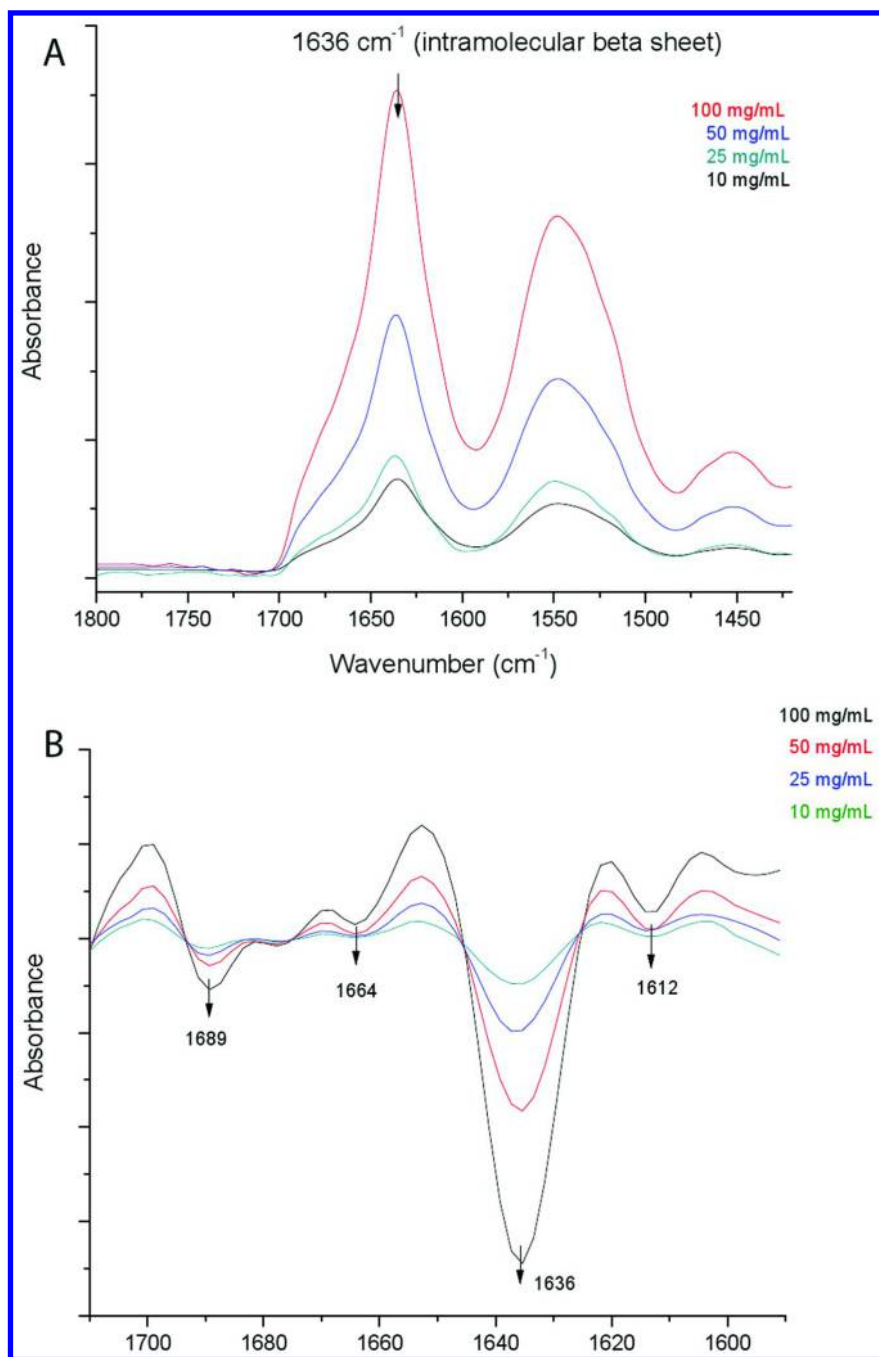


Figure 11. (A). Absorbance spectra of the NISTmAb samples at different concentrations in histidine buffer with 150 mM salt. (B) Second derivative plots of the corresponding NISTmAb samples in the amide I region. (see color insert)

## Circular Dichroism

CD is a spectroscopic technique utilizing circularly polarized light in the ultraviolet (electronic) or infrared (vibrational) region. CD spectroscopy measures the difference between the left-handed and right-handed circularly polarized light absorption of chirally active samples as a function of wavelength. The difference in these absorbances is called “ellipticity.” Proteins are optically active because their basic building blocks are L-amino acid subunits. In addition, protein secondary structures give rise to specific wavelength-dependent ellipticity changes arising from electronic transitions:  $n \rightarrow \pi^*$  and  $\pi \rightarrow \pi^*$ . The peptide bonds adopt a specific, relative chiral orientation depending on the secondary structure, for example,  $\alpha$ -helices (ellipticity signal maximum at 193 nm, minimum at 208 and 222 nm) (62),  $\beta$ -sheets (ellipticity maximum at 195 nm, minimum at 218 nm) (63), or random coil (ellipticity maximum  $\sim$ 215 nm, minimum at 195 nm) (64). The region we use to study the changes in secondary structure is the far-UV, and it extends from 180 to 250 nm.

Furthermore, other UV-active chromophores that affect ellipticity are aromatic residues due to four electronically excited valence states ( $^1L_b$ ,  $^1L_a$ ,  $^1B_b$ , and  $^1B_a$ ) (65). Fixed chiral orientations of the aromatic side chains, such as tryptophan (266 nm), tyrosine (265–290 nm), and phenylalanine ( $\sim$ 265 nm), give rise to ellipticity signals in the near-UV range (250–340 nm) (66). If these side chains are involved in hydrophobic contacts, then they might be sensitive to tertiary structure because changes in the chiral orientation of these moieties could result in alterations to the spectrum. Although histidine and disulfide bonds also weakly contribute to near-UV CD, they are generally not considered.

Lastly, the sensitivity of these two regions are different. The far-UV region is orders of magnitude more sensitive than near-UV. Practically, we acquire data from these two regions in separate experiments in which the concentration and cuvette path length are optimized. CD is sensitive to all chiral molecules in the sample, which typically precludes the use of a histidine buffer system because the histidine-contributed background signal would dominate the overall spectrum. However, the NISTmAb formulation has a low histidine concentration (25 mM) that can be compatible with CD if the cuvette path length is sufficiently short and the protein concentration sufficiently high. In other buffer systems, desalting may be required.

The preparation of mAb samples will generally depend on the wavelength region studied; far-UV is approximately 100-fold more sensitive than near-UV. In the case of the NISTmAb, the sample was used neat at a concentration of 10 mg/mL for both far- and near-UV regions. The far-UV CD was measured using a very short path length cuvette, 0.001 cm, compared to the near-UV, which required a 0.1 cm cuvette.

The data collected were analyzed as follows. A blank spectrum measured using a buffer only sample was subtracted from the raw sample spectrum. The resulting difference spectrum was then offset-corrected such that the maximum wavelength is zero. Finally, the ellipticity data were normalized to mean residue ellipticity (deg  $\text{cm}^2/\text{dmol}$  residues), or MRE, and molar ellipticity (deg  $\text{cm}^2/\text{dmol}$ ) for the far- and near-UV, respectively.

The far-UV CD spectrum (190–250 nm) is consistent with a predominantly  $\beta$ -sheet protein, which was expected for the mAb standard (Figure 12). The key feature is a minimum at 218 nm. Due to the increased overall UV absorbance and corresponding high-tension voltage increase, the range from 190 to 195 nm is less reliable and will depend on the quality of the instrument. The near-UV CD spectrum (250–350 nm) has spectral qualities attributable to the many tryptophan, tyrosine, and phenylalanine amino acids in the protein (Figure 12). There are some minor noise spikes at 272–275 nm.

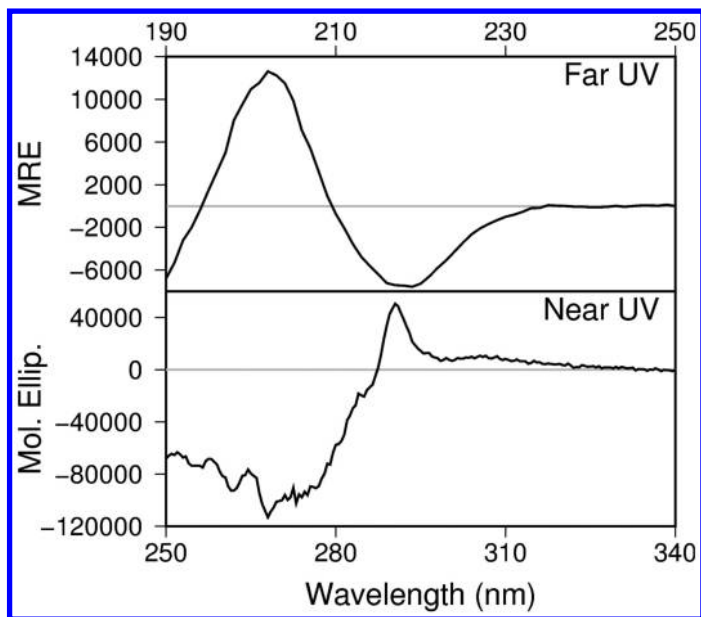


Figure 12. Molecular ellipticity for the proposed IgG1 NISTmAb at the far- and near-UV wavelength range spanning 190 to 340 nm.

## Conclusions

In this collaboration, researchers from different institutions and industries used multiple techniques for characterizing the HOS of mAb in different buffers at varying concentrations. Each methodology informs on various biophysical parameters of a protein therapeutic, the theory and application of which are exemplified herein with optimized results collected on the NISTmAb.

The global size and mass of the NISTmAb were first measured by AUC and light scattering, which rely on orthogonal first principles. Such methodologies are critical to ensuring proper quaternary structure of mAb-based therapeutics as well as evaluating propensity for aggregation during formulation screening. Protein charge also plays a major role in determining the developability of high-concentration liquid formulation. One of the true first-principle methods to determine charge is MCE, and using this technique, the NISTmAb is found to be positively charged at the measured experimental conditions.

Thermal stability is one of the key product attributes that determine the inherent stability of the molecule. DSC remains as an unparalleled technique to assess the thermodynamic stability of proteins in a given buffer condition. As observed in DSC, the NISTmAb exhibited reversal order of unfolding transition that is characteristic of a higher order structural imprint. Also, salt-induced destabilization was observed for all three domains in DSC. By using the DSF technique, temperature-dependent unfolding of the mAb at different lower concentrations was monitored. The DSF profile of the NISTmAb was typical of an IgG profile with a melting temperature between 60 and 70 °C, and the presence of salt slightly decreased the thermal stability of the mAb. The results from both DSC and DSF are consistent, implying that this mAb undergoes similar tertiary conformational changes at increased temperatures.

CD spectroscopy is a well-known, rapid technique for secondary structure and tertiary structure characterization at 10 mg/mL concentration. Both far-UV and near-UV CD spectrum of NISTmAb samples showed spectral qualities attributable to predominant  $\beta$ -sheet protein structure. For high-concentration samples, a complementary tool, ATR-FTIR, was used, and at all tested experimental conditions, mAb samples retained their native secondary structures with or without the addition of sodium chloride.

The techniques described herein are critical to all life cycle stages of a drug product. Initial developability screening of candidate molecules hinges greatly on its biophysical stability. A subsequent characterization of these attributes further informs the higher order structural identity. Such methods also play a critical role in identifying optimum formulations to assure stability and, therefore, ultimately a consistent supply of safe and efficacious product to the patient.

## Acknowledgments

The authors would like to thank John Gabrielson (Amgen Inc.) for his contributions during the preparation of this chapter. The authors also would like to acknowledge Michael R. DeFelippis (Eli Lilly and Company) for careful review of the chapter.

## References

1. Nelson, A. L.; Dhimolea, E.; Reichert, J. M. *Nat. Rev. Drug Discovery* **2010**, *9*, 767–774.
2. Kelley, B. *Biotechnol. Prog.* **2007**, *23*, 995–1008.
3. Carter, P. J. *Nat. Rev. Immunol.* **2006**, *6*, 343–357.
4. Sliwkowski, M. X.; Mellman, I. *Science* **2013**, *341*, 1192–1198.
5. Sievers, E. L.; Senter, P. D. *Annu. Rev. Med.* **2013**, *64*, 15–29.
6. Frankel, S. R.; Baeuerle, P. A. *Curr. Opin. Chem. Biol.* **2013**, *17*, 385–392.
7. Watts, R. J.; Dennis, M. S. *Curr. Opin. Chem. Biol.* **2013**, *17*, 393–399.
8. Schellekens, H.; Jiskoot, W., Immunogenicity of Therapeutic Proteins. In *Pharmaceutical Biotechnology*; Crommelin, D. J. A.; Sindelar, R. D.; Meibohm, B., Eds.; Springer: New York, 2013; pp 133–141.

9. Cole, J. L.; Lary, J. W.; Moody, T.; Laue, T. M. Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **2008**, *84*, 143–179.
10. Zhao, H.; Lomash, S.; Glasser, C.; Mayer, M. L.; Schuck, P. *PLoS one* **2013**, *8*, e83439.
11. Schuck, P. *Biophys. J.* **2000**, *78*, 1606–1619.
12. Fujita, H. *Foundations of Ultracentrifugal Analysis*; Wiley: New York, 1975.
13. Ralston, G. *Introduction to Analytical Ultracentrifugation*; Beckman, 1993.
14. Philo, J. S. *Anal. Biochem.* **2006**, *354*, 238–246.
15. Stafford, W. F.; Braswell, E. H. *Biophys. Chem.* **2004**, *108*, 273–279.
16. Gabrielson, J. P.; Arthur, K. K.; Stoner, M. R.; Winn, B. C.; Kendrick, B. S.; Razinkov, V.; Svitel, J.; Jiang, Y.; Voelker, P. J.; Fernandes, C. A.; Ridgeway, R. *Anal. Biochem.* **2010**, *396*, 231–241.
17. Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G. *J. Pharm. Sci.* **2010**, *99*, 2200–2208.
18. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York, 1992.
19. Blanco, M. A.; Sahin, E.; Li, Y.; Roberts, C. J. *J. Chem. Phys.* **2011**, *134*, 225103.
20. Hiemenz, P.; Rajagopalan, R. *Principles of Colloid and Surface Chemistry*; Marcel Dekker: New York, 1997.
21. McQuarrie, D., *Statistical Mechanics* University Science Books: Sausalito, CA, 2000.
22. Van Holde, K.; Johnson, W.; Ho, P. *Principles of Physical Biochemistry*; Pearson/Prentice Hall: Upper Saddle River, NJ, 2006.
23. Wyatt, P. J. *Anal. Chim. Acta* **1993**, *272*, 1–40.
24. Zimm, B. *J. Chem. Phys.* **1948**, *16*.
25. Wu, H. *Chem. Phys.* **2010**, *367*, 44–47.
26. Zhao, H.; Brown, P. H.; Schuck, P. *Biophys. J.* **2011**, *100*, 2309–2317.
27. Guinier, A. *Ann. Phys.*, *12*, 161–237.
28. Li, Y.; Weiss, W. F. t.; Roberts, C. J. *J. Pharm. Sci.* **2009**, *98*, 3997–4016.
29. Attri, A. K.; Minton, A. P. *Anal. Biochem.* **2005**, *337*, 103–110.
30. Minton, A. P. *Biophys. J.* **2007**, *93*, 1321–1328.
31. Some, D. *Biophys. Rev.* **2013**, *5*, 147–158.
32. Chu, B. *Laser Light Scattering: Basic Principles and Practice*; Academic Press, 1991.
33. Schmitz, K. *An Introduction to Dynamic Light Scattering by Macromolecules*; Academic Press, 1990.
34. Connolly, B. D.; Petry, C.; Yadav, S.; Demeule, B.; Ciaccio, N.; Moore, J. M.; Shire, S. J.; Gokarn, Y. R. *Biophys. J.* **2012**, *103*, 69–78.
35. Yadav, S.; Laue, T. M.; Kalonia, D. S.; Singh, S. N.; Shire, S. J. *Mol. Pharm.* **2012**, *9*, 791–802.
36. Rallison, J. M.; Hinch, E. J. *J. Fluid Mech.* **1986**, *167*, 131–168.
37. Harding, S. E.; Johnson, P. *Biochem. J.* **1985**, *231*, 543–547.
38. Scherer, T. M.; Liu, J.; Shire, S. J.; Minton, A. P. *J. Phys. Chem. B* **2010**, *114*, 12948–12957.

39. Laue, T. M.; Shepard, H. K.; Ridgeway, T. M.; Moody, T. P.; Wilson, T. J. *Meth. Enzymol.* **1998**, *295*, 494–518.
40. Durant, J. A.; Chen, C.; Laue, T. M.; Moody, T. P.; Allison, S. A. *Biophys. Chem.* **2002**, *101-102*, 593–609.
41. Jordon, K. *Real-Time Electrophoretic Mobility in Membrane Confined Electrophoresis*; University of New, 2014.
42. Laue, T. *J. Mol. Recognit.* **2012**, *25*, 165–173.
43. Chang, B. S.; Kendrick, B. S.; Carpenter, J. F. *J. Pharm. Sci.* **1996**, *85*, 1325–1330.
44. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. *J. Biomol. Screening* **2001**, *6*, 429–440.
45. Lo, M. C.; Aulabaugh, A.; Jin, G. X.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. *Anal. Biochem.* **2004**, *332*, 153–159.
46. Ericsson, U. B.; Hallberg, B. M.; Detitta, G. T.; Dekker, N.; Nordlund, P. *Anal. Biochem.* **2006**, *357*, 289–298.
47. Malawski, G. A.; Hillig, R. C.; Monteclaro, F.; Eberspaecher, U.; Schmitz, A. A. P.; Crusius, K.; Huber, M.; Egner, U.; Donner, P.; Muller-Tiemann, B. *Protein Sci.* **2006**, *15*, 2718–2728.
48. He, F.; Hogan, S.; Latypov, R. F.; Narhi, L. O.; Razinkov, V. I. *J. Pharm. Sci.* **2010**, *99*, 1707–1720.
49. Philo, J. S. *Biophys. J.* **1997**, *72*, 435–444.
50. Fesinmeyer, R. M.; Hogan, S.; Saluja, A.; Brych, S. R.; Kras, E.; Narhi, L. O.; Brems, D. N.; Gokarn, Y. R. *Pharm. Res.* **2009**, *26*, 903–913.
51. Wakankar, A. A.; Feeney, M. B.; Rivera, J.; Chen, Y.; Kim, M.; Sharma, V. K.; Wang, Y. *J. Bioconj. Chem.* **2010**, *21*, 1588–1595.
52. Remmele, R. L., Jr.; Bhat, S. D.; Phan, D. H.; Gombotz, W. R. *Biochemistry* **1999**, *38*, 5241–5247.
53. Menzen, T.; Friess, W. *J. Pharm. Sci.* **2013**, *102*, 415–428.
54. Shire, S. J.; Shahrokh, Z.; Liu, J. *J. Pharm. Sci.* **2004**, *93*, 1390–1402.
55. Wartewig, S.; Neubert, R. H. *Adv. Drug Delivery Rev.* **2005**, *57*, 1144–1170.
56. Sarroukh, R.; Goormaghtigh, E.; Ruyschaert, J. M.; Raussens, V. *Biochem. Biophys. Acta* **2013**, *1828*, 2328–2338.
57. Sathya Devi, V.; Coleman, D. R.; Truntzer, J. *Protein J.* **2011**, *30*, 395–403.
58. Martin, I.; Goormaghtigh, E.; Ruyschaert, J. M. *Biochem. Biophys. Acta* **2003**, *1614*, 97–103.
59. Glassford, S. E.; Byrne, B.; Kazarian, S. G. *Biochem. Biophys. Acta* **2013**, *1834*, 2849–2858.
60. Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469–487.
61. Pelton, J. T.; McLean, L. R. *Anal. Biochem.* **2000**, *277*, 167–176.
62. Holzwart, G.; Doty, P. *J. Am. Chem. Soc.* **1965**, *87*, 218.
63. Greenfie, N.; Fasman, G. D. *Biochemistry* **1969**, *8*, 4108.
64. Venyaminov, S. Y.; Baikalov, I. A.; Shen, Z. M.; Wu, C. S. C.; Yang, J. T. *Anal. Biochem.* **1993**, *214*, 17–24.
65. Rogers, D. M.; Hirst, J. D. *Biochemistry* **2004**, *43*, 11092–11102.
66. Pain, R. *Curr. Protoc. Protein Sci.* **2005**, Chapter 7, Unit 7.6.

## Chapter 7

# Developability Assessment of a Proposed NIST Monoclonal Antibody

**Dorina Saro,<sup>\*,1</sup> Audrey Baker,<sup>1</sup> Robert Hepler,<sup>1</sup>  
Stacey Spencer,<sup>1</sup> Rick Bruce,<sup>2</sup> Steven LaBrenz,<sup>2</sup>  
Mark Chiu,<sup>1</sup> Darryl Davis,<sup>1</sup> and Steven E. Lang<sup>1</sup>**

<sup>1</sup>Analytical Discovery Group, Biologics Research,  
Biotechnology Center of Excellence,  
Janssen Research & Development, LLC,  
Spring House, Pennsylvania 19477, United States

<sup>2</sup>Drug Product Development, Parentals and Liquids Formulation,  
Janssen Research & Development, LLC,  
Malvern, Pennsylvania 19355, United States

\*E-mail: dsaro@its.jnj.com

Developability is a term used to describe a process of evaluating potential lead candidate molecules for chances of being successfully developed into drug products. The developability evaluation encompasses a set of biochemical and biophysical assays used to characterize biotherapeutics and provide information on the integrity; purity; and conformational, structural, and chemical stability of the molecules. Candidates that do not present a favorable profile during the developability evaluation are eliminated at early stages in the drug discovery process or are redesigned to avoid potential issues that may occur during the lifecycle of a biotherapeutic. The proposed National Institute of Standards and Technology monoclonal antibody (NISTmAb) was evaluated through a developability process designed for mAb drug candidates. The NISTmAb is not intended to be a therapeutic molecule, and it is used in this developability assessment as a class-specific, representative IgG1 mAb. The goal is to provide a starting point for our development organization for tracking and trending purposes and a data set for senior managers to highlight potential risks. There are no set criteria for each assay and each data point is



treated as part of a holistic data set designed to tell as much about the whole molecule as possible in a 4-week timeframe.

The biochemical and biophysical attributes of the NISTmAb were evaluated under different conditions for stability and integrity. The molecule was concentrated to levels above 100 mg/mL and was stable in phosphate-buffered saline (PBS) buffer when stored at 4 °C up to 1 month as judged by microfluidic electrophoresis, capillary isoelectric focusing (cIEF), size-exclusion chromatography–multi-angle light scattering (SEC-MALS), and dynamic light scattering (DLS) methods. The NISTmAb was shown to be thermally stable by differential scanning calorimetry (DSC). The NISTmAb has minimal self-interactions or association with other polyclonal IgG (poly IgG) molecules as estimated by self-interaction chromatography (SIC) and cross-interaction chromatography (CIC) methods. The NISTmAb is compatible with sera as demonstrated by the Phase Separation assay. The testing of NISTmAb for post-translational modifications included the effects of thermal stress, deamidation induced by high pH, and photo-oxidation. The NISTmAb was resistant to photo-oxidation but showed increased heterogeneity after thermal and high pH stress as evaluated by orthogonal analytical assays. The colloidal stability and the solubility of the NISTmAb in buffers with varying pH values was tested using high-throughput UV and DLS analysis. The molecule was stable under buffer conditions of various pH values that were tested. Based upon the complete data package, the NISTmAb has a favorable developability profile and the changes observed in the molecule after stress conditions do not represent risks that cannot be mitigated with the right formulation and control strategy at later stages in the drug discovery and development process.

## Introduction

Developability is a process of evaluating the potential of manufacturing human monoclonal antibodies (mAbs) into biological therapeutic drugs in liquid or lyophilized formulation by confirming their biochemical and biophysical properties at release and under forced degradation conditions (*1*). These properties often include but are not limited to solubility, thermal stability, structural and conformational integrity, minimal association with itself and other molecules, and resistance to stress-inducing agents/factors during production and storage. Developability assessment is seen as a critical component of the discovery or pre-clinical stages of drug development. The typical biochemical and biophysical assays that are employed to determine the aforementioned properties of a human mAb comprise “concentratability” or solubility, microfluidic electrophoresis,

capillary isoelectric focusing (cIEF), size-exclusion chromatography–multi-angle light scattering (SEC-MALS), dynamic light scattering (DLS), cross-interaction chromatography (CIC), self-interaction chromatography (SIC), differential scanning calorimetry (DSC), and Phase Separation in serum analysis. The developability assessment is an integral part of discovery programs to help to ensure that molecules entering clinical trials have biochemical and biophysical properties that are amenable to upstream and downstream processing, formulation, and storage under acceptable conditions. The heterogeneity or lack of a suitable profile in any of the properties is then either related to a change in functional activity, clearance, or safety (*I*). Testing at early drug discovery stages for stability and integrity is important in reducing overall costs of the drug development process by reducing the failure rate due to nonclinical outcomes or less than optimal clinical outcomes as a result of product-related impurities that cause a loss of efficacy or safety. Multiple candidates are typically assessed at this stage with a small amount of material required.

Developability is a term used to encompass several aspects of pharmaceutical production as it relates to the chances of arriving at a successfully marketed therapeutic, discovery activities, early expression and purification (small-scale production), protein engineering, and finally large-scale good manufacturing practice (GMP) protein product (manufacturability). As such, it involves determining key process steps or attributes (process knowledge) and relating those to key protein attributes or outcomes (product knowledge). Early in the life cycle, one cannot in general predict all of the critical quality attributes of a protein. Based upon internal and external knowledge, companies make strategic decisions on at what stage the developability process takes place and what the focus of the outcome is. For this reason, each biopharmaceutical company pursues an assessment of the attributes of a potential protein therapeutic in relation to the necessary processing steps to produce a drug product differently. Common aspects of protein therapeutic stability assessments are *in silico* risk assessment, protein engineering of high-risk attributes, platform assessment (process and assay fit), and stability assessment. Each of these steps can be performed during both screening and selection phases.

Many of the techniques or assays used in the developability process have been highlighted in other chapters of this volume. A key aspect of the process is that the disconnected temporal and spatial aspect of a retrospective review is removed and, therefore, a consistent chain of custody allows one to make inferences that otherwise would be open to interpretation. For instance, if two orthogonal techniques that both assess aggregation provide results that are deemed not comparable, a logical conclusion would be that the techniques themselves are the root cause. All too often, however, root cause analysis arrives at the conclusion that storage or sample treatment is different and, therefore, the sample itself is different.

One often-overlooked critical component of developability is the work flow and how it is organized in order to maximize the information from which key decisions will be made. As an example, one could organize the work flow and results around the attributes being tested. Other possibilities are mimicking the needs of subfunctions (quality control [QC], release, stability indicating, primary

structure, and modifications) or phases (early-phase needs leading into late-phase testing).

National Institute of Standards and Technology (NIST) mAb was evaluated through the same developability process designed to evaluate clinical candidates although the molecule itself is not planned to be developed into a therapeutic molecule. The NISTmAb serves as an isotype-specific molecule to highlight a developability assessment protocol. For this assessment, the work was organized around general attributes and was focused on providing information that could be used downstream (i.e., at later phases) of the discovery process. The assays used in this work test the molecule for purity, homogeneity, conformational stability, and resistance to stress. The ability of the molecule to interact with itself, as well as poly IgG molecules and sera, also were evaluated. The solubility of the NISTmAb also was addressed in this work. Because there was no target product profile (TPP) to work from, it should be noted up front that the data are treated as data without a predetermined limit or criteria. There was no assessment of criteria or analytical outcome for the NIST reference material (RM). In short, there is no acceptable or unacceptable limit of modification or assay shift. It depends on where the modification is; what the modification is; how it affects other outcomes such as binding; and, ultimately, on the risk each program is willing to accept. Some programs are willing to take more risk; therefore, the outcome is looked at holistically across all assays, and a risk assessment can be produced for each molecule. Just as importantly, the desired attributes of both the product and the process are tied to the TPP, which ultimately dictates how drug product will be administered (intravenously or subcutaneously) and packaged (high concentration in a vial or self-injectable syringe). Specifically in the case of the NISTmAb, there were no functional assays to compare alongside the forced degradation conditions. For this reason, it would be hard to say whether a certain change in, for example, deamidation is relevant to activity.

Overall, the data indicate that the NISTmAb is stable and can be concentrated at levels above 100 mg/mL in phosphate-buffered saline (PBS) buffer and should be able to be processed and stored in an appropriate formulation with minor deamidation risk.

## Materials and Methods

NIST candidate RM 8670 (humanized mAb, IgG1 $\kappa$ , lot # 3F1b) material was dialyzed extensively in 1x Dulbecco's phosphate-buffered saline (DPBS) (2.7 mM KCl, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, 136.9 mM NaCl, 8.9 mM Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, no Ca<sup>2+</sup>/Mg<sup>2+</sup>, purchased from Invitrogen [Gibco, Grand Island, NY]) buffer to remove any trace of formulation buffer prior to analysis.

### Solubility and Stability

The NIST sample was concentrated using a Millipore concentrator with 30K molecular weight cutoff (MWCO) membrane and centrifuged at speeds of 4500 revolutions per minute (RPM) at 23 °C. The sample was checked intermittently

by visual inspection for any signs of precipitation until the desired volume was reached to obtain concentration > 100 mg/mL. Concentration was determined at this step by diluting a sample 100-fold in PBS. Triplicate measurements were performed. Concentration was determined on an Agilent 8453 diode array spectrophotometer using the extinction coefficient 1.4 calculated from the primary amino acid sequence at 280 nm. After the concentration step, the sample was stored overnight at 4 °C. The following morning, the concentrated sample was placed at room temperature for equilibration. The sample was visually inspected for any signs of precipitation, and a fraction was removed for analytical characterization of the NIST candidate by micro-sodium dodecyl sulfate (SDS), cIEF, DLS, SEC-MALS, and serum compatibility. Stability of the sample was analyzed at 4 °C after incubating for different time periods (0, 1, and 4 weeks). At each time point, the sample were visually inspected for any signs of apparent aggregation and characterized by micro-SDS, cIEF, SEC-MALS, and DLS for any sample variation, including soluble or insoluble aggregation.

### Microfluidic Electrophoresis

Microfluidic electrophoresis Caliper LabChip GXII (PerkinElmer, Waltham, MA) was used following the manufacturer's instructions. For the reducing conditions, a 5  $\mu$ L aliquot of NIST sample diluted to 1 to 2 mg/mL was added to a well in a 96-well plate. A 30  $\mu$ L portion of reducing sample buffer (10x NuPAGE sample reducing buffer diluted 1:10 in HT Protein Express Sample Buffer) was added to each sample. The sample was heated at 70 °C for 15 minutes on a Caliper LabChip GXII (PerkinElmer, Waltham, MA). A 4  $\mu$ L portion of 100 mg/mL iodoacetamide (in 1 M Tris) solution was added to each sample, and then 71  $\mu$ L of distilled water was added for a final sample volume of 110  $\mu$ L.

For the nonreducing conditions, a 5  $\mu$ L protein sample at 1 to 2 mg/mL was added to a well in a 96-well plate. A 30  $\mu$ L portion of HT Protein Express Sample Buffer was added to each sample. A 4  $\mu$ L aliquot of 100 mg/mL iodoacetamide (in 1 M Tris) solution was added to each sample. The sample was heated at 70 °C for 15 minutes. A 71  $\mu$ L portion of distilled water was added to each sample for a final sample volume of 110  $\mu$ L.

A 12  $\mu$ L sample of the proprietary ladder (PerkinElmer, Waltham, MA) was heated at 70 °C for 15 minutes. A 120  $\mu$ L portion of distilled water was then added to the ladder. The analysis proceeded as according to the HT Protein Express 200 assay, using a Protein Express chip prepared according to manufacturer's instructions (PerkinElmer, Waltham, MA).

### Capillary Isoelectric Focusing

The NIST sample and an internal control mAb were characterized on an iCE280 Bioanalyzer (Protein Simple, San Jose, CA). The NIST sample was analyzed at 0.25 mg/mL in 0.35% methyl cellulose, 4% pH 3–10 Pharmalytes, and 4% pH 5–8 Pharmalytes (GE Healthcare). Calibration markers at pI 5.1 and pI 9.5 (Sigma Aldrich) were included at 1% (v/v). The samples were analyzed with a 1 minute prefocusing time at 1500V and 9 minute focusing time at 3000V

at 10 °C. Results were integrated using ChromPerfect software. The pI of the samples was assigned based on the intensity of the highest peak.

### Size-Exclusion Chromatography–Multi-Angle Light Scattering

Injections of 100 µg to 200 µg of NIST sample were made onto a pre-equilibrated Tosoh TSKgel BioAssist G3SWxl column, 7.8 mm x 30 cm, 5 µm, PEEK housing (Tosoh) column in 1x DPBS (Gibco, Grand Island, NY) buffer. The protein species were eluted at a flow rate of 0.35 mL/min at room temperature in 1x DPBS buffer. The absorbance at 280 nm was monitored using an Agilent MWD detector. The refractive index signal at 685 nm was monitored using an OptiLab rEX differential refractometer (Wyatt Technology Inc., Santa Barbara, CA), and light scattering signals were monitored at 690 nm using a DAWN-EOS MALS detector (Wyatt Technology Inc., Santa Barbara, CA). Two injections of 100 µg of bovine serum albumin (BSA) were similarly analyzed to serve as a system control for MALS signal normalization. Molecular weight calculations were performed with Astra software (Wyatt Technology Inc., Santa Barbara, CA) using a generic protein  $dn/dc$  value of 0.185 mL/g (2).

### Dynamic Light Scattering

DLS, also known as photon correlation spectroscopy (PCS), provides information about the size and shape of molecules in solution. To differentiate the presence (or absence) of various sized species, the hydrodynamic radii ( $R_h$ ) and percent polydispersity (Pd) of the samples were determined. Large percent Pd values are indicative of molecular heterogeneity. Particle sizes and size distributions of all samples were determined using a DynaPro Plate Reader (Wyatt Technology Inc., Santa Barbara, CA) at 23 °C. A 30 µL portion of each sample at a nominal concentration of 1 mg/mL was placed in Corning® 384-well black polystyrene plates with clear flat bottoms (CLS3540). Triplicate measurements were performed for the NIST sample, with each measurement consisting of 20 runs. The refractive index of 1.333 at 589 nm for PBS buffer at 20 °C was used (a standard value embedded in the software by the manufacturer). The method of cumulants was used to analyze the data.

### Differential Scanning Calorimetry

DSC experiments were performed using a MicroCal Auto VP-capillary DSC system (GE Healthcare) in which temperature differences between the reference and sample cell were continuously measured and converted to power units. Samples were heated from 25 °C to 95 °C at a rate of 1 °C/min. A prescan time of 10 minutes and a filtering period of 10 seconds were used for each run. DSC measurements were made at a NISTmAb concentration of approximately 0.5 mg/mL in 1x DPBS buffer in duplicates. Analysis of the resulting data was performed using MicroCal Origin 7 software.

## Phase Separation in Serum

The experimental setup for this assay was the same as previously described (3). Briefly, the NIST sample at 146.5 mg/mL in PBS buffer was diluted 1:1 with human, mouse, and cynomolgus monkey (cyno) sera purchased from Bioreclamation (Hicksville, NY). PBS buffer and a proprietary mAb control at 50 mg/mL concentration were used in the same setup as described (3). Two  $\mu\text{L}$  of the mAb/serum mixture were pipetted into 4  $\mu\text{L}$  wells of an HR8-142 96-Well Corning 3554 Crystal EX microplate (Hampton Research). Wells were imaged using a light microscope (Optical Apparatus Company) fitted with a Nikon SMZ 100 lens at 10x magnification.

## Cross-Interaction Chromatography

Purified polyclonal human IgG was obtained from Sigma and resuspended in 0.1 M  $\text{NaHCO}_3$ , pH 8.1, 0.5 M NaCl. The protein was then coupled to a 1 mL HiTrap NHS-activated column (GE Healthcare # 17-0716-01) for 5 hours at room temperature. Coupling efficiency was estimated by measuring the protein concentration before and after the coupling step. Remaining reactive groups on the resin were blocked overnight at 2 to 8  $^\circ\text{C}$  in 0.5 M ethanolamine, pH 8.1, 0.5 M NaCl. A control column was prepared by blocking all reactive groups with ethanolamine, as described above.

Separations were performed on both IgG-coupled and control columns. NISTmAb samples were prepared by diluting to 0.1 mg/mL in 1x DPBS. A 20  $\mu\text{L}$  aliquot of each sample and control were injected onto both columns at a flow rate of 0.1 mL/min. Peaks were visualized by monitoring the absorbance at 215 nm. Retention times were determined using the HPLC system software.

A correction factor was determined for each protein sample using the retention times of the protein and acetone on the non-IgG-containing column. The retention factor on the IgG-coupled column was then adjusted with the correction factor to determine the chromatographic retention factor ( $k'$ ) using the relationship given below, where  $V_R$  is the elution volume of the sample on the protein coupled column,  $V_0$  is the elution volume from the control column,  $t_R$  is the retention time on the protein-coupled column, and  $t_M$  is the retention time on the control column. The retention time is defined as the elution time of the peak maximum (4):

$$k' = \frac{V_R - V_0}{V_0} = \frac{t_R - t_M}{t_M}$$

## Self-Interaction Chromatography

The NIST sample was dialyzed overnight in coupling buffer (0.1 M  $\text{NaHCO}_3$ , pH 8.1, 0.5 M NaCl). A 1 mg sample of NISTmAb was coupled to 0.1 mL of resin (0.025 g dry Toyopearl AF-tresyl-650M chromatography resin, Tosoh Biosciences, pre-swollen with deionized water and washed in coupling buffer). The coupling reaction was allowed to proceed for 5 hours at room temperature before the resin was washed twice with 1 mL of coupling buffer. The remaining

tresyl groups were capped by incubating the resin in 0.5 M ethanolamine, pH 8.1, 0.5 M NaCl, overnight at 2 to 8 °C. The coupled resin was packed in an 18 cm chromatography column and equilibrated in PBS using a flow rate of 60 µL/min until the bed height stabilized. The integrity of the packing procedure was verified by injecting a 20 µL sample of 10% acetone (aqueous). Experiments were carried out on an Agilent HPLC equipped with an autosampler. NISTmAb samples were prepared by diluting to 0.1 mg/mL in PBS. Each sample (and the acetone control) was injected at 20 µL onto the column and eluted at a flow rate of 60 µL/min in PBS. Peaks were visualized by monitoring the absorbance at 215 nm. Retention times were determined using Chemstation software. The chromatographic retention factor,  $k'$ , was calculated as described previously in the CIC section above. The second virial coefficient ( $B_{22}$ ) was determined using the relationships given below as described by Payne et al. (5):

$$B_{22} = \frac{6.0225 \times 10^{23}}{MW^2} \left( BHS - \frac{k'}{\text{phi}2 * \text{rho}} \right)$$

where MW is the molecular weight of the protein, phi2 is the phase ratio or the available surface area per stationary phase volume calculated from the properties of the matrix based on a published method (6), rho is the number of peptide molecules per unit surface area, and BHS is the calculated exclusion volume:

$$BHS = 4 * V_m * 1 \times 10^{-24} \left( \text{units in } \frac{\text{mL}}{\text{molecule}} \right)$$

$$V_m = 1.205 * MW - 1665$$

## Post-Translational Modification Risk Assessment

Post-translational modification (PTM) risk assessment of NISTmAb candidate was performed by determining the susceptibility of methionine and tryptophan to photo-oxidative stress and asparagine residues to deamidative stress at elevated pH. Stressed samples were submitted for analysis by microfluidic electrophoresis, cIEF, SEC-MALS and DLS.

*Photo-oxidative stress* consisted of exposure to white light for 5 days and UV light for 1 day. One sample of NIST in PBS at about 1 mg/mL was placed in a validated Photostability Chamber (Caron, Model 6545-1) at 25 °C. Efficiency of photo-exposure was determined by performing quinine chemical actinometry as recommended by ICH guidelines (7). Two samples containing 2% quinine, one wrapped with aluminum foil and the other unwrapped, were placed in the photo-chamber.

For *deamidative stress at high pH*, a sample of NISTmAb was dialyzed into pH 8.9 phosphate buffer and incubated at 37 °C for 5 days. After incubation, the sample was dialyzed back into PBS, pH 7.4. An unstressed sample at 4 °C in PBS was used as a control for both photo-oxidation and deamidation stress samples.

*Thermal Stress.* In order to check susceptibility of the NISTmAb to long-term storage, the sample at 10 mg/mL concentration was subjected to thermal stress at 40 °C for 2 weeks in PBS, pH 7.4.

### **Pre-Formulation Buffer Screen of the NIST Candidate Using High-Throughput Formulation with UV and DLS Analysis**

The NIST sample was desalted using Zeba columns (Thermo Scientific, 40K MWCO) and buffer exchanged to final buffer (10 mM histidine, pH 6.3) prior to the screen. The 96-well plate used for data collection was prepared with a Freeslate CM3 High-Throughput Biologics Formulation System. A NIST stock solution of 2 mg/mL and a series of 8 buffers (200 mM each, with pH varying from 3.5 to 8.4) were used. The buffers used for the screen were: 200 mM citrate (pH 3.5), 200 mM acetate (pH 4.2), 200 mM acetate (pH 4.9), 200 mM histidine (pH 5.6), 200 mM histidine (pH 6.3), 200 mM phosphate (pH 7.0), 200 mM phosphate (pH 7.7), and 200 mM Tris/acetate (pH 8.4). The final NIST concentration in the well was 0.2 mg/mL, and each buffer was 50 mM. A 100  $\mu$ L paraffin oil aliquot was placed over each sample to avoid evaporation during the run. Absorbance at 350 nm was measured before and after the DLS temperature ramp (TRamp) using a Molecular Devices SpectraMax plate reader instrument. The DLS was measured with a DynaPro (Wyatt Technology Inc., Santa Barbara, CA) plate reader instrument. The scans were recorded as 17 connected isotherms starting at 26 °C and continuing step-wise through successive isotherms, ending at 70 °C.

## **Results and Discussion**

### **Concentratability**

Therapeutic antibody molecules are often administered in humans at high concentration due to large dose requirements for treatment and poor tolerance of volumes administered in excess of 1.5 mL (8, 9). To test the ability of NIST material to reach high concentration for proper administration, the sample was dialyzed in PBS buffer to remove all traces of formulation buffer and concentrated from 10 mg/mL initially (the nominal concentration) to 146.5 mg/mL (> 100 mg/mL). The rationale for using PBS buffer for all the mAb samples during the developability assessment is based on the necessity to create the same testing conditions for all the molecules that go through the process in order to make the comparisons and conclusions more meaningful. Table 1 summarizes the concentration of NIST material. The sample was visibly clear and can be concentrated to > 100 mg/mL in PBS without the formation of visible protein aggregates or precipitants. 100 mg/mL was chosen as a target because for subcutaneous administration, concentrations of drug product as high as this could be expected. The concentrated sample was kept at 4 °C for the duration of the assessment. At the specified time points, the concentrated sample was taken out and allowed to equilibrate at room temperature, and the concentration was measured for the diluted aliquots in triplicates. The stability of the NISTmAb was determined in PBS after incubation at 4 °C for 1 day (week 0), 1 week,



and 4 weeks. Except for the sample used in the Phase Separation assay, all of the samples that were used for further characterization were diluted from the concentrate to levels of 0.5 to 1 mg/mL in PBS buffer with serum that comes from the concentrated material.

**Table 1. Summary of Protein Concentration of NISTmAb in PBS<sup>a</sup>**

Lot ID	Nominal Concentration (mg/mL)	Final Concentration (mg/mL)			
		Upon Concentration	Week 0	Week 1	Week 4
3F1b	10	146.5 ± 0.9	151 ± 8.2	150.9 ± 1.8	152.0 ± 5.1

<sup>a</sup> The values in the Table represent average values and standard deviations from triplicate measurements. NISTmAb: National Institute of Standards and Technology monoclonal antibody, PBS: phosphate-buffered saline.

## Microfluidic Electrophoresis

Electrophoresis is a widely used analytical method used to determine a macromolecule's size and purity based on the principles of separation on a gel matrix (10). Technological advancements of the electrophoresis methods using capillary and microfluidic devices have allowed for increased sensitivity and throughput of the method (11). We have used a microfluidic-based electrophoresis system for determining the size and the purity levels of the NISTmAb. Briefly, the method is based on the separation of the samples in a microfluidic chip that mimics the gel environment (LabChip GXII platform). The protein is fluorescently labelled while it is running in the chip, allowing for increased sensitivity of the detection. The method can be run under native, nonreducing conditions and reducing conditions. For a mAb sample under native conditions, a single band is detected, corresponding to the molecular weight (MW) of the intact molecule. Under reducing conditions, both the heavy chain (HC) and light chain (LC) are detected. The microfluidic method allows for less sensitivity compared to capillary electrophoresis method but offers higher throughput (12).

The size and purity of the NIST sample was verified by microfluidic electrophoresis analysis of the material before concentration and of the concentrated material at different time points in a 1 month period. Figure 1 shows the microfluidic electrophoresis profiles of NIST material after incubation at 4 °C for week 0, week 1m and week 4 at > 100 mg/mL. The profile of NISTmAb was typical of that of a mAb: a single band in nonreducing condition, corresponding to the MW of an intact mAb, and two bands in reducing conditions, corresponding to the MW of HC and LC. Similar bands were observed in the NIST sample microfluidic electrophoresis profiles performed before the concentration step (Table 2). The percent mAb or purity results from this assay are well within the limits needed for clinical work. For phase 1 studies, purity above 90% is a typical cutoff.

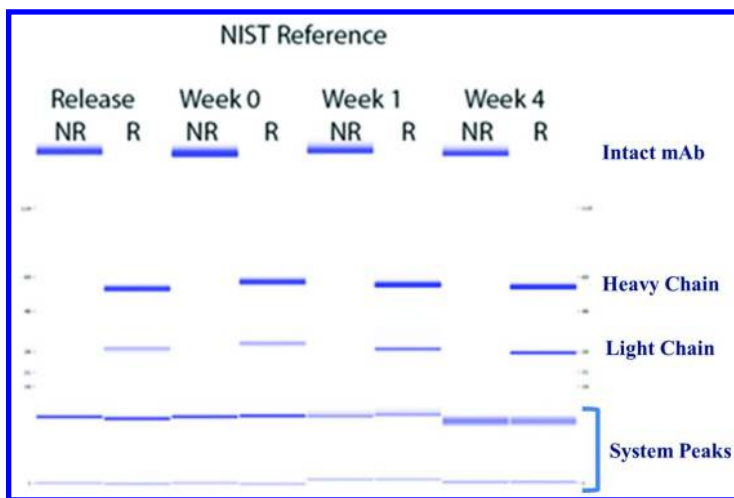


Figure 1. Microfluidic electrophoresis profiles of National Institute of Standards and Technology monoclonal antibody (NISTmAb) material before and after incubation at high concentration at 4 °C. Release, Week 0, Week 1, and Week 4 were analyzed at 1 mg/mL concentration. NR: nonreducing, R: reducing conditions.

**Table 2. Summary of Microfluidic Electrophoresis Quantification of NISTmAb after Incubation at High Concentration at 4 °C<sup>a</sup>**

Time Point	% mAb NR	% mAb R
Release	99.4	99.4
Week 0	99.6	99.6
Week 1	99.8	99.7
Week 4	98.9	99.3

<sup>a</sup> The values in the Table for the reducing conditions (R) represent the sum of the values for the light chain and heavy chain. NISTmAb: National Institute of Standards and Technology monoclonal antibody, NR: non-reducing, R: reducing conditions.

### Capillary Isoelectric Focusing

cIEF is the method of choice used to determine the charge distribution of biological molecules, especially of the mAb samples (13). The method allows detection of different glycosylation variants of the mAbs (14, 15), deamidation, engineered sites in the molecules (16), and any other modifications that affect the molecule's charge (17, 18).

The charge distribution of the NIST sample was verified by cIEF analysis of the material. Figure 2 shows the cIEF profile of NISTmAb. The cIEF profile of NIST indicates that the experimental pI of the material is 9.17, representing

71% of the peak area. The predicted pI value of the molecule based on the primary amino acid sequence is calculated to be 8.9 (19), but these calculations do not take into account the potential PTMs in the molecule that result in a charge change. Importantly, the NIST material did not have significant changes upon concentration and storage at 4 °C up to a 1-month period, as indicated by the values of experimental pI detected at different time points (weeks 0, 1, and 4). A slight increase in sample heterogeneity was observed after a 1 month incubation at 4 °C, as indicated by a decrease of the main peak area.

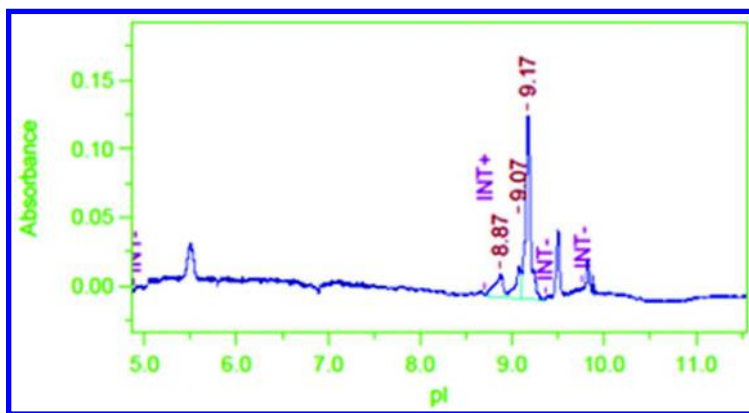


Figure 2. Capillary isoelectric focusing (cIEF) profile of National Institute of Standards and Technology monoclonal antibody (NISTmAb) at 1 mg/mL.

### Size-Exclusion Chromatography–Multi-Angle Light Scattering

SEC is the method of choice used to determine sample size distribution based on separation of species in a gel matrix, presence of aggregates, fragmentation, and other modifications that result in size changes in molecules. The method is quantitative and, typically, the absorbance is used to determine the presence of size variants in the sample.

We have used a SEC system coupled to Multi-Angle Light Scattering detector (SEC-MALS) to determine the size and the distribution of species present in the NIST sample. The NIST samples were analyzed before and after concentration at different time points, up to a 1 month period. Figure 3 and Table 3 show the levels of main-monomeric species present in the sample after incubation at high concentration. SEC-MALS profiles of all of the NISTmAb samples exhibited > 99% of the main peak (monomer) of the expected size of 150 kDa. Moreover, the incubation at 4 °C for up to a 1 month period did not have any significant effect on the size distribution of the samples as indicated by the percent monomeric peak present in all of them. These results are well within acceptable typical acceptable levels for monomeric content. Although it does depend on product and phase, greater than 99% purity would be acceptable throughout most of the development lifecycle.

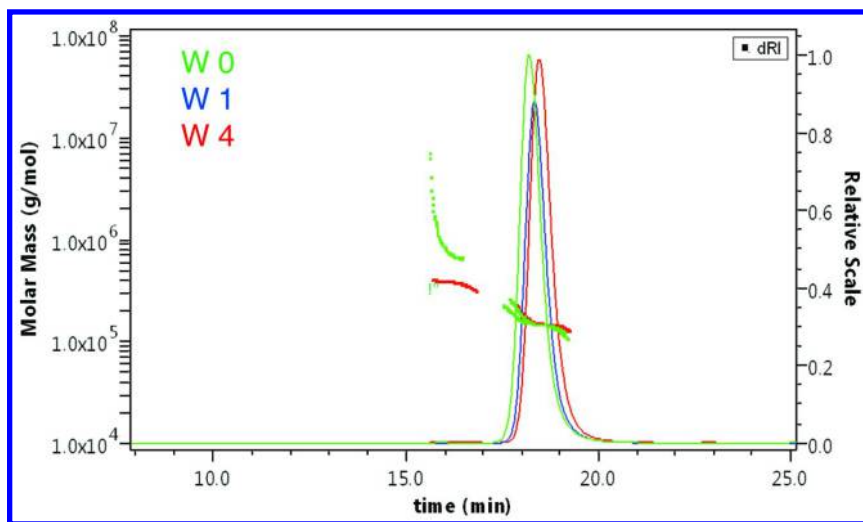


Figure 3. Size-exclusion chromatography–multi-angle light scattering (SEC-MALS) profiles of National Institute of Standards and Technology monoclonal antibody (NISTmAb) after incubation at high concentration at 4 °C. Chromatograms for Week 0 (green), Week 1 (blue), and Week 4 (red) samples are shown.

**Table 3. SEC-MALS Data of NISTmAb after Incubation at High Concentration at 4 °C<sup>a</sup>**

<i>Time Point</i>	<i>Monomer (%)</i>
Release	100
Week 0	100
Week 1	99.8
Week 4	99.4

<sup>a</sup> Data show the amount of main species present in the release samples and stability samples at high concentration after week 0, week 1, and week 4 incubation at 4 °C. NISTmAb: National Institute of Standards and Technology monoclonal antibody, SEC-MALS: size-exclusion chromatography–multi-angle light scattering.

### Dynamic Light Scattering

DLS is a method used to detect the particle sizes in solution based on the scattering of light. The scattering intensity is a function of the particle size; therefore, the method is very sensitive to the presence of aggregates in the sample. The development of plate readers and the relatively low sample requirements for the DLS measurements have increased the use and the throughput of the method in pharmaceutical drug discovery processes. The scattering intensity change is measured as function of time and used to determine the diffusion coefficient,

from which the hydrodynamic radius,  $R_h$ , is calculated. DLS results are highly sensitive to the presence of aggregates due to a strong dependence of the light scattering intensity on the hydrodynamic radius (20, 21).

Soluble aggregates and subvisible particles in the release and stability samples were determined by DLS after incubation at 4 °C for up to a 1 month period. Table 4 shows the  $R_h$ -average, percent Pd values, and percentage mass of the NIST candidate at different time points. All of the samples yielded a peak with a radius of 5.2 to 5.4 nm, with percent mass values of 100%. These results are consistent with the presence of monomeric IgG species with the expected MW range for a ~150 kDa molecule. The percent Pd was determined to be less than 8.4%, indicative of monodispersed, homogeneous species. Typical values for mAbs would be  $R_h$  values at approximately 7 and percent Pd values less than 15%.

**Table 4. DLS Data of NISTmAb after Incubation at 4 °C at High Concentration<sup>a</sup>**

<i>Time Point</i>	<i>R<sub>h</sub> (nm)</i>	<i>Pd (%)</i>	<i>Mass (%)</i>
Control	5.4	7.9	100.0
Week 0	5.4	8.4	100.0
Week 1	5.3	7.0	100.0
Week 4	5.2	4.7	100.0

<sup>a</sup> Values represent averages of triplicate measurements at 1 mg/mL concentration. DLS: dynamic light scattering, NISTmAb: National Institute of Standards and Technology monoclonal antibody, Pd: polydispersity,  $R_h$ : hydrodynamic radius.

### Differential Scanning Calorimetry

DSC is the standard method used to determine the thermal stability of therapeutic molecules. The method is quantitative (22), and the automation has helped to increase the method throughput (23). In general, it has been observed that the low-temperature thermal transitions determined by DSC correlate with long storage conditions for the therapeutic molecules (24). For multidomain molecules like mAbs, several transitions are observed during a DSC experiment, corresponding to the unfolding events of each of the domains. Typically in an IgG1 molecule, the lowest transition belongs to the C<sub>H2</sub> domain and the highest transition to the C<sub>H3</sub> domain. The Fab domain of the IgG1 molecule shows temperature transitions that vary depending on the sequence of the variable region (25). Understanding the thermal stability of the Fab region is important in the candidate screening process because it has been suggested that stabilization of the thermal unfolding of the Fab domain through formulation can be more effective in improving the long storage stability of therapeutic molecules than the increase of the first unfolding transition (26).

The thermal stability of the proposed NIST candidate mAb samples was determined by DSC thermal unfolding of the sample from 25 to 95 °C in PBS buffer. Representative DSC thermograms indicating the transitions are shown in

Figure 4 for duplicate runs. The data were analyzed using the “*non two state*” function in Origin software and are summarized in Table 5. The NIST sample shows three main thermal transitions assigned as the C<sub>H2</sub> domain at 71.2 °C, the C<sub>H3</sub> domain at 84.1 °C, and the Fab domain at 88.9 °C. An analysis of 17 IgG1 mAbs by DSC shows that the Fab region exhibited thermal unfolding transitions with midpoints (T[M]s) varying from 57 to 82 °C (25). Compared to this panel of IgG1 molecules, the NIST Fab is more stable (88.9 °C); therefore, it can be concluded that the molecule has a favorable developability profile based on the DSC data.

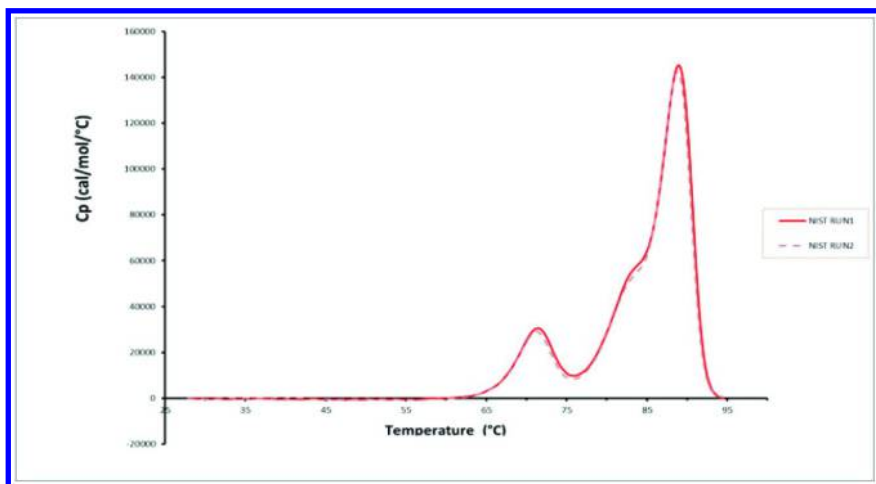


Figure 4. Thermograms of the thermal unfolding of National Institute of Standards and Technology monoclonal antibody (NISTmAb). Results from duplicate runs are shown (red: run 1, dashed pink: run 2).

Table 5. Summary of the DSC Data for the NISTmAb Sample<sup>a</sup>

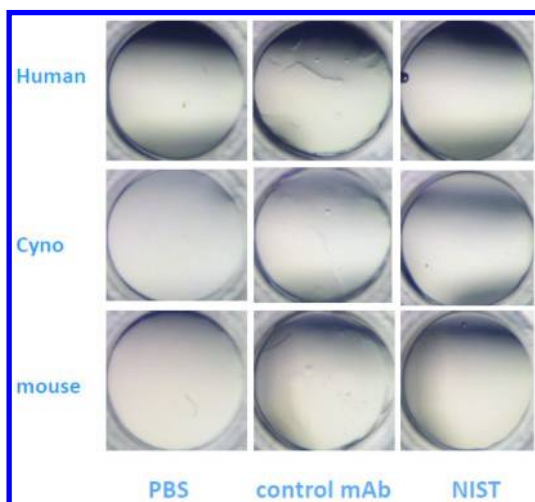
Run	T <sub>m</sub> 1 (°C)	ΔH1 (cal/mol)	T <sub>m</sub> 2 (°C)	ΔH1 (cal/mol)	T <sub>m</sub> 3 (°C)	ΔH3 (cal/mol)	Total ΔH (cal/mol)
1	71.2	1.86E+05	84.0	4.95E+05	88.9	5.38E+05	1.22e+06
2	71.2	1.69E+05	84.1	4.65E+05	88.9	5.29E+05	1.16e+06

<sup>a</sup> DSC: Differential Scanning Calorimetry, NISTmAb: National Institute of Standards and Technology monoclonal antibody.

### Phase Separation in Serum

Because therapeutic mAb molecules are administered at high concentrations, it is important to assess the process of mixing of the mAb candidates with serum to avoid problems that can arise at later stages during in vivo testing. For this reason, we have developed a miniaturized assay (3) that allows the detection of phase separation during the process of mixing the mAb with serum. The assay

mimics physiological conditions where the mAb in formulation buffer at high concentration is injected and the initial concentration at the injection site is high. If the mixing with the serum is not homogeneous, this could lead to ineffective mAb distribution and even cause irritation at the injection site (27). The phase separation of the NISTmAb in sera was evaluated as part of the developability assessment, as we would typically do with therapeutic candidates at discovery stages. NISTmAb sample (at 146.5 mg/mL) in PBS was mixed in equal volume with human, cyno, or mouse serum, as described in Materials and Methods. Figure 5 shows images of the mixtures of NISTmAb with human, cyno, and mouse serum. PBS was used as a negative control, and a proprietary serum-incompatible mAb (control mAb) was used as a positive control and showed phase separation when combined with sera. The images show that NISTmAb when mixed with sera at high concentration forms a homogeneous mixture without any phase separation observed. In comparison to the control mAb images, which have clear phase separation (droplets), the NISTmAb images are clear.



*Figure 5. Phase separation testing of National Institute of Standards and Technology monoclonal antibody (NISTmAb) with human, cynomolgus monkey, and mouse sera. A monoclonal antibody that is known to be incompatible with sera (control mAb) was used as a positive control.*

### Cross-Interaction Chromatography

CIC is a method developed by our scientists (5) as a high-throughput screening tool to predict the apparent solubility of a protein relative to the solubility of known ones. The principle of the method is the measurement of weak protein-protein interactions between a mAb in solution and human polyclonal antibodies coupled to a resin. The method is used to predict the ability of an antibody to be concentrated at levels above 100 mg/mL without precipitation. The data from this method correlates well with the ultrafiltration method but with the benefit of only using microgram amounts of material, which is important in early

stages of drug development. A chromatographic retention factor ( $k'$ ) is calculated from the retention times of the sample in the poly IgG column compared to the poly IgG-free control. The formula used in calculations of the  $k'$  parameter is shown in CIC section of Materials and Methods. If the  $k'$  values calculated from the retention maxima of the peaks in both the poly IgG-coupled and poly IgG-free control columns are greater than 0.6, it is an indication that the molecules are generally significantly less soluble. Also, the most soluble antibodies tested by CIC have considerable lower  $k'$  values (5). CIC was measured using a column coupled with polyclonal human IgG at 25 mg/mL (Figure 6). Samples were loaded on the column at  $\sim 0.1$  mg/mL. The chromatographic retention factor value ( $k'$ ) of NISTmAb is calculated from this method to be  $-0.04$  (Table 6), indicative of minimal protein-protein interactions. A proprietary mAb was used as an interacting positive control (positive standard), with calculated  $k'$  values of 0.40 and 1.09. Another proprietary mAb molecule was used as a non-interacting negative control (negative standard), with a calculated  $k'$  value of 0.08.

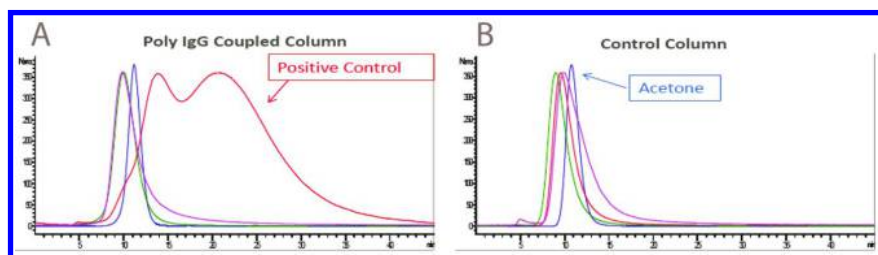


Figure 6. Cross-interaction chromatography (CIC) profile of the National Institute of Standards and Technology monoclonal antibody (NISTmAb). Panel A represents the traces obtained with a polyclonal IgG (poly IgG)-coupled column.

Panel B shows the same samples run over a poly IgG-free control column.

NISTmAb is shown in pink, interacting positive control in red, non-interacting negative control in green, and acetone in blue. CIC results show no interaction with the column for the NISTmAb.

**Table 6. Cross-Interaction Chromatography Data**

Sample	Retention Time (min)		$k'$
	Poly IgG-Coupled Column	Poly IgG-Free Control Column	
NIST monoclonal antibody	9.87	9.95	-0.04
Negative standard	10.09	9.01	0.08
Positive standard	13.90, 20.82	9.59	0.40, 1.09



## Self-Interaction Chromatography

The second virial coefficient ( $B_{22}$ ) is a measure of protein-protein interactions. Several methods that measure the  $B_{22}$  coefficient based on scattering intensities and nanoparticle spectroscopy methods have been described (28, 29). We have used here a chromatographic-based method, as described in the SIC section of Materials and Methods.

Figure 7 and Table 7 show the SIC data for the NISTmAb on protein-coupled and uncoupled (control) columns, along with the calculated  $B_{22}$  values. The  $B_{22}$  value for the NISTmAb was calculated to be 0.39, indicative of minimal self-interactions.

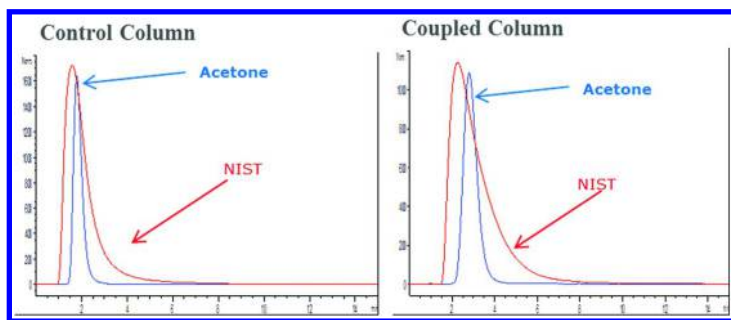


Figure 7. Self-interaction chromatography (SIC) profile of the National Institute of Standards and Technology monoclonal antibody (NISTmAb). The left panel shows the control column, and the right panel shows the protein-coupled column profiles. NISTmAb is shown in red and acetone in blue.

Table 7. Summary of SIC Data of the NISTmAb Candidate<sup>a</sup>

Retention Time(min)		$B_{22}$
NISTmAb Column	Control Column	
2.3	1.6	0.39

<sup>a</sup> NISTmAb: National Institute of Standards and Technology monoclonal antibody, SIC: self-interaction chromatography.

## Results and Discussion: Forced Degradation

### Post-Translational Modifications Risk Assessment of the NISTmAb Candidate

An important component of the developability testing for the therapeutic drug candidates is the assessment of the risks for PTMs of residues, especially those involved in the function of the molecule. Typically, most of the side chains of the 20 naturally occurring amino acids are stable under a variety of different conditions, but some are more reactive than others. For example, the side chains

of methionine and tryptophan residues amino acids are known to be susceptible to oxidation (30), whereas those of asparagine and glutamine are prone to deamidation under certain pH conditions (31). Isomerization of aspartate residues and fragmentation of the mAb molecules at Asp-Pro and Asp-Gly sequences also has been demonstrated (32, 33).

For the purpose of the NISTmAb developability assessment, we decided to evaluate the risks of the molecule undergoing oxidation and deamidation after stress-inducing conditions, as these are considered among the most common types of chemical modifications that occur in therapeutic molecules. To test the chemical stability of the NISTmAb for oxidation and deamidation, the molecule was subjected to photo-oxidation and high pH stresses. Photo-oxidation was performed according to ICH guidelines. The efficiency of photo-exposure was verified by quinine chemical actinometry. The effects of thermal exposure at 40 °C for 2 weeks in PBS, pH 7.4, also were tested. The thermal stress is designed to accelerate sample aggregation that correlates with the propensity to aggregate under long-term storage conditions of the therapeutic molecule. The assays performed to assess the potential changes in the NIST molecule upon exposure to the photo-exposure, high pH, and thermal stresses were microfluidic electrophoresis, cIEF, SEC-MALS, and DLS.

The results of the microfluidic electrophoresis analysis of the stressed samples are shown in Figure 8 and summarized in Table 8. The presence of lower MW species in the nonreduced samples that underwent thermal, deamidation, and photo-oxidation stress is noticeable compared to the untreated NISTmAb material control sample (first two lanes). The high pH stress (deamidation) appears to make the molecule more susceptible to degradation, as observed in the change of the percentage of intact IgG to 81% compared to 99% in the untreated control (Table 2). Thermal stress also seems to promote degradation of the NIST sample, as judged by the 92% of intact molecule remaining after treatment compared to 99% of the control. The presence of the degradation bands for the thermal- and high pH- (deamidation) treated samples is shown by an asterisk in Figure 8. Values above 90% purity or less than a 10% change in purity would be considered acceptable for this type of forced degradation conditions. For those instances when purity was grossly below 90%, an explanation of where the degradation is occurring would be warranted.

The cIEF profiles of the stressed NIST samples are shown in Figure 9 and Table 9. Profiles indicate greater differences in the thermal and high pH samples compared to the control. The sample that was subjected to photo-oxidation seems very similar to the untreated control sample. The observed pI of the main peak for these samples did not change as result of PTM stress, but the percent area of the main peak decreased as result of minor peaks appearing in those PTM samples compared to the control, indicating the presence of heterogeneous species in the PTM samples. The decrease in the main species with an increase in acidic species, which is consistent with deamidation, is expected under these conditions. The pI values and the percent area of the highest peaks for untreated control and PTM NISTmAb samples are shown in the summary table.

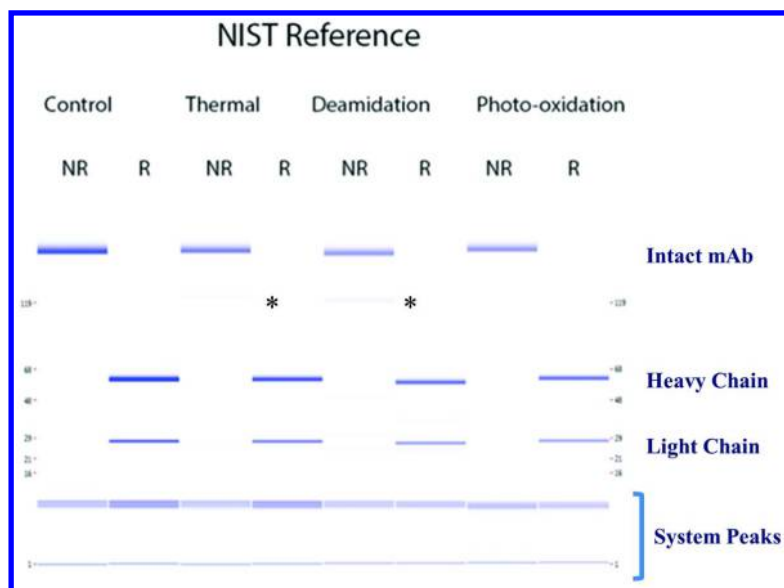


Figure 8. Microfluidic electrophoresis profiles of National Institute of Standards and Technology monoclonal antibody (NISTmAb) material before and after thermal, high pH (deamidation), and photo-oxidation stress exposure at 1 mg/mL. NR: nonreducing, R: reducing conditions, \*: degradation products noted in the Figure.

**Table 8. Summary of Microfluidic Electrophoresis Quantification of NISTmAb after Thermal, High pH (Deamidation), and Photo-Oxidation Stress Exposure at 1 mg/mL<sup>a</sup>**

Sample	mAb (%)	
	NR	R
Thermal	91.5	97.4
High pH	80.7	93.2
Photo-oxidative	96.1	98.7

<sup>a</sup> The values in the Table for the reducing conditions (R) represent the sum of the values for the light chain (LC) and heavy chain (HC). NISTmAb: National Institute of Standards and Technology monoclonal antibody, NR: non-reducing.

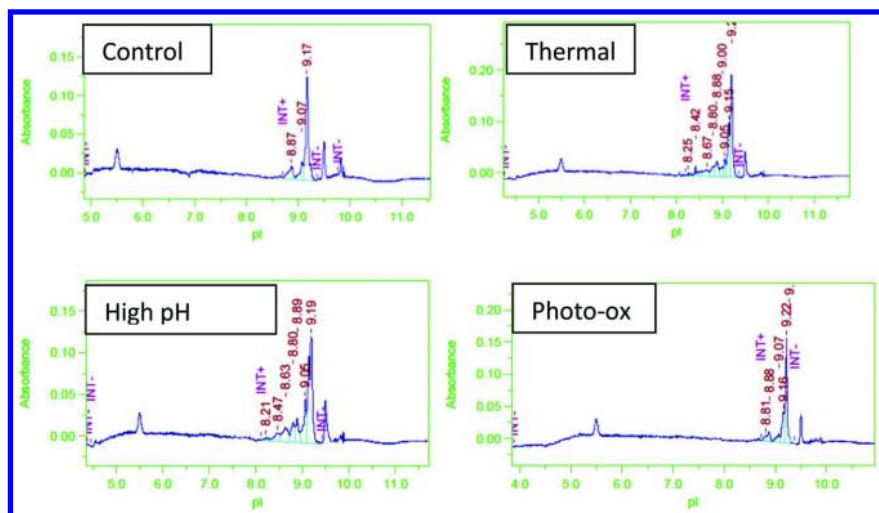


Figure 9. Capillary isoelectric focusing (cIEF) profiles of National Institute of Standards and Technology monoclonal antibody (NISTmAb) samples before (control) and after thermal, deamidation (High pH), and photo-oxidation (Photo-ox) stress exposure at 1 mg/mL.

**Table 9. Summary of cIEF Profiles of NISTmAb after Thermal, High pH (Deamidation), and Photo-Oxidation Stress Exposure<sup>a</sup>**

Sample	<i>pI</i>	Area of Highest Peak (%)
Control	9.2	71
Thermal	9.2	60
High pH	9.2	52
Photo-oxidation	9.2	78

<sup>a</sup> cIEF: capillary isoelectric focusing, NISTmAb: National Institute of Standards and Technology monoclonal antibody.

The results of the SEC-MALS analysis of the control and stressed NIST samples are shown in Figure 10 and Table 10. Overlaid in the Figure 10 are the traces of the thermal (green), high pH (deamidation) (blue) and photo-oxidation sample (red). The presence of lower MW species in the high pH and thermally stressed samples is noticeable in the blue and green traces. Also, the main peak representing the monomeric species is only 76% and 96% for the high pH (deamidation) and thermally stressed samples, respectively, compared to 100% of the untreated control and photo-oxidized ones. As with the other assays used to assess the outcome from high pH incubation, the results indicate that elevated pH can have the most impact on the stability of the NISTmAb.

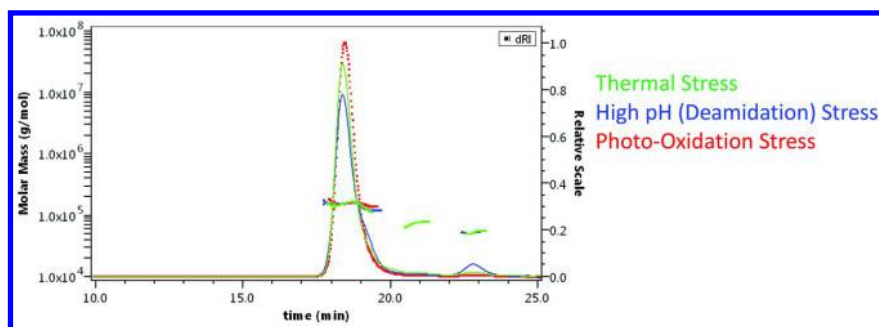


Figure 10. Size-exclusion chromatography–multi-angle light scattering (SEC-MALS) profiles of National Institute of Standards and Technology monoclonal antibody (NISTmAb) stressed samples. Chromatograms for samples exposed to thermal stress (green), high pH (deamidation) stress (blue), and photo-oxidation stress (red) are shown.

**Table 10. SEC-MALS Data of NISTmAb Stability Samples Showing the Amount of Main Species Present in the Release and PTM Samples<sup>a</sup>**

Sample	Monomer (%)
Release	100
Thermal stress	96
High pH stress	76
Photo-oxidation	100

<sup>a</sup> NISTmAb: National Institute of Standards and Technology monoclonal antibody, PTM: post-translational modification, SEC-MALS: size-exclusion chromatography–multi-angle light scattering.

The results of the DLS analysis for the stressed NIST samples are shown in Table 11. A small but significant reduction of the observed  $R_h$  value was observed for the NIST samples that were subjected to thermal and high pH (deamidation) stress. No gross changes were observed under these stress conditions, suggesting good colloidal stability and absence of self-association of NISTmAb.

**Table 11. DLS Data of the Release and PTM NISTmAb Samples<sup>a</sup>**

<i>Sample</i>	<i>R<sub>h</sub> (nm)</i>	<i>Pd (%)</i>	<i>Mass (%)</i>
Control	5.4 ± 0.0	7.9	100.0
Thermal stress	5.0 ± 0.1	8.5	100.0
High pH stress	5.0 ± 0.2	7.8	100.0
Photo-oxidation stress	5.2 ± 0.1	5.1	100.0

<sup>a</sup> Values represent averages of triplicate measurements with standard deviation. DLS: dynamic light scattering, NISTmAb: National Institute of Standards and Technology monoclonal antibody, Pd: polydispersity, PTM: post-translational modification, R<sub>h</sub>: hydrodynamic radius.

### **Preformulation (Buffer pH) Using HTP Formulation with UV and Dynamic Light Scattering Analysis**

The screen is designed to test the solubility and the colloidal stability of therapeutic candidates under buffer conditions varying in pH values from 3.5 to 8.4. Typically, mAb samples have low solubility and high aggregation propensity in pH conditions close to the pI of the molecule. The NISTmAb was tested in the screen using a 96-well plate first subjected to UV measurements at 350 nm. These measurements allow for general particulate light scattering to be detected. A second UV measurement was again taken after the TRamp study (described below) to compare large particulates before and after the temperature ramp. This was done on a Molecular Devices SpectraMax plate reader instrument.

A Wyatt Dynamic light scattering instrument was used to perform real-time DLS on the samples during the TRamp screen. The samples were equilibrated in temperature increments from 26 °C to 70 °C. The light scattering was detected at each temperature to assess aggregation as temperature increased.

Results of the turbidity measured at 350 nm and DLS TRamp screen for the NISTmAb candidate in buffers with different pH are shown in Figures 11 and 12. The results indicate that the NISTmAb is very stable under the conditions that were tested in the screen. No major increase in the turbidity of the samples at eight different buffers with varying pH was detected after heating up the samples to 70 °C. Minor effects on the aggregation of the NISTmAb under the set of buffers tested in this screen were seen from the scattering data. Taken together, the turbidity and scattering data measured in this setup indicate that NISTmAb is stable over a wide pH range varying from 3.5 to 8.4.

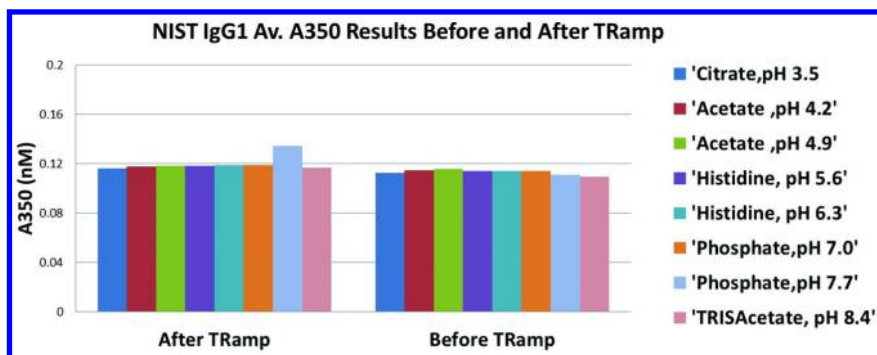


Figure 11. The A350 nm (turbidity) values for the National Institute of Standards and Technology monoclonal antibody (NISTmAb) sample in different buffers are shown before and after the temperature ramp (TRamp) experiment. Values in the graph represent averages from triplicate runs.

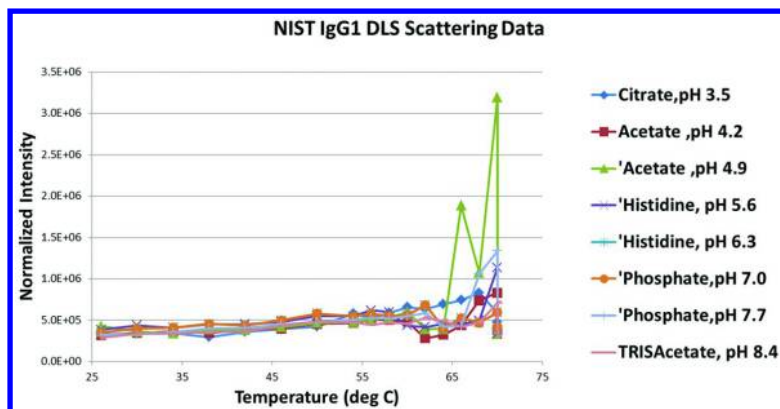


Figure 12. Normalized intensity values as measured by dynamic light scattering (DLS) for the National Institute of Standards and Technology monoclonal antibody (NISTmAb) in different buffers as function of increasing temperatures during the temperature ramp (TRamp) experiment are shown in this graph. The values represent averages of triplicate runs.

## Conclusions

The developability assessment of the NISTmAb candidate molecule by a set of biochemical and biophysical assays as described above indicated that the molecule can be concentrated at levels above 100 mg/mL in PBS buffer and maintains its stability over incubation up to a 1 month period. The evaluation of the NISTmAb was performed in physiological PBS buffer and not in the original formulation of the sample. The NIST material is monomeric, as determined

by SEC-MALS, with a size distribution ( $R_h$ ) typical for a mAb determined by DLS. The concentration of the sample and incubation at 4 °C does not affect the size distribution of the molecule, indicating that it maintains integrity and conformational stability over this period of time. These conclusions are based on the microfluidic electrophoresis, cIEF, SEC-MALS, and DLS analysis of samples at different time points. NIST material is stable, as determined by the DSC profile of the molecule, with thermal transitions starting at 71.2 °C. NIST sample is compatible with human, cyno, and mouse sera because it does not phase separate as does the positive control mAb used for the phase separation assay described in this chapter. The NIST molecule also has minimal self-interaction, as determined by the SIC method, and also minimal cross-interaction, as determined by the CIC method.

An important component of the developability analysis is the assessment of the biochemical and biophysical attributes of the molecule under stress conditions. The types of stress we induced were thermal; photo-oxidation of methionine and tryptophan residues; and high pH, which results in deamidation of asparagine and glutamine residues. The combined results of orthogonal analytical assays indicate that although photo-oxidation under the specific conditions tested does not result in significant oxidation in the molecule, both thermal and high pH stress induce increases in both sample heterogeneity and fragmentation of the molecule. These results indicate potential problems that the molecule might experience during changes in temperature or higher pH either during production or the storage life cycle of the molecule. For a more detailed analysis of the residues that are prone to changes due to stress, high resolution peptide mapping methods are utilized routinely as part of the developability assessment of preclinical candidates. The data from the peptide mapping efforts are not included in this chapter but are considered very important in evaluating the overall developability profile of the candidates and the decision process. For the details on screening for PTMs in the NISTmAb with peptide mapping, refer to the PTMs chapter/Volume 2, Chapter 3 in this series. Collectively, although slight changes were observed under some conditions in this study, the NISTmAb was relatively stable to changes upon stress applied.

Another important component of the developability assessment that is not discussed in this chapter is the analysis of the stressed samples in binding and/or functional assays. That requires knowledge of the therapeutic target and development of assays that are sensitive and reproducible to be able to reliably determine whether the changes observed on the molecular level are significant or not at the binding and/or functional level. Depending on the outcome of the binding and/or functional assays, preclinical candidates can be deprioritized or redesigned to meet specific program-based criteria.

Based on the data presented in this chapter on the characterization of the NISTmAb, it can be concluded that the molecule has a favorable developability profile with high solubility (above 100 mg/mL in PBS), high stability after 1 month incubation at high concentration, and resistance to the stress-inducing conditions that were evaluated in this work.



## References

1. Strohl, W. R.; Strohl, L. M. *Therapeutic Antibody Engineering*; Woodhead Publishing, 2011; Vol. 11.
2. Wen, J.; Arakawa, T.; Philo, J. S. *Anal Biochem* **1996**, *240*, 155–166.
3. Baker, A. E.; Mantz, A. R.; Chiu, M. L. *mAbs* **2014**, *6*, 1509–1517.
4. Jacobs, S. A.; Wu, S. J.; Feng, Y.; Bethea, D.; O'Neil, K. T. *Pharm. Res.* **2010**, *27*, 65–71.
5. Payne, R. W.; Nayar, R.; Tarantino, R.; Del Terzo, S.; Moschera, J.; Di, J.; Heilman, D.; Bray, B.; Manning, M. C.; Henry, C. S. *Biopolymers* **2006**, *84*, 527–533.
6. DePhillips, P.; Lenhoff, A. M. *J. Chromatogr. A* **2000**, *883*, 39–54.
7. Q1b Photostability Testing of New Drug Substances and Products, Q1a(R2). In *ICH Harmonised Tripartite Guideline, Step 4 version*; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, Switzerland, 2003.
8. Shire, S. J.; Shahrokh, Z.; Liu, J. J. *Pharm. Sci.* **2004**, *93*, 1390–1402.
9. Wang, W. *Int. J. Pharm.* **1999**, *185*, 129–188.
10. Shapiro, A. L.; Vinuela, E.; Maizel, J. V., Jr. *Biochem. Biophys. Res. Commun.* **1967**, *28*, 815–820.
11. Creamer, J. S.; Oborny, N. J.; Lunte, S. M. *Anal. Methods* **2014**, *6*, 5427–5449.
12. Chen, X.; Tang, K.; Lee, M.; Flynn, G. C. *Electrophoresis* **2008**, *29*, 4993–5002.
13. Lin, J.; Tan, Q.; Wang, S. *J. Sep. Sci.* **2011**, *34*, 1696–1702.
14. Quan, C.; Alcalá, E.; Petkovska, I.; Matthews, D.; Canova-Davis, E.; Taticek, R.; Ma, S. *Anal. Biochem.* **2008**, *373*, 179–191.
15. Raymond, C.; Robotham, A.; Spearman, M.; Butler, M.; Kelly, J.; Durocher, Y. *mAbs* **2015**, *7*, 571–583.
16. Chen, X. N.; Nguyen, M.; Jacobson, F.; Ouyang, J. *mAbs* **2009**, *1*, 563–571.
17. Vlasak, J.; Ionescu, R. *Curr. Pharm. Biotechnol.* **2008**, *9*, 468–481.
18. Dick, L. W., Jr.; Qiu, D.; Mahon, D.; Adamo, M.; Cheng, K. C. *Biotechnol. Bioeng.* **2008**, *100*, 1132–1143.
19. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M.; Appel, R.; Bairoch, A., Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J., Ed.; Humana Press, 2005; pp 571–607.
20. Demeester, J.; De Smedt, S.; Sanders, N.; Hausteraete, J., Light Scattering. In *Methods for Structural Analysis of Protein Pharmaceuticals*; Jiskoot, W.; Crommelin, D., Eds.; AAPS, 2005; pp 245–277.
21. *Methods for Structural Analysis of Protein Pharmaceuticals*. AAPS, 2005; p 678.
22. Privalov, G. P.; Privalov, P. L. *Methods Enzymol.* **2000**, *323*, 31–62.
23. Plotnikov, V.; Rochalski, A.; Brandts, M.; Brandts, J. F.; Williston, S.; Frasca, V.; Lin, L. N. *Assay Drug. Dev. Technol.* **2002**, *1*, 83–90.
24. Seeliger, D.; Schulz, P.; Litzenburger, T.; Spitz, J.; Hoerer, S.; Blech, M.; Enenkel, B.; Studts, J. M.; Garidel, P.; Karow, A. R. *mAbs* **2015**, *7*, 505–515.

25. Garber, E.; Demarest, S. J. *Biochem. Biophys. Res. Commun.* **2007**, *355*, 751–757.
26. Brader, M. L.; Estey, T.; Bai, S.; Alston, R. W.; Lucas, K. K.; Lantz, S.; Landsman, P.; Maloney, K. M. *Mol. Pharm.* **2015**, *12*, 1005–1017.
27. Dimitrov, D. S. *mAbs* **2010**, *2*, 347–356.
28. Lehermayr, C.; Mahler, H. C.; Mader, K.; Fischer, S. *J. Pharm. Sci.* **2011**, *100*, 2551–2562.
29. Tessier, P. M.; Jinkoji, J.; Cheng, Y. C.; Prentice, J. L.; Lenhoff, A. M. *J. Am. Chem. Soc.* **2008**, *130*, 3106–3112.
30. Stadtman, E. R. *Annu. Rev. Biochem.* **1993**, *62*, 797–821.
31. Stephenson, R. C.; Clarke, S. *J. Biol. Chem.* **1989**, *264*, 6164–6170.
32. Sreedhara, A.; Cordoba, A.; Zhu, Q.; Kwong, J.; Liu, J. *Pharm. Res.* **2012**, *29*, 187–197.
33. Stackhouse, N.; Miller, A. K.; Gadgil, H. S. *J. Pharm. Sci.* **2011**, *100*, 5115–5125.

## Chapter 8

# Protein Particles (0.1 $\mu\text{m}$ to 100 $\mu\text{m}$ )

Dean C. Ripple<sup>\*1</sup> and Linda O. Narhi<sup>2</sup>

<sup>1</sup>Biomolecular Measurement Division,  
National Institute of Standards and Technology,  
Gaithersburg, Maryland, 20899, United States

<sup>2</sup>Research and Development, Amgen Inc.,  
Thousand Oaks, California 91320, United States

\*E-mail: dean.ripple@nist.gov

Protein molecules in solution can form proteinaceous particles by a variety of aggregation processes. The size and concentration of these particles is an important quality attribute for therapeutic monoclonal antibody (mAb) solutions. In this chapter, we describe the techniques commonly used to determine size and count of particles in solution for the size range 2  $\mu\text{m}$  to 100  $\mu\text{m}$ . After first discussing general principles of particle formation and properties, we present general suggestions on sample handling and particle measurement, and then give detailed information on the application of the two most common techniques: light obscuration and flow imaging. The chapter concludes with a description of advanced techniques that extend the measurement size range down to 0.1  $\mu\text{m}$  or characterize the particles more fully.

## Introduction

In this chapter, we will focus primarily on the methods for analyzing and characterizing the subvisible particles (between 2  $\mu\text{m}$  and 100  $\mu\text{m}$ ) in therapeutic protein solutions, as determined by the light obscuration or flow imaging methods described below (1, 2), with some discussion of technologies for analyzing the submicrometer particles (0.1  $\mu\text{m}$  to 2  $\mu\text{m}$ ) as well. The chapter begins with a brief overview of the mechanisms of protein aggregation and how these mechanisms can generate protein aggregates that can range in size from dimers to particles visible to the human eye. We then move on to provide information about the

need and strategic approaches for analysis of the micrometer and submicrometer particles. The discussion focuses primarily on practical techniques for the sizing and counting of protein particles with effective diameters ranging from 2  $\mu\text{m}$  to 100  $\mu\text{m}$ . A number of reviews discuss the relative merits of commonly used instruments (3–5). This chapter complements these reviews by discussing general metrology issues relevant for particle analysis, sample handling issues relevant for common antibody solutions, and detailed considerations for the two most commonly used techniques, light obscuration and flow imaging. To conclude the chapter, we look at the latest advances in counting and characterization techniques, including those that have been applied to the sizing and counting of protein particles in the size range from 0.1  $\mu\text{m}$  to 2  $\mu\text{m}$ .

The smallest protein aggregates, or oligomers, (usually less than 0.1  $\mu\text{m}$  and often called soluble aggregates) are addressed in the Aggregation chapter/Volume 3, Chapter 5. Special techniques for microscopic and spectral characterization are outside the scope of this chapter, but are also discussed in the Aggregation chapter/Volume 3, Chapter 5 and several reviews (4, 6). By providing information on the composition and molecular structure of particles, microscopic and spectral characterization can complement the results of the sizing and counting instruments discussed here.

### Sources of Protein Subvisible Particles (2 $\mu\text{m}$ to 100 $\mu\text{m}$ )

An important product quality attribute monitored for all parenteral therapeutics are subvisible particles, defined here as particles both foreign and from the therapeutic itself in the 2  $\mu\text{m}$  to 100  $\mu\text{m}$  size range. With protein therapeutics, especially at high protein concentrations, many of these particles are actually protein aggregates. Protein self-association can result in protein species ranging from dimers to n-mers that can be greater than 100  $\mu\text{m}$  in size.

As shown in Figure 1 (a combination of the authors' work and Refs. (7–9)), there are many different pathways by which protein aggregates can form, including those driven by colloidal stability, conformational stability, modification, and interactions with available surfaces; all of these are applicable to antibodies (9–12). The fundamentals of protein aggregation are described in detail in the Aggregation chapter/Volume 3, Chapter 5 of this series, so we will provide just a brief overview here. The reactions on the right-hand side of the figure are primarily driven by interactions between different states of the native sequence protein molecule itself, while those on the left-hand side involve interactions with other surfaces or chemical modification of the native protein.

The colloidal stability of a protein (shown as assembly processes in the figure) is the measure of the propensity for the folded molecule to self-associate under specific solution conditions (such as salt type and concentration, pH, and other excipients) based on its molecular properties and especially the surface characteristics of the protein (12, 13). This is an equilibrium reaction between the monomer and aggregate, often starting with association to dimer and building to larger species, including the subvisible particles, in which the initial reaction occurs when the protein is in the native, folded state. This type of self-association is often reversible and increases with increasing protein concentration. Many

therapeutic monoclonal antibodies (mAbs) require protein concentrations of higher than 50 mg/mL, concentrations at which short-range interactions become very important and significant self-association of native protein can occur. Colloidal stability of mAb solutions is an important factor in the formation of micrometer and larger sized particles with long-term storage. A common measure of colloidal stability is the second virial coefficient,  $B_{22}$ , which is a measure of protein-protein interactions (7, 9). Although the initial aggregate is reversible and involves surface interactions (e.g., association between charge patches or hydrophobic patches on the surface), with time in close proximity, irreversible aggregates can form. This can result from the formation of covalent links like disulfide bonds under the appropriate solution conditions, structural rearrangements that result in irreversible non-covalent interactions, or growth in size to where the aggregate is essentially irreversible.

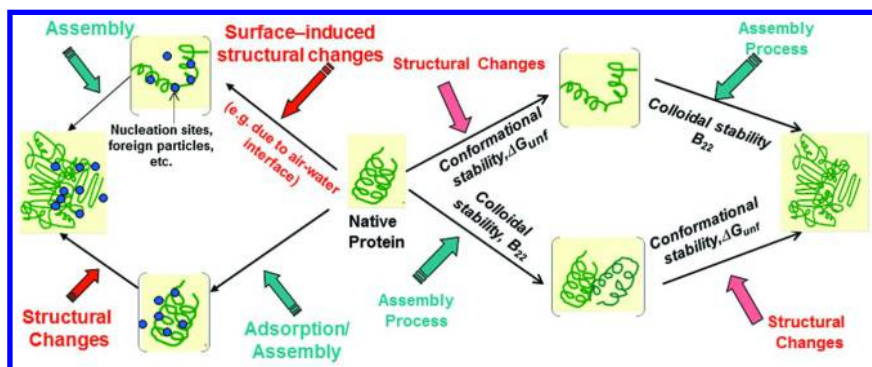


Figure 1. Schematic of the multiple pathways for formation of protein aggregates and subvisible particles. Sources: Adapted from Refs. 7–9 and the authors' work. (see color insert)

The conformational stability of proteins can also contribute to the formation of protein aggregates (8, 9). Proteins exist in an equilibrium between native folded structure and unfolded state(s); this is defined as conformational stability. This is a thermodynamic equilibrium reaction that is often studied and defined by the free energy of unfolding,  $\Delta G_{\text{unf}}$  (7, 9). Protein unfolding can involve multiple intermediate states with surfaces exposed that are normally protected from the solution and from other protein molecules. Exposure of these sites can facilitate protein disulfide cross-linking, hydrophobic interactions, and so forth, which can lead to the formation of irreversible aggregates as well.

Exposure to conditions during process and storage of antibodies can also result in protein modifications (like oxidation) that can sometimes increase the amount of protein aggregate formed. Interactions with the different surfaces present at every step of processing or delivery can also occur and induce aggregate formation (14). The air-liquid interface is a very unique surface that is extremely hydrophobic; protein molecules located right at this interface often unfold. With mixing or agitation, these unfolded molecules are pulled into the solution, where they can interact with the bulk of the protein population present as native molecules, while a different population of molecules migrates to the interface.

These different reactions can occur during manufacture and storage of protein therapeutics and, as shown in Figure 1, can build on top of each other. For example, driven by colloidal stability, native protein can self-associate, and then the protein in these aggregates can partially unfold due to conditions affecting the conformation. The result of all of these interactions is a very heterogeneous population of protein aggregates across a size continuum from dimer to hundreds of micrometers, with multiple different conformations and other covalent and non-covalent forces holding them together. It is this heterogeneous population of aggregates that needs to be characterized.

To make the analytical challenge even more complicated, the total subvisible particle population can also include non-proteinaceous species, such as silicone oil droplets from vial caps or pre-filled syringes, and occasionally, cellulose and other contaminants. Nano- or micro-sized particles shed from containers or processing equipment can also serve as nucleation sites for protein aggregation, resulting in particles that are chemically heterogeneous (8). The resulting particles will be proteinaceous to varying degrees; for brevity, we will refer to all particles that are predominantly protein as “protein particles.”

As a result of the multiple interactions between the inherent mAb properties, the formulation, the container, and the stresses and history of a mAb sample, formation of protein aggregates and particles is not driven only by the properties of the mAb itself, but is instead a result of many factors (see Figure 2). Figure 3 shows optical images of typical protein particles.

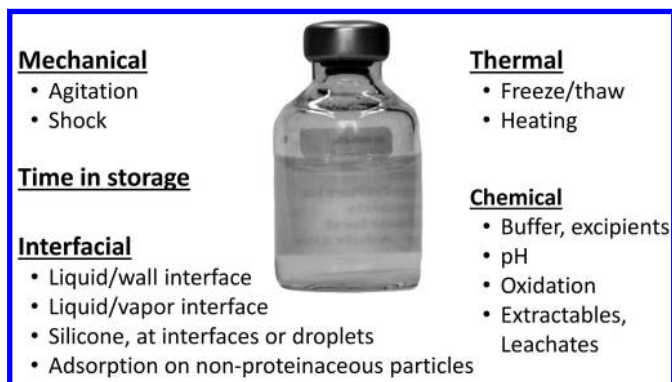
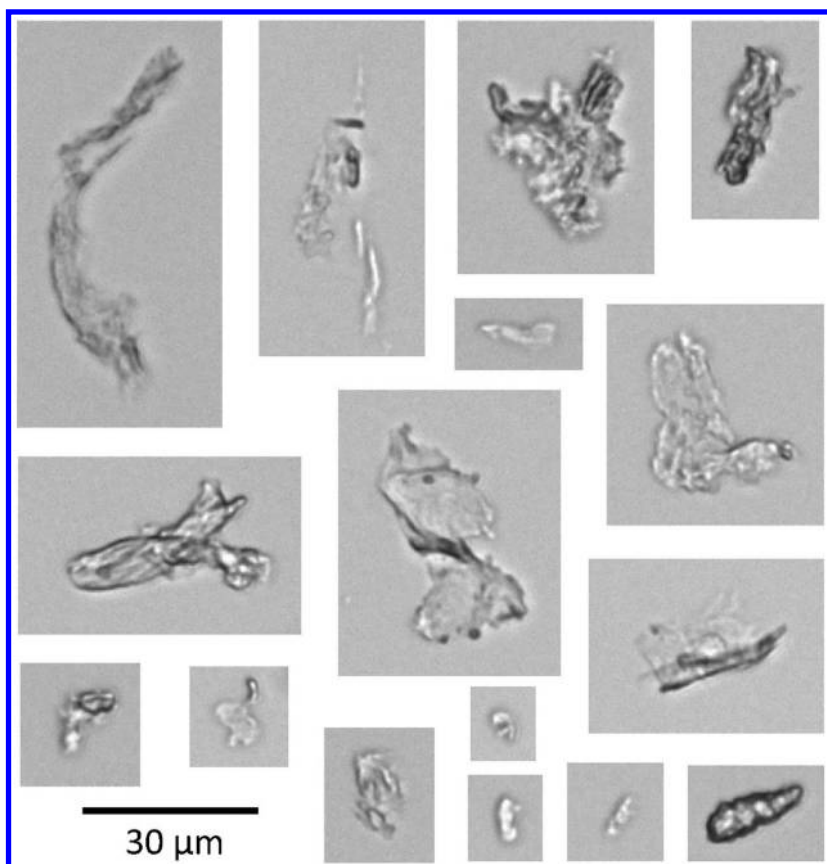


Figure 2. Factors that influence the formation of protein particles.



*Figure 3. Optical images of protein particles created by thermal stress (37 °C for 1 week followed by storage at –80 °C and thawing prior to measurement) of the NIST monoclonal antibody (mAb).*

### Classification of Protein Aggregates

As described above, the total subvisible particle population is very heterogeneous, both in terms of types of molecules in the particles as well as characteristics of the protein particle population. This heterogeneity, coupled with the use of different terms and definitions of these terms, has historically made it very hard to compare results across different proteins and laboratories. Characterization of the subvisible particle population (especially the protein particles) is an important part of product and process characterization and comparability assessments, as well as an essential part of root cause analyses if the subvisible particle profile differs significantly from what was expected. The potential biological consequences of subvisible particles in protein therapeutics have been the subject of much debate and investigation.

Monitoring the number and size of subvisible particles has long been part of the lot release requirements for parenteral drugs. For traditional small molecule pharmaceuticals, particles in this size range are foreign material, derived from the environment or from the surfaces to which the drug is exposed during production and packaging. The compendial limits were established based on historical levels of subvisible particles found in commercial products, with primary safety concerns around possible capillary occlusion from foreign particles in this size range (15). For protein parenterals, especially at high concentration, the majority of the subvisible particles are protein aggregate. Thus with the advent of protein-based therapeutics, the concern shifted to the potential immunogenicity of protein aggregate above 1  $\mu\text{m}$ , especially between  $\approx 1 \mu\text{m}$  and  $\approx 10 \mu\text{m}$ . The uncertainty associated with this potential biological consequence of subvisible protein particles results in the regulatory agencies assigning a high risk to these species (1, 16).

The results of several different studies using *in vitro* and *in vivo* model systems to assess the ability of different types of subvisible protein to activate immune cells and so forth have been published in the last few years (see, for example, (17–20)). Taken together, this work suggests that a subpopulation of protein particles in the approximate range of 1  $\mu\text{m}$  to 10  $\mu\text{m}$  can activate both early- and late-stage responses from the immune cells, but only at very high doses of protein aggregate, and the response is much weaker than that seen with positive controls. The response obtained depends on the size and amount of the protein particles present, as well as other characteristics of the subvisible particle population. Characterization of these species requires analytical tools that can determine particle size distribution and also, as much as possible, the morphology and the degree of dis-sociability of the particles; for protein particles, the conformation of the protein, as well as the nature of any chemical modifications that might be included (2, 19), can also provide important information.

During product and process development, the subvisible particle population should be characterized as thoroughly as appropriate. This information can inform decisions on process steps and even candidate selection. Collecting this data early in the development lifecycle will also help set the baseline for the lot-to-lot variability of this population, the basis for determining comparability across lots, and “what is normal” for the particle size distribution of this product. With this information, the particle size distribution can be determined at lot release and provide enough information to know when a particular lot is significantly different from the usual, warranting further investigation before release. Particle counting and sizing are the most basic analyses; these analyses are required for comparison across samples and ensuring consistency of commercial product. The particle size distribution is also the most important parameter in understanding the biological consequences of these particles, with threshold determinations being of critical importance. As discussed above, these techniques will be the focus of the rest of this chapter.



# Analytical Techniques: General Discussion

## Overview

There are two general approaches for determining particle concentration: ensemble methods and those that measure individual particles. Ensemble techniques integrate a signal from many particles at once. Deconvolution of this signal produces a particle size distribution and total concentration. Dynamic light scattering (DLS) and turbidity are two common ensemble methods (21). For highly polydisperse mixtures in the size range of interest here, the deconvolution process is often imperfect, and ensemble techniques provide only a semiquantitative particle size distribution. Even with this limitation, ensemble methods are often useful for relative, high-throughput measurements of aggregation propensity in the initial stages of product, process, and formulation development. For quantitative characterization of polydisperse suspensions of protein particles, methods that measure individual particles are generally preferred over ensemble methods. These single-particle methods have a number of common traits: each instrument measures physical parameters of individual particles, translates the measurement result to a reported diameter by an instrument-specific algorithm, and derives particle concentration from the number of particles detected per unit volume of fluid sampled. Although these methods do not suffer from deconvolution problems, they do have limitations; for example, no instrument provides a true, three-dimensional representation of the actual particle morphology. Instead, the effective diameter is inferred from a measured parameter of the particle, such as projected area (imaging methods), optical cross section (light scattering methods), or displaced volume (electrical sensing zone [ESZ] method).

Figure 4 illustrates the range of effective diameters covered by the single-particle methods discussed in this chapter, as well as ensemble methods coupled to the size-separation method of asymmetric flow field-flow fractionation (AF4). Table 1 lists some of the advantages and limitations of these methods.

In this section, we discuss measurement issues that are generic for a broad variety of single-particle counting and sizing instruments and are relevant for both established and emerging measurement technologies. To make the discussion easier to follow and less abstract, it is useful to describe the principles of the two most common measurement techniques, light obscuration and flow imaging.

In a light obscuration particle counter, a syringe pump draws sample through a flow cell through which a focused laser passes (see Figure 5A). A particle traversing the beam will scatter and absorb light, reducing the transmitted beam intensity. The reduction in the beam intensity is converted to an equivalent diameter, based on calibration of the instrument with polystyrene latex (PSL) beads of several known diameters. Both aggregated protein and PSL bead calibration standards have negligible optical absorption at the visible or near-infrared wavelengths used in light obscuration, so the reduction in beam intensity is directly related to the scattering of light by the particle or bead under test.

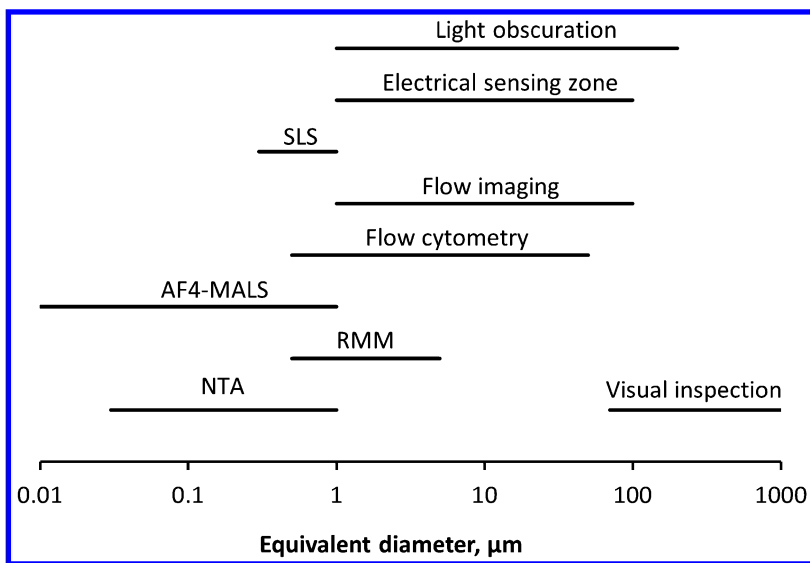


Figure 4. Approximate effective-diameter range for a variety of particle detection methods. Methods depicted are asymmetric flow field-flow fractionation with multi-angle light scattering detection (AF4-MALS), nanoparticle tracking analysis (NTA), resonant mass measurement (RMM), and static light scattering (SLS) (range shown for single-particle SLS). AF4-MALS is an ensemble method; all other methods can be used in a single-particle mode.

**Table 1. Summary of Size Ranges, Advantages, and Limitations of Various Particle Detection Methods**

Method	Advantages	Limitations
Light obscuration (LO)	Rapid, compendial method; substantial history of drug product testing with method	Reported diameter not accurate for particles with low optical contrast; no ability to distinguish particle subpopulations
Flow imaging (FI)	Large dynamic range; morphology analysis	Morphological analysis limited to particles >5 μm
Resonance mass measurement (RMM)	Distinguishes silicone oil from other particles	Reports buoyant mass only; sensitive to channel blockage
Electrical sensing zone (ESZ)	Capable of measuring particles in opalescent or turbid samples	Measurements may require increased ionic conductance; possible interference by electronic noise

*Continued on next page.*

**Table 1. (Continued). Summary of Size Ranges, Advantages, and Limitations of Various Particle Detection Methods**

<i>Method</i>	<i>Advantages</i>	<i>Limitations</i>
Flow cytometry	Capable of distinguishing multiple particle subpopulations	Fluorophores not developed for tagging many particle types; diameter scale not quantitative
Field-flow fractionation (FFF, AF4)	Broad size range; multiple optical detectors after separation	Proteins may adsorb on separation membranes, with possible shift in particle size distribution; methods not optimized for protein particles
Nanoparticle tracking analysis (NTA)	Simplicity of dynamic light scattering (DLS) with improved particle size resolution	Limited dynamic range; low sampling efficiency

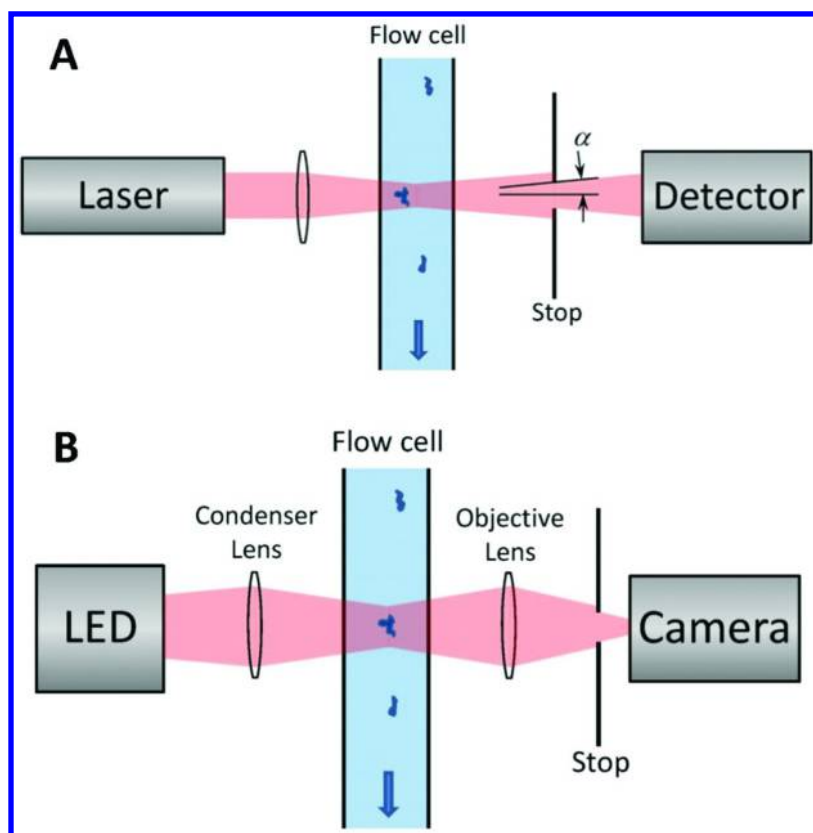


Figure 5. Schematics of (A) light obscuration and (B) flow imaging particle counters. LED is light-emitting diode. (see color insert)

Flow imaging works by capturing images of species as they pass through a flow cell mounted in an automated microscope (see Figure 5B). The light source, objective, and stop may be optimized to increase image contrast and depth of field. The diameter of each particle is then calculated based on the diameter of a circle of the same area as the detected particle, or by similar morphological algorithms.

## Reference Materials

The most commonly used reference materials for particle sizing and counting instruments are suspensions of PSL beads. Commercial standards are available consisting of suspensions of beads of known diameter and concentration. Use of these beads requires care that the beads are well dispersed; if the beads have partially agglomerated, a shoulder will appear next to the main peak of monodisperse beads. An effective procedure to bring beads into suspension is the following: shake the bottle vigorously for 10 s, sonicate for 20 s, let the bottle sit for 10 min to allow air bubbles to rise, and tip the bottle 10 times prior to sampling.

PSL beads work well to calibrate instruments that rely on the displaced volume (ESZ), displaced mass (resonant mass measurement [RMM]), or hydrodynamic radius (DLS or nanoparticle tracking analysis [NTA]) of the particle as the fundamental measurement. (These techniques will be discussed below.) From the point of view of optical counting techniques, PSL beads differ from actual protein particles in two respects: the difference between the refractive index of the particle or bead and the matrix liquid,  $\Delta n$ , is much lower for protein particles than for PSL beads, and protein particles have a much more irregular morphology than PSL beads. For PSL in water,  $\Delta n$  0.25. An upper limit of the refractive index of protein particles can be estimated as 1.40, the refractive index of amorphous protein adsorbed on a surface (22). If the particle is not highly compacted, the refractive index may be even lower, leading to  $\Delta n < 0.07$ . Small values of  $\Delta n$  lead to reduced optical contrast, which causes inaccurate particle sizing for light obscuration (23, 24) and possibly flow imaging (23, 24). Thus, calibration of optical instruments with PSL beads may not give accurate diameter values for particles with much lower optical contrast than these standards.

Silica beads have a refractive index much lower than PSL ( $n \approx 1.42$ – $1.47$ , depending on manufacturing process). In optical instruments, such as flow imaging and light obscuration, silica beads immersed in a liquid of nearly matching refractive index can be used to assess the repeatability and accuracy of the reported diameter (23). Water-glycerol, water-sucrose, or water-(2-pyridinemethanol) mixtures may be used as a matrix fluid of tunable refractive index. Silica beads are not as inherently monodisperse as PSL beads, especially at diameters greater than 10  $\mu\text{m}$ . We recommend that the beads be inspected for diameter uniformity under a microscope prior to use.

The use of silica beads as a diameter standard is only partially satisfactory as a calibration standard. Although the beads do mimic the low optical contrast of protein particles, the bead morphology is quite different. To address this limitation, reference materials made from partially fluorinated polymers are being developed

(25). Once these reference materials become available, they will be a valuable addition or substitute for silica beads.

### Repeatability, Linearity, Accuracy, and Limit of Quantification

Particle counting instruments measure two separate parameters: (1) particle count and (2) particle size. The precision, linearity, and bias of an instrument can be affected by errors in either of these two quantities. The repeatability, linearity, and accuracy of particle count are readily determined by measuring suspensions of monodisperse PSL beads multiple times and at different concentrations. The influence of reduced optical contrast can be assessed using silica beads in matrix fluids of varying refractive index. Use of beads will characterize the fundamental limits of performance of an instrument. Repeatability of actual protein samples may be much worse due to lack of repeatability of the protein particles themselves.

At low concentrations, all of the particle counting instruments in common use (e.g., light obscuration, flow imaging, ESZ instruments) are designed to have excellent linearity; deviations from linearity are likely due to either background particles or adsorption/desorption of particles from tubing or cell walls.

At high concentrations, deviations from linearity will occur when two particles are sensed as a single particle, commonly referred to as a coincidence error. When this occurs, the counter will register these particles as a single particle that is larger in diameter than the individual particles. As a consequence, the reported concentration may be too low for small-diameter particles but too high for large-diameter particles. Coincidence of particles can be a significant error for solutions at very high particle concentrations, and especially when the particles of interest (e.g., protein particles) are surrounded by a higher concentration of another variety (e.g., silicone oil droplets).

Particle diameter of irregular particles is a more difficult parameter to assess for repeatability and accuracy. Errors in diameter can affect particle counts significantly, and these errors are often larger than either the instrument linearity or repeatability (23, 25). Suppose that we are measuring the concentration of particles greater than a diameter limit  $d$ ,  $N(d)$ . For polydisperse particle distributions, an error in the measured diameter,  $\Delta d$ , results in an error in  $N$  of  $\Delta N = N(d + \Delta d) - N(d)$ . Near the lowest diameter limits of many types of particle counters, including both flow imaging and light obscuration instruments, the error in  $N$  can be substantial. See Figure 6 for a graphical explanation.

Commercial flow imaging and light obscuration instruments in good repair generally have a negligible rate of false-positive particle counts. What is often not negligible is the background particle concentration due to detection of real but unintended particles. In addition to particles from the blank sample or from the container (discussed in the Sample Handling and Interferences sections, below), particles may also desorb from tubing walls. Note that these background sources are a characteristic of the recent usage history of the instrument and not an intrinsic characteristic of the instrument itself.

Room air is commonly assumed to be a major contributor to background counts. In fact, other particle generation processes may introduce higher levels of contaminants than short-term exposure to laboratory air of typical quality.

Important sources of extraneous particles include abrasion products created by mating plastic fittings (especially if the mating occurs with dry plastics) and particle hold-up on surfaces and crevices. These sources can be minimized by mating plastic fittings that are prefilled with water and by disassembling, wiping, and rinsing tubing and fittings suspected of holding particles.

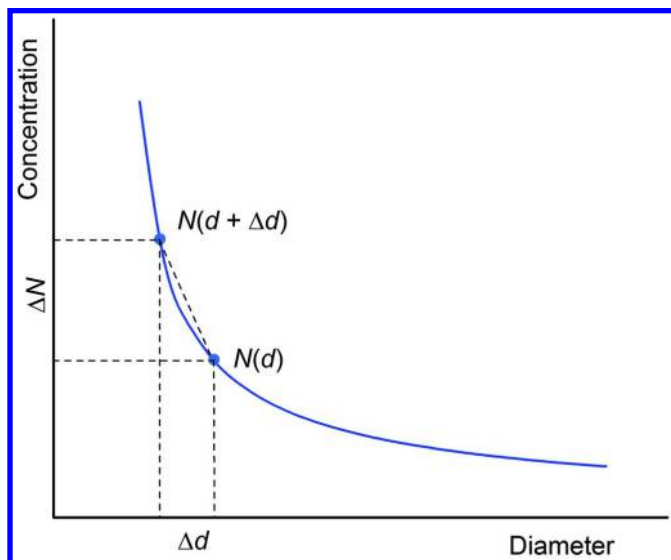


Figure 6. Relationship between a diameter error  $\Delta d$  and a count error  $\Delta N$ . The curved line indicates an arbitrary particle size distribution. The error in concentration is given by  $\Delta N = -m\Delta d$ , where  $m$  is the slope of the diagonal dashed line.

Many published papers have documented the poor agreement between different particle counting methods (3, 26, 27). Different methods give particle concentrations that often differ by a factor of 2 or more, and in some cases, differences can exceed a factor of 10. Untangling the causes of these discrepancies requires understanding how each instrument determines particle concentration and size and how diameter is defined for each instrument. Some of the causes of differences in measured diameter are discussed below in the sections on method accuracy.

### Particle Orientation

When a homogeneous suspension of particles travels through a flow cell, hydrodynamic effects can lead to particle rotations and displacements that can affect the apparent particle size and count. One relatively minor effect is that particles in a flow cell will move away from the wall (28). This effect can cause

a small error in flow imaging instruments; the error can be minimized by either imaging the full width of the flow cell or by setting the imaged area at the center of the flow cell, and by calibrating the instrument with PSL beads of known count with the imaged area in the same location. Another effect is that fibrous or plate-like particles often preferentially align in the high-shear fields of a flow cell (29). Although this alignment improves the quality of the particle images, another consequence of this alignment is that the true volume cannot be obtained from the projected area of a particle without assuming that the particle has a particular shape. For example, a randomly oriented prolate spheroid (a sphere elongated along one axis with aspect ratio defined as the width divided by length) has a projected area that differs from an aligned spheroid by 15% for an aspect ratio of 0.5 and 20% for an aspect ratio of 0.25 (30). At present, there are no controlled studies of particle orientation effects on actual protein particles.

## Sampling

The variance in particle counts routinely exceeds the variance expected simply from the statistical fluctuations due to the finite number of counted particles. There are two causes of this variability. First, formation of particles, and the count and size of those particles, is not a predictable process from vial to vial. It is quite common to observe large variations between different vials of nominally identical contents and processing. Second, each vial may have a heterogeneous distribution of particles throughout its volume, and full homogeneity may be hard to achieve without altering the particles. Consequently, variability of samples should be assessed by measuring multiple vials. The relative contributions of statistical fluctuations, heterogeneity within a single vial, and vial-to-vial differences can be assessed by comparing expected uncertainties from the Poisson distribution, intra-vial repeatability, and vial-to-vial repeatability.

Statistical fluctuations and the observed heterogeneity of a single vial are related to the volume actually measured from each vial. At one extreme, 100% testing of a vial will completely eliminate any effects of heterogeneity within the vial. The volume tested for particle counting applications varies greatly depending on the method used. Visible particle inspection, although only semiquantitative, has the advantage of inspecting 100% of samples, which may themselves have volumes of many milliliters. Pharmacopeial methods using light obscuration require a sample of 1–5 mL, with sampling of up to 100% of that volume. Flow imaging typically tests less than 1 mL per run, with a sampling efficiency ranging from 20% to 85% for that volume. Methods used for counting particles below 1  $\mu\text{m}$  in diameter may have very low sampling efficiencies when drawing from milliliter-size samples. For example, NTA and RMM methods typically sample only tens of microliters per run. When the size range of interest is below 1  $\mu\text{m}$ , even small volumes of sample contain enough particles to achieve statistical uncertainties below a few percent of the total count. The more significant concern is that a small sample will not be representative of the average distribution of particle throughout the sample. To address this concern, samples may be drawn either from multiple locations of a single vial or, preferably, from multiple vials.

## Filtration for Characterization

Filtration of several types is also used to harvest protein particles to enable subsequent spectral or microscopic characterization. The simplest method of concentrating protein particles is to allow the particles to gravitationally sediment overnight. Placing the vial in a thermally insulated container reduces natural convection and can increase the sedimentation rate of small particles. When images of particles are not sought but only identification of secondary structure, particles may be concentrated by centrifugation. The supernatant is discarded, the concentrated particles resuspended in buffer, and the resulting slurry then analyzed as is or air dried on an appropriate substrate for the characterization method used (19). When particles are to be imaged, great care is needed for sample preparation (31, 32). Because protein particles are highly hydrated and flexible, attempts to capture protein particles on membrane filters may lead to large changes in the particle geometry (4). The membrane should be chosen to give a low background for the characterization method being used, and great care is needed to capture particles, gently wash the particles of excipients, and then dry the sample. Spatial resolution of  $\approx 50 \mu\text{m}$  may be achieved with micro-spectroscopy methods (32). For particles larger than  $100 \mu\text{m}$ , *in situ* spectral characterization may be possible, reducing the difficulty in sample handling (33).

## Sample Characteristics and Handling

### Overview

Compared to many other particles, protein particles are particularly delicate and require special sample handling to obtain repeatable, accurate counts. This section discusses methods for handling samples and known interferences common in protein solutions.

### Handling

Handling of protein solutions must strike a balance between ensuring a homogeneous sample (which requires mixing of the sample and removal of microscopic bubbles) and minimizing changes in the particle size distribution (which can occur on sample agitation) (4). Because protein particles have a density slightly higher than the matrix fluid, they will slowly sediment to the bottom of the storage vial. Slow natural convection of samples will generally keep particles of diameter near  $1 \mu\text{m}$  in suspension indefinitely; sedimentation is primarily an issue for larger protein particles. Mixing the samples by slow tipping effectively eliminates initial sedimentation of particles. The rate of sedimentation is generally small enough that measurements are not greatly affected by sedimentation occurring during the course of the measurement.

As a general guide, protein solutions can be effectively mixed by gently tipping the vial back and forth 10–20 times. By repeatedly tipping the sample and measuring the particle size distribution, it is possible to assess the sensitivity of any particular sample to the tipping process. For some samples, even gentle



tipping can cause changes in the particle size distribution, and tipping fewer times or at less of an angle may give the best repeatability.

## Sample Stability

A protein aggregate is said to be reversible if the aggregate breaks down into soluble protein monomer or small oligomers. This dissociation can occur spontaneously during storage or as the result of a change in solution conditions. The degree to which this occurs depends on the protein, the formulation, and the nature of the stress that created the particle (19, 34). There is no standard length of storage time over which reversibility is determined, although typical time scales are approximately 1 day. Conversely, storage conditions may promote the growth of particles.

Protein particles may also change upon dilution (including dissociating or increasing in amount and size). To test for changes in particle size or concentration upon dilution, measure the particle size distribution for a series of dilutions and determine if the distributions scale as expected from the dilution factors (26). Dilution may be necessary to ensure that particle concentrations remain within the linear range of a particular type of instrument or to reduce the solution viscosity or opalescence to manageable levels. Such dilutions should only be done when necessary and with a range of dilution factors so that unexpected systematic effects may be identified. Dilutions may be made into buffer or filtered protein solution. Again, examining the consistency of the results with the results anticipated from the dilution factor can guide the choice of diluent.

## Degassing

Protein solutions are known to be susceptible to the formation of air bubbles, and these bubbles can lead to erroneously high particle counts (35). Traditional methods of removing air bubbles are sonication of the sample or simply letting the sample sit for 1 or 2 hours (15). However, sonication of protein samples can often generate or otherwise alter protein particles, and letting the sample sit is sometimes ineffective. As an alternative, degassing under a partial vacuum has been found to be an effective means of eliminating air bubbles (35).

A liquid maintained at constant temperature under an air atmosphere will become saturated with air over time. If this liquid is then raised in temperature, the solubility of air in the liquid is lower, the liquid becomes oversaturated with air, and any small vapor nucleation sites will grow into air bubbles within the liquid. This scenario can result in a large concentration of air bubbles in samples that have been removed from refrigerated storage conditions or even have simply been transferred from one laboratory to another slightly warmer laboratory. An additional source of air bubbles is the reconstitution of lyophilized protein samples. Suspensions of protein particles are particularly susceptible to trapped air bubbles because the high viscosity of some drug products inhibits the gravitational rise of bubbles to the air-liquid surface, and possibly because the irregular morphology of protein particles may provide locations for nucleation or capture of air bubbles.

The goal of degassing is not to remove all of the dissolved air; it is sufficient to remove enough air so that the liquid is no longer oversaturated. Any existing air bubbles will then readily be absorbed by the liquid. Thus, exposing the samples to only partial vacuum and for limited amounts of time still suffices in preventing air bubble formation. For best results, the degree of vacuum and the amount of time need to be determined empirically for each protein. Typical conditions are a vacuum of 10 kPa (approximately 0.1 bar) applied for a period of approximately 1 hour.

## Interferences

There are several known interferences in counting particles, including silicon oil droplets, air bubbles, and high particle load in nominally particle-free diluents or buffers.

Air bubbles can be diagnosed in flow imaging data by the circularity of the image and by the darkness of the image relative to typical protein particles (26). For other methods, air bubbles may be diagnosed by comparing particle measurements before and after sample degassing. An understanding of the mechanism of air bubble formation (see the Sample Handling and Degassing sections, above) can assist the analyst in minimizing the occurrence of air bubbles.

Silicone oil droplets are commonly found in samples taken from prefilled syringes containing drug product, and sometimes from the small amount of silicone oil found on the caps of vials (36). When measuring samples containing silicone oil droplets, the primary goal is to determine the particle size distribution of the two subpopulations: the oil droplets and the protein particles. Light obscuration is not capable of distinguishing oil droplets from other particles. Flow imaging combined with an analysis of morphological parameters has been shown to be effective in differentiating silicone oil droplets from protein particles for diameters of 3–5  $\mu\text{m}$  and greater (36, 37). For smaller droplets (up to 5  $\mu\text{m}$ ), the relatively new method of RMM (described in Advanced Techniques, below) is very useful because silicone oil droplets appear as particles of positive buoyant mass, whereas all other common particles appear as particles of negative buoyant mass (38, 39).

Understanding the response of a protein to a particle-forming stress or understanding the linearity of an instrument on dilution both require buffers and containers that do not contribute an appreciable concentration of particles. Containers can shed particles as a result of inadequate initial cleaning, chemical degradation of surfaces, abrasion of stirrers or other objects on surfaces, and opening and closing the container closure. Solid buffer salts may contain non-soluble particles. Cleaning of containers is a compromise: extensive cleaning can remove extraneous contaminants but at the same time may damage the container surface, leading to increased particle shedding. Filtration can be effective in reducing particle loads in stock buffer solutions, but the user must confirm that the filters used do not themselves shed significant concentrations of particles (40). Testing the particle concentration of blanks with the same formulation, container, and applied stresses (i.e., samples identical in all aspects

except with no protein content) is critical in assessing the fraction of particles arising from sources other than the protein solutions being tested.

At high concentrations, protein solutions have a higher refractive index, leading to less optical contrast between protein particles and the surrounding fluid, and are often opalescent or turbid due to microscopic variations in refractive index caused by protein-protein interactions. Sharma and colleagues (26) have shown that compared to light obscuration, flow imaging is relatively insensitive to solution opalescence. Reduced optical contrast is known to cause inaccurate counts for both light obscuration (23, 24) and flow imaging (23) techniques. Non-optical methods such as ESZ (27, 41) or RMM (38, 39) are completely insensitive to opalescence or to reduced optical contrast.

## Metrology: Light Obscuration

### Overview

As a result of its historical use in pharmacopeial methods for particle measurement and its ease-of-use, light obscuration is the most common method of particle measurement in the biopharmaceutical industry, (15, 42). In this section, we examine particular measurement issues arising with light obscuration.

Light obscuration has several advantages as a particle counting method: the instruments are easy to use, measurement speed is high, industry has much experience with the method, and historical values can be used for comparison. Light obscuration is the preferred method of counting protein particles in accordance with existing pharmacopeial standards. However, there are two fundamental difficulties with light obscuration. First, the reported particle diameter differs significantly from the physical particle diameter when measuring protein particles. Second, the method cannot discriminate between different types of particles, and an independent method is needed to ascertain the relative concentrations of different particle subpopulations.

Light obscuration instruments are available with a variety of sensor heads and can be calibrated at a variety of flow rates. The most common industry practice is to use sensor heads suitable for the approximate diameter range of 1–100  $\mu\text{m}$ , and flow rates of 10 or 25 mL/min. It is possible to extend the lower range of a light obscuration instrument to approximately 0.5  $\mu\text{m}$  diameter by directly detecting the scattered light instead of inferring the amount of scattering by measuring the transmitted beam intensity.

### Sample Handling

Because of the speed of measurement, sample sedimentation is rarely an issue, provided the measurement is initiated immediately after inserting the sampling needle into the sample. Sedimentation effects can also be minimized by either measuring the complete test sample or by locating the tip of the sampling needle halfway into the test sample.

Air bubbles can cause significant errors in light obscuration due to the lack of specificity of the method combined with the prevalence of air bubbles in the typical

size range covered by light obscuration. The United States Pharmacopeia (USP) method for degassing protein therapeutics (42) should be followed to minimize the effects of air bubbles.

Manufacturers specify an upper limit of liquid viscosity for reliable operation. When attempting to measure liquids of higher viscosity, volumetric flows may be inaccurate due to air leaks or cavitation. Two methods of enabling measurement of higher viscosity liquids are: (1) set the instrument to run at a smaller volumetric flow rate and recalibrate the instrument response for this flow rate; or (2) pressurize the sample inlet (43).

### **Linearity, Repeatability, Accuracy, and Limit of Quantitation**

Coincidence errors in light obscuration occur when more than two particles are simultaneously within the path of the laser. This effect becomes negligible if the particle concentration is sufficiently dilute. Typical concentration limits for less than 10% coincidence errors are in the range of  $10^4$  to  $10^5$  particles per milliliter. In practice, the protein particle counts for drug products rarely reach these limits. However, coincidence errors may be important in the testing of intentionally stressed samples or in the testing of samples from prefilled syringes with a high concentration of oil droplets.

Because a light obscuration counter counts all particles that pass through the flow cell, errors in count are predominantly due to errors in particle diameter, provided that the cell is unblocked and the syringe pump and tubing have no leaks. Running tests with PSL beads of known concentration suffices to demonstrate count accuracy.

When applied to suspensions of protein particles, the diameter accuracy of light obscuration counters is poor throughout the size range of these instruments, and the resulting error in diameter can give large errors in particle concentration. As an example, the discrepancies between light obscuration and flow imaging data from Singh et al. (Table 4 in Ref. (44)) are equivalent to approximately a factor of two difference in diameters as determined by the two techniques. The diameter reported by a light obscuration counter is equal to the diameter of a PSL bead of the same equivalent light scattering as the tested particle. This reported diameter is approximately equal to the physical particle diameter, provided that the particles are larger than approximately 10  $\mu\text{m}$  diameter and have a refractive index difference from the surrounding fluid of 0.1 or more. For protein particles, the refractive index difference is typically much smaller than 0.1, and PSL beads do not provide an accurate calibration basis. The inadequacy of PSL beads at mimicking the optical scattering of protein particles leads to a high sensitivity of the measurement to the refractive index difference between particles and the matrix fluid and substantial errors in the reported diameter. One example of this high sensitivity is the apparent drop in particle concentration as a given number of particles are spiked into protein solutions of increasing protein concentration (23, 26). As the protein concentration rises, the refractive index of the matrix fluid rises as well, decreasing the refractive index difference between particles and fluid. In spite of the low diameter accuracy, light obscuration is still widely used because of

its role in compendial methods, its ease-of-use, and the experience and historical data accumulated with this technique.

The limit of quantification is very good with light obscuration and is limited only by the background count levels. Background counts can arise due to particles found in nominally particle-free water or buffers, electrical or optical noise, surface contamination of the illuminated area of the flow cell, desorption of particles from wetted surfaces, or diffusion of particles trapped in crevices. Because the source of background counts is so diverse, the analyst should run appropriate blanks at the beginning and interspersed among other samples.

For several lots of IgG particle suspensions, Cao et al. (35) obtained a repeatability of 9% of particle count and an intermediate precision of 11% of the count. These relative standard deviation (RSD) values reflect the combined variability of both diameter and count for these polydisperse samples.

## Quality Assurance and Troubleshooting

Pharmacopeial methods require that users run both blanks and PSL count standards to assure the operation of light obscuration apparatus. To adapt these methods to the smaller sizes of interest to regulatory bodies and industry, we suggest an additional test of running PSL beads of a known diameter that is close to the lower limit of the instrument.

The most common problems with light obscuration instruments, along with recommended solutions are:

- Inaccuracy in particle count. The flow cell may be blocked (see below), or the syringe pump or tubing may be leaking.
- Drift in diameter indication. Recalibrate with PSL beads as recommended by the manufacturer.
- High count level in the lowest size bin. Contamination of the illuminated portion of the flow-cell walls can lead to a high level of noise and spurious counts in the lowest size bin. Backflushing and cleaning of the cell as recommended by the manufacturer may help. We have found running small-diameter (5  $\mu\text{m}$  or less) PSL bead samples through the instrument after cleaning to be helpful: we surmise that the polymer beads help to scrub the surface.
- Blocked, or partially blocked cell. Any blockage of the cell can cause poor repeatability of counts and inaccurate readings of count standards. Backflushing the cell, followed, if necessary, by cleaning, removes most blockages.
- Bubbles in the syringe pump. Bubbles can be caused by either leaks in the fluid lines or cavitation, especially when measuring high-viscosity liquids. Check for integrity of fluid lines, and refer to the section on Sample Handling.

# Metrology: Flow Imaging

## Overview

Flow imaging offers a number of advantages over light obscuration: a more direct determination of particle diameter, greater sensitivity to particles with low optical contrast (23), and differentiation of different particle types by morphological analysis (36, 37). Although not as fast or as easy to use as light obscuration counters, the development of both robotic and turn-key flow imaging systems have increased the throughput and ease-of-use of these devices. Although light obscuration remains invaluable as a lot release test, flow imaging is increasingly important for product development and characterization. In this section, we examine the specific measurement issues of flow imaging.

Flow imaging systems are appealing because the morphological parameters provide a mechanism to distinguish between particles of different types. As described in the Introduction, the heterogeneity of the subvisible particle population makes it very difficult to compare results across laboratories and samples, as well as tie particular species to potential biological consequences, without advanced techniques that can distinguish between particle or droplet subpopulations. Flow imaging does enable identification of subpopulations by morphological analysis, and the method has been extremely effective in the analysis of subvisible particles.

Typical flow imaging systems use cell depths of 80  $\mu\text{m}$  to 300  $\mu\text{m}$  for 4 $\times$  magnification or 80  $\mu\text{m}$  to 100  $\mu\text{m}$  for 10 $\times$  magnification. Use of higher magnification and numerical aperture does not necessarily give improved performance. Increased magnification leads to lower contrast and a smaller depth of field, both of which can lead to counting and sizing errors. There is a fundamental trade-off between good optical resolution and large depth of field, and it is therefore difficult for a single instrument to provide both optimal images and optimal count accuracy (45).

## Sample Handling and Sample Characteristics

In identifying particles, it is useful to generate a library of images for particles of known chemical composition generated by a known method. For particles below 10  $\mu\text{m}$  in equivalent diameter, the library will be most useful to generate typical image attributes (e.g., intensity, aspect ratio) characteristic of particular particle compositions.

Protein particles themselves are only slightly denser than the surrounding protein solution, and sedimentation is usually not a major problem. Some measurable effects can be seen for particles greater than approximately 20  $\mu\text{m}$  in diameter. One method to assess the magnitude of these effects is to export the data and then determine the count rate for particles above a certain threshold as a function of time throughout the run. Sedimentation effects can be minimized by priming the system over as short a period as possible and by not terminating runs until nearly all of the tested volume has passed through the flow cell.

Because the run volumes are relatively small (<1 mL) for flow imaging, a significant number of particles may adhere to the walls of the tubing and other

components of the flow path above the flow cell. One technique of minimizing the potential loss of particles by wall adsorption is to not rinse the system with water or particle-free buffer between counting runs. Varying the rinsing protocols and examining the resulting variation in counts can be useful to quantify this effect.

Commercial flow imaging instruments acquire a set of background images to obtain a light-intensity map over the image plane. If the sample has a significantly different refractive index or opalescence than the liquid used for acquiring the background images (i.e., a difference in  $n$  of  $\approx 0.01$  or more), particle detection will not be optimally sensitive. Measuring the background using filtered protein solution can be a useful approach.

### Linearity, Repeatability, Accuracy, and Limit of Quantitation

Coincidence errors in a flow imaging system can occur when the projected areas of two particles overlap, leading to a slight non-linearity in the instrument response versus particle concentration. In effect, an error occurs when two particles are within an approximate volume  $At$ , where  $A$  is the projected area of the larger particle and  $t$  is the cell thickness. Coincidence errors are negligible provided that  $NAt \ll 1$ . In practice, the linearity of flow imaging systems is excellent and counts as high as  $10^6 \text{ mL}^{-1}$  can have negligible coincidence errors for typical polydisperse protein particle samples.

Although many publications report standard deviations for flow imaging particle counts, it is difficult to discern if the counts were obtained in quick succession or for several independent vials. The intermediate precision attainable with flow imaging can be inferred from the data reported by Sharma et al. (26) for a dilution study in which known quantities of aggregated mAb were spiked into nearly particle-free mAb solutions. Over a range of dilutions greater than 1600 (from concentrations of  $100 \text{ mL}^{-1}$  to  $160\,000 \text{ mL}^{-1}$ ), flow imaging of protein particles gave total counts that agreed with the expected counts within an RSD of 9% for all particles greater than  $2 \mu\text{m}$  diameter and  $\approx 20\%$  for particles within the size ranges  $2\text{--}10 \mu\text{m}$ ,  $10\text{--}25 \mu\text{m}$ , and  $25\text{--}50 \mu\text{m}$ . As with the light obscuration precision values, these RSD values reflect the combined variability of both diameter and count for these polydisperse samples.

Unlike light obscuration instruments or other non-imaging instruments, the accuracy of flow imaging instruments does not depend greatly on the accuracy of the flow rate through the instrument. Every time an image is acquired, the system samples a volume equal to the thickness of the flow cell times the imaged area. The lateral dimensions of the image are equal to the physical dimensions of the imaging optical sensor divided by the objective magnification. The magnification can be confirmed by running large-diameter PSL beads of known diameter through the system. What is much harder to calibrate is the thickness of the flow cell, which can vary from one cell to another. The flow cell thickness can be verified by measuring the concentration of commercial count standards; if the cell thickness differs appreciably from its nominal value, the counts will be in error.

There are additional issues with the handling of particles that are near the edge of the field of view, which is primarily a problem for accurate counts of large particles. Inclusion of particles that are truncated at the image edge will result in

a count that is slightly too high for small particles and slightly too low for large particles. There are two ways of assessing the effects of truncated particles. One method is to export the data and generate particle size distributions with truncated particles included, truncated particles included only for two sides (e.g., particles that intersect the top or left edge only are counted), and all truncated particles excluded. Another method is to export the data and count only particles with a center of mass that is at least a “guard” distance  $g$  from the edge. When deriving the particle size distribution from these counts, the imaged area per exposure is now  $(w_x - 2g)(w_y - 2g)$  instead of  $w_x w_y$ , where  $w_x$  and  $w_y$  are the dimensions of the imaged portion of the flow cell. International standards on particle counting give additional details (46, 47).

The lower size limit of detection for flow imaging systems is generally near the optical resolution of the system, which is limited to 1–3  $\mu\text{m}$  by the small numerical aperture of the objectives or by the limited depth of field of the objective. Optical systems are capable of detecting particles smaller than the optical resolution, but images of these particles will have little morphology information, and the diameter of these images will not correspond to the actual physical diameter of the particle. Some flow imaging systems partially compensate for this effect. To further complicate the situation, as the refractive index drops, the measured diameter of small particles will also drop, which is an effect opposite in sign to the optical resolution effect. The combined magnitude of both effects may be assessed by measuring silica beads of a diameter close to the lower limit of the system suspended in a water-glycerol matrix (23).

As discussed in Analytical Techniques: General Discussion section, for particles of diameter close to the lower limit of detection, these sizing errors can lead to large errors in particle counts.

The decrease of refractive index difference can lead to an additional effect for large particles: extended protein particles may consist of relatively dense regions connected by nearly invisible low-density regions. The particle recognition software may detect such particles as two or more smaller particles.

## Quality Assurance and Troubleshooting

There are several common quality assurance techniques applicable to flow imaging that are straightforward: monitor repeatability of replicate samples and run blanks, and count PSL beads of known concentration. Since the count accuracy depends on the cell thickness, it is also important that each flow cell used be checked for count accuracy with PSL beads of known concentration.

There is a less obvious method that can identify partially blocked cells and particles that are stuck on a surface. Assume that the variable  $x$  measures the distance across the field of view transverse to the flow direction, and  $N_i(x)$  is the cumulative number of particles captured for transverse pixel values between 0 and  $x$ . If the particles are randomly distributed, as they should be, then a plot of  $N_i(x)/N$ , where  $N$  is the total number of particles measured, will be a straight line. Steps in the line indicate stuck particles. Smooth deviations from a straight line are indicative of a clog or partial clog in the flow cell.



The most common problems with flow imaging, along with recommended solutions, are the following:

- Improper focus, as indicated by incorrect diameter of PSL beads. Refocus the instrument according to manufacturer's recommendations.
- Blocked cell. Backflush the cell or replace it.
- High counts in blanks. Replace the polymer tubing or fittings upstream of the flow cell; wipe off and rinse with particle-free water any mating metal tubing.

## Advanced Techniques

### Overview

As discussed above, optical methods may not accurately count particles suspended in formulations that are either opalescent or have a refractive index close to that of the particles. Another weakness of both light obscuration and flow imaging is their inability to count particles of diameter below  $\approx 2 \mu\text{m}$ ; flow imaging also cannot reliably differentiate between different particle populations below  $\approx 5 \mu\text{m}$ . Exploration of methods for counting and characterizing particles in the size range  $0.1\text{--}2 \mu\text{m}$  is motivated both by a concern that smaller particles could have an impact on the safety and efficacy of biotherapeutics, as well as by an interest in understanding the mechanisms of how larger particles are formed.

In this section, we discuss a number of advanced techniques that are complementary to the more common techniques. Some of these techniques have only recently been developed; others are established but newly applied to particles in biotherapeutics. Some of the techniques discussed for soluble aggregates in the Aggregation chapter/Volume 3, Chapter 5, including electron and atomic force microscopy, are also of interest in characterizing larger particles. All of these tools can be applied, as appropriate, during product, process, and formulation development to help tease out the characteristics of the subpopulations of subvisible particles present in the protein therapeutic being studied.

Although all of the below methods can be implemented with commercial equipment, application of these methods to protein particles is still the subject of much research.

### Resonant Mass Measurement

Prefilled syringes containing drug product often have a thin layer of silicone oil added as a lubricant to aid dispensing, and on storage, silicone oil may leach into the drug product as small droplets. A pervasive challenge in measuring particles in these drug products is distinguishing protein particles from silicone droplets, especially at small sizes ( $< 5 \mu\text{m}$ ), where morphological analysis of flow imaging data is not reliable. RMM is a recently developed technique that is ideal for distinguishing oil droplets from protein particles over a size range of approximately  $0.5\text{--}5 \mu\text{m}$  (38, 39).

RMM directly measures the buoyant mass of a particle,  $\Delta m$ , which is the mass of the displaced matrix fluid minus the mass of the particle. A fraction of the test sample passes through a micro-machined channel on a vibrating cantilever arm. When the buoyant mass is negative, the cantilever mass increases, and the resonance frequency drops. This method provides very clear differentiation of silicone oil droplets versus protein particles or manufacturing impurities because of all common impurities, only silicone oil has a positive  $\Delta m$ . For this reason, RMM is invaluable in characterizing drug products that contain silicone oil droplets. RMM does not have the capability to distinguish between different particle types with negative  $\Delta m$ .

RMM has several practical limitations. An RMM measurement requires flow of sample through a small-diameter channel. The user must take care that the channels and channel entrances are not clogged by adsorbed protein particles, which adhere to surfaces much more readily than polymer beads. For protein particles, conversion of the measured buoyant mass into an effective diameter is hampered by the lack of knowledge of the density of protein particles. The reporting of particle counts by diameter bins is conventional, but there is presently no reason to believe that a particle size distribution is more relevant to product assessment than a particle mass distribution.

Literature studies of protein particles with RMM show reduced numbers of particles at diameters less than approximately 0.5  $\mu\text{m}$  equivalent diameter. This drop-off is likely the result of the limited sensitivity of RMM to particles of this size or smaller with the density of aggregated protein.

### Electrical Sensing Zone Particle Counters

In ESZ particle counters (also termed Coulter counters (48)), particles are detected by pumping the test sample through an orifice, and then monitoring the electrical conductance across the orifice. The presence of a non-conducting particle in the orifice reduces the conductance; the magnitude of the conductance drop is a quantitative measure of the particle volume, from which an equivalent effective particle diameter can be inferred. The most significant advantage of ESZ is that particles may be detected even when the optical contrast approaches zero or in highly opalescent solutions, which may occur in high-concentration protein solutions (27, 41). A disadvantage is that the method works best at ionic conductivities equivalent to 150 mol/L of NaCl, which can be much larger than the conductivities of some formulated protein solutions.

Comparing the particle diameter obtained from an ESZ counter with that obtained by other methods is difficult for protein particles because the particle volume for an irregular, highly hydrated particle is not simply related to the optical scattering or projected area. More work needs to be done to understand the relation between protein particle structure and the ESZ signal and correlate the ESZ observations with microscopy results. A useful validation exercise may be to compare the reported diameters for irregular particles of known density (e.g., an abraded polymer) using both ESZ and RMM. For materials of known density, the particle mass reported by RMM is readily converted to the diameter of an equivalent sphere, which is the same diameter definition used by ESZ.

Particles may be measured over the size range 0.5  $\mu\text{m}$  to 50  $\mu\text{m}$  with ESZ, but for any given orifice, particles may be detected only down to a diameter of approximately 6% of the orifice diameter. Near this lower limit, false positives from electrical noise may be hard to distinguish from true particles (27). Methods to reliably distinguish electrical noise from true particles are needed to take full advantage of ESZ. Compounding this problem, addition of salt to improve the signal-to-noise ratio of the electrical signal may alter the concentration of aggregates or alter the aggregate structure.

ESZ has great value for the measurement of particles in high-concentration or high-opalescence drug product in which particles may be impossible to detect by optical methods. To address the limitations of ESZ, measurements should be conducted with multiple aperture sizes and different ionic conductances to verify that electrical noise, sample dilution, and salt addition do not significantly alter the ESZ results.

## Flow Cytometry

Flow cytometry was developed to identify subpopulations of cells by measuring the scattering of light off of the cells and the simultaneous fluorescence of labels that are specific for certain cell properties (such as the abundance of particular receptors). In these instruments, the sample flows in a thin, focused stream that is traversed by one or more lasers. Detectors measure fluorescence emitted from individual particles labeled with appropriate fluorescent dyes, or measure scattering from particles in the forward or side directions. Recent research has shown that flow cytometry is also applicable to the detection of particles commonly found in biotherapeutics, including both silicone oil droplets and protein particles (4, 49–51). Flow cytometry has two potential advantages. First, it may be possible to detect particles with diameters less than 1  $\mu\text{m}$ . Second, multiple scattering and fluorescence channels can provide a means for distinguishing between different particle subpopulations below the 5  $\mu\text{m}$  limit of flow imaging for morphological analysis. It is also possible to use this technique to collect subpopulations of aggregate for further characterization.

Protein particles have been fluorescently labeled using hydrophobic dyes. Stronger fluorescent intensities can be achieved using dyes that fluoresce upon intercalation in cross-beta-sheet protein structures. In an experiment probing the interactions between protein aggregates and silicone oil (49), it was possible to label the protein covalently with an amine-reactive fluorophore, and label the silicone oil with a non-polar dye. However, this labeling was undertaken prior to combining the protein particles and oil droplets. Thus, the labeling was useful for a controlled study, but techniques to selectively label silicone oil droplets and protein particles in a combined state (e.g., test samples from a stability study) remain beyond the present state-of-the-art.

There are two main challenges in the use of flow cytometry. First, there are no standard methods or reference materials for converting the scattering and fluorescence signals in flow cytometry into absolute measures of particle size or other physicochemical attributes of the detected particles. There are also no standard methods for the optimal labeling of particles with fluorescent dyes.

Research is needed to understand the relative affinities of dyes for different types of particles, as well as the affinity and binding kinetics of dyes for heterogeneous particles (e.g., protein coating an inorganic particle). Research also is needed to ensure that the dyes themselves do not change the nature of the protein particle population.

### **Asymmetric Flow Field-Flow Fractionation Plus Optical Detectors**

AF4 is not a particle detection method itself but is instead a method to fractionate polydisperse particle suspensions by size using carefully designed hydrodynamic flows. Once a polydisperse suspension has been fractionated, particle size and optical properties may be ascertained using a variety of detection techniques, including multi-angle light scattering and DLS. In effect, field-flow fractionation enables the use of several ensemble detection methods that are sensitive to particle sizes below 1  $\mu\text{m}$  but are unable to accurately size unfractionated polydisperse mixtures. Potential advantages of AF4 are the very wide size range of particles measured in a single run, the ability to characterize particles by multiple detection modes, and the ability to collect fractions with different aggregate subpopulations.

In AF4, particles are segregated by size along a porous membrane using fluid flows both parallel and through the membrane. The greatest challenge in developing field-flow fractionation is to optimize the segregation process to provide sufficient size separation while minimizing adsorption of protein particles onto the membrane, which could lead to loss of particles or changes in the particle size distribution (52, 53). Changes in particle concentration (decreased concentration during injection, increased concentration during the separation stage, and dilution during the elution phase) may cause changes in particle size or morphology. Initial attempts have shown that separation of protein particles is possible, although the recovery of protein can be as low as 60%, indicating that membrane adsorption still is a challenge (52, 53). Exposing protein particle suspensions to multiple separation cycles, with measurements of the particle size distribution after each cycle, would assist in understanding the impact of the separation process on the measured size distribution.

### **Nanoparticle Tracking Analysis**

The extent of Brownian motion of small particles depends on the sample viscosity and the hydrodynamic radius of the particle. DLS measures hydrodynamic radius by examining fluctuations in scattering caused by Brownian motion. NTA measures Brownian motion in a much more direct method by tracking either scattered laser light (54) or the fluorescence of single particles (55). Provided that the scattering cross section or fluorescence intensity is high enough, NTA can determine the size of particles below the resolution of optical imaging. Because NTA is fundamentally a single-particle measurement, NTA works well with polydisperse mixtures. Limitations of NTA are that the sampling volume is low and the range of particle concentrations is limited ( $10^6$  to  $10^9$   $\text{mL}^{-1}$ ); at low

concentrations, there are few particles visible, whereas at high concentrations, the Brownian motion tracks of different particles can be indistinguishable.

Advantages of the method are that the measured hydrodynamic radius is approximately comparable to the effective radius obtained by flow imaging methods for larger particles, and that the method is relatively easy to implement relative to flow cytometry or AF4. From this perspective, NTA provides a good general-purpose tool for analysis of particle size distributions below 1  $\mu\text{m}$ , provided that the particle concentrations are within the optimum range. A challenge of the method is to differentiate particle translation motion from the internal motion of flexible protein particles. Another limitation is that at high concentration of protein monomer, the background light scattering increases, and small particles may be more difficult to detect and track. The method may allow categorization of different subpopulations by examining the scattering intensity.

## Conclusions

This chapter has focused on methods for measurement of the size and amount of the subvisible particles in protein therapeutics, especially light obscuration and flow imaging, covering their strengths, weaknesses, and applications. Particle size distribution is an important parameter that can be used to help develop protein candidates, processes, and formulations that minimize particle formation. These techniques can also be used to help understand which particle characteristics are most important in eliciting a biological response. These are important tools to have in the analytical tool box. The compendial method, light obscuration, has the advantage of years of use and extensive historical data. Characterization performed at the earlier stages of development can be used to support the use of a single lot release method, with a clearly defined strategy for further characterization when appropriate. As emerging technology continues to improve and expand, the underlying mechanism of protein aggregation will be better understood. Such advances will likely allow observation of the subpopulations that comprise the total subvisible particle population, which should result in ever more homogeneous drug product.

## References

1. Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G.; Fan, Y. X.; Kirshner, S.; Verthelyi, D.; Kozłowski, S.; Clouse, K. A.; Swann, P. G.; Rosenberg, A.; Cherney, B. *J. Pharm. Sci.* **2009**, *98*, 1201–1205.
2. Narhi, L. O.; Schmit, J.; Bechtold-Peters, K.; Sharma, D. *J. Pharm. Sci.* **2012**, *101*, 493–498.
3. Narhi, L. O.; Jiang, Y.; Cao, S.; Benedek, K.; Shnek, D. *Curr. Pharm. Biotechnol.* **2009**, *10*, 373–381.
4. USP/NF General information <1787>. Particulate Matter in Protein Injections. In *United States Pharmacopeia and The National Formulary (USP–NF)*; United States Pharmacopeial Convention: Rockville, MD, 2014.

5. Zölls, S.; Tantipolphan, R.; Wiggenhorn, M.; Winter, G.; Jiskoot, W.; Friess, W.; Hawe, A. *J. Pharm. Sci.* **2011**, *101*, 914–935.
6. *Analysis of Aggregates and Particles in Protein Pharmaceuticals*; Mahler, H. C., Jiskoot, W., Eds.; Wiley: Hoboken, NJ, 2012.
7. Chi, E. Y.; Krishnan, S.; Kendrick, B. S.; Chang, B. S.; Carpenter, J. F.; Randolph, T. W. *Pharm. Res.* **2003**, *20*, 1325–1336.
8. Mahler, H. C.; Friese, W.; Grauschopf, U.; Kiese, S. *J. Pharm. Sci.* **2009**, *98*, 2909–2934.
9. Chi, E. Y.; Krishnan, S.; Kendrick, B. S.; Chang, B. S.; Carpenter, J. F.; Randolph, T. W. *Protein Sci.* **2003**, *12*, 903–913.
10. Wang, W. *Int. J. Pharm.* **1999**, *185*, 129–188.
11. *Misbehaving Proteins: Protein (Mis)folding, Aggregation, and Stability*; Murphy, R.; Amos, T., Eds.; Springer-Verlag: New York, 2006.
12. Cleland, J. L.; Powell, M. F.; Shire, S. J. *Crit. Rev. Ther. Drug Carrier Syst.* **1993**, *10*, 307–377.
13. Gokarn, Y. R.; Fesinmeyer, R. M.; Saluja, A.; Cao, S.; Dankberg, J.; Goetze, A.; Remmele, R. L., Jr.; Narhi, L. O.; Brems, D. N. *Protein Sci.* **2009**, *18*, 169–179.
14. Bee, J. S.; Randolph, T. W.; Carpenter, J. F.; Bishop, S. M.; Dimitrova, M. *N. J. Pharm. Sci.* **2011**, *100*, 4158–4170.
15. USP/NF General Chapter <788>. Particulate Matter in Injections. In *United States Pharmacopeia and The National Formulary (USP–NF), USP35–NF-30*; United States Pharmacopeial Convention: Rockville, MD, 2012.
16. Rosenberg, A. S.; Verthelyi, D.; Cherney, B. W. *J. Pharm. Sci.* **2012**, *101*, 3560–3567.
17. Kiese, S.; Pappenberger, A.; Friess, W.; Mahler, H-C. *J. Pharm. Sci.* **2008**, *97*, 4347–4366.
18. Fradkin, A. H.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2009**, *98*, 3247–3264.
19. Joubert, M. K.; Luo, Q.; Nashed-Samuel, Y.; Wypych, J.; Narhi, L. O. *J. Biol. Chem.* **2011**, *286*, 25118–25133.
20. Bi, V.; Jawa, J.; Joubert, M. K.; Kaliyaperumal, A.; Eakin, C.; Richmond, K.; Pan, O.; Sun, J.; Hokom, M.; Goletz, T. A.; Wypych, J.; Zhou, L.; Kerwin, B. A.; Narhi, L. O.; Arora, T. *J. Pharm. Sci.* **2013**, *102*, 3545–3555.
21. Philo, J. S. *Curr. Pharm. Biotechnol.* **2009**, *10*, 359–372.
22. Vörös, J. *Biophys. J.* **2004**, *87*, 553–561.
23. Zölls, S.; Gregoritzka, M.; Tantipolphan, R.; Wiggenhorn, M.; Winter, G.; Friess, W.; Hawe, A. *J. Pharm. Sci.* **2013**, *102*, 1434–1446.
24. Knollenberg, R. G.; Gallant, R. C. In *Proceedings of the International Conference on Particle Detection, Metrology and Control*; Institute of Environmental Sciences: , 1990; pp 154–182.
25. Ripple, D.; Wayment, J. R.; Carrier, M. J. *Am. Pharm. Rev.* **2010**, *14*, 90–96.
26. Sharma, D. K.; Oma, P.; Pollo, M. J.; Sukumar, M. *J. Pharm. Sci.* **2010**, *99*, 2628–2642.
27. Barnard, J. G.; Rhyner, M. N.; Carpenter, J. F. *J. Pharm. Sci.* **2012**, *101*, 140–153.
28. Yahiaoui, S.; Feuillebois, F. *J. Fluid Mech.* **2010**, *662*, 447–474.

29. Einarsson, J.; Johansson, A.; Mahato, S. K.; Mishra, Y. N.; Angilella, J. R.; Hanstorp, D.; Mehlig, B. *Acta Mech.* **2013**, *224*, 2281–2289.
30. Jennings, B. R.; Parslow, K. *Proc. R. Soc. London, Ser. A* **1988**, *419*, 137–149.
31. Garidel, P.; Herre, A.; Kliche, W. In *Analysis of Aggregates and Particles in Protein Pharmaceuticals*; Mahler, H. C., Jiskoot, W., Eds.; Wiley: Hoboken, NJ, 2012; pp 269–302.
32. Garidel, P. *Spectrosc. Eur.* **2013**, *25*, 19–22.
33. Cao, X.; Wen, Z. Q.; Vance, A.; Torraca, G. *Appl. Spectr.* **2009**, *63*, 830–834.
34. Kiese, S.; Pappenberger, A.; Friess, W.; Mahler, H.-C. *J. Pharm. Sci.* **2010**, *99*, 632–644.
35. Cao, S.; Jian, Y.; Narhi, L. I Forum (PF). United States I Convention. <http://www.usp.org/usp-nf/pharmacopeial-forum>.
36. Strehl, R.; Rombach-Riegraf, V.; Diez, M.; Egodage, K.; Bluemel, M.; Jeschke, M.; Koulov, A. V. *Pharm. Res.* **2012**, *29*, 594–602.
37. Vandesteeg, N.; Kilbert, C. *J. Pharm. Sci.* **2013**, *102*, 1696–1700.
38. Patel, A. R.; Lau, D.; Liu, J. *Anal. Chem.* **2012**, *84*, 6833–6840.
39. Weinbuch, D.; Zölls, S.; Wiggenhorn, M.; Friess, W.; Winter, G.; Jiskoot, W.; Hawe, A. *J. Pharm. Sci.* **2013**, *102*, 2152–2165.
40. Liu, L.; Randolph, T. W.; Carpenter, J. F. *J. Pharm. Sci.* **2012**, *101*, 2952–2959.
41. Demeule, B.; Messick, S.; Shire, S. J.; Liu, J. *AAPS J.* **2010**, *12*, 708–715.
42. USP/NF General Chapter <787>. Particulate Matter in Therapeutic Protein Injections. In *United States Pharmacopeia and The National Formulary (USP–NF), USP37–NF-32*; United States Pharmacopeial Convention: Rockville, MD, 2014.
43. Weinbuch, D.; Jiskoot, W.; Hawe, A. *AAPS J.* **2014**, *16*, 1128–1131.
44. Singh, S. K.; Afonina, N.; Awwad, M.; Bechtold-Peters, K.; Blue, J. T.; Chou, D.; Cromwell, M.; Krause, H.-J.; Mahler, H.-C.; Meyer, B. K.; Narhi, L.; Nesta, D. P.; Spitznagel, T. *J. Pharm. Sci.* **2010**, *99*, 3302–3321.
45. Zölls, S.; Weinbuch, D.; Wiggenhorn, M.; Winter, G.; Friess, W.; Jiskoot, W.; Hawe, A. *AAPS J.* **2013**, *15*, 1200–1211.
46. *Particle Size Analysis — Image Analysis Methods, Part 2: Dynamic Image Analysis Methods*; ISO (International Organization for Standardization): Geneva, 2006; ISO 13322-2.
47. *Particle Size Analysis — Image Analysis Methods, Part 1: Static Image Analysis Methods*; ISO (International Organization for Standardization): Geneva, 2004; 13322-1.
48. Certain commercial equipment, instruments, or materials are identified in this document. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.
49. Ludwig, D. B.; Trotter, J. T.; Gabrielson, J. P.; Carpenter, J. F.; Randolph, T. W. *Anal. Biochem.* **2011**, *410*, 191–199.
50. Mach, H.; Bhambhani, A.; Meyer, B. K.; Burek, S.; Davis, H.; Blue, J. T.; Evans, R. K. *J. Pharm. Sci.* **2011**, *100*, 1671–1678.

51. Nishi, H.; Mathäs, R.; Fürst, R.; Winter, G. *J. Pharm. Sci.* **2014**, *103*, 90–99.
52. Cao, S.; Pollastrini, J.; Jiang, Y. *Curr. Pharm. Biotechnol.* **2009**, *10*, 382–390.
53. Hawe, A.; Romeijn, S.; Filipe, V.; Jiskoot, W. *J. Pharm. Sci.* **2012**, *101*, 4129–4139.
54. Filipe, V.; Hawe, A.; Jiskoot, W. *Pharm. Res.* **2010**, *27*, 796–810.
55. Filipe, V.; Poole, R.; Kutscher, M.; Forier, K.; Braeckmans, K.; Jiskoot, W. *Pharm. Res.* **2011**, *28*, 1112–1120.



## Chapter 9

# Analytical Methods for the Measurement of Host Cell Proteins and Other Process-Related Impurities

Kesh Prakash<sup>\*,1</sup> and Weibin Chen<sup>2</sup>

<sup>1</sup>Emergent BioSolutions, 300 Professional Drive,  
Gaithersburg, Maryland 20879, United States

<sup>2</sup>Waters Corporation, 34 Maple Street,  
Milford, Massachusetts 01757, United States

\*E-mail: PrakashK@ebsi.com

Host cell proteins (HCPs) and other process-related impurities such as Protein A, insulin, and residual host cell DNA are routinely encountered during the manufacture of biopharmaceutical products. HCP impurities are of significant interest to the biopharmaceutical industry because therapeutic proteins (monoclonal antibodies [mAbs] and recombinant proteins) need to be highly purified from the production host cell line that is utilized to manufacture them. It is imperative that HCPs are removed during the purification (downstream) process and brought down to as low a level as possible in the final drug substance. This is done to ensure product purity and safety. Similarly, clearance of other process-related impurities (host cell DNA, protein A, and insulin) must be demonstrated during the manufacture of biopharmaceutical products. This chapter will discuss analytical methods that cover the detection and clearance of these impurities. In addition, the application of these analytical methods to the quantitation of various process-related impurities in the submitted NISTmAb is described. The results obtained show that the test sample is pure and would meet specification requirements.

## General Introduction to Host Cell Proteins (HCPs) and Other Process-Related Impurities

The production of a biopharmaceutical such as a monoclonal antibody (mAb) involves a variety of steps, typically divided into upstream and downstream processes (1). From any stage of the process, process-related impurities, which can include raw materials, host cell components, media components, leachables from the type of purification process used, and chemical additives, may be derived (2). For example, low levels of process-related impurities such as HCPs and residual DNA (from the host cell) may persist during various stages of the biopharmaceutical manufacturing process. Downstream processing steps intended for removal of such impurities (e.g., Protein A that may leach from the Protein A resin) may also be introduced. Based on the manufacturing process used, other cell culture media-derived impurities (e.g., insulin, bovine serum albumin [BSA], transferrin,  $\beta$ -glucans) also may be present and will need quantification (2). Although there are a number of such potential impurities to consider, this chapter will focus mainly on HCPs, Protein A, insulin, and residual DNA.

Cell expression system-derived impurities will vary, depending on the host cell used in the production process. Biopharmaceutical products are generally produced utilizing recombinant technology in host cells ranging from bacteria (e.g., *Escherichia coli* [*E. coli*]), yeast (e.g., *Pichia pastoris*), or cell lines of various origins such as mammalian (e.g., Chinese hamster ovary [CHO] cells), insect (e.g., Sf9 cells), or even plant (e.g., tobacco). Host cell-derived proteins or HCPs arise as a result of cell lysis or secretion during cell culture and harvesting, and they are released into the medium in addition to the product. HCPs are a complex set of proteins that need to be monitored during the development and commercialization of biopharmaceutical products. It is essential that HCPs be removed during the purification process and brought down to as low a level as possible in the final drug product to ensure product purity and patient safety (2, 3). HCPs should be few in number and low in quantity. One of the important properties of HCPs is that they are specific and unique to the host cells and the particular culture process used for manufacture (4, 5). HCPs can vary in isoelectric point (pI) (~3–11) and hydrophobicity, and display a broad range of molecular weights (5 kD to > 250 kD), depending on the host cell and manufacturing process utilized. The number of HCPs in conditioned media (upstream) samples can vary from several hundred (e.g., *E. coli*) to more than a thousand proteins (e.g., CHO cells), depending on the host cell system used (4, 5). Most of the experience gathered to date has been obtained utilizing *E. coli* and mammalian cells like CHO and NS0 (2, 5).

Additional impurities must also be considered. Purification-derived impurities include host cell DNA. Low levels of host cell DNA originating from cellular DNA also may remain following purification (3, 6).

Cell culture and harvest-derived impurities usually consist of raw materials like insulin and others added as growth media supplements for stimulating cell growth. The purification process must be designed appropriately to bring these impurity levels down. In addition, downstream purification process-derived

impurities such as Protein A may be present as a result of leaching from Protein A resin (2) used as the capture column during purification of mAbs.

The goal across the biopharmaceutical industry is to develop a production process that results in low levels of all process-related impurities—HCPs, host cell DNA, Protein A, insulin, and so forth. All of this is done to ensure product purity and safety prior to administration of these products to patients.

## Regulatory Guidance

There is no clear regulatory guidance on HCP, Protein A, and insulin levels. However, most biopharmaceutical products reviewed by the U.S. Food and Drug Administration (FDA) contain enzyme-linked immunosorbent assay (ELISA)-based HCP levels of 1 to 100 ng/mg of product (3, 7–10). The ability to set a single numerical limit for an appropriate level of HCP for all products is difficult because the HCPs present can vary from process to process and product to product, and the sensitivity of detection and quantitation is closely linked to the quality of the immunoreagents from which the assay is constructed.

According to both U.S. (FDA) and European Union (European Medicines Agency [EMA]) regulations, HCPs are labelled as process-derived impurities, and guidance to lower their presence by the use of appropriate well-controlled manufacturing processes is given. In particular, 21 CFR 610.13 (6) states that biological products be “... free of extraneous material except that which is unavoidable...,” while ICH Q6B guidance (10) specifically states that biologic product manufacturers should evaluate impurities which may be present and develop meaningful acceptance criteria for the impurities based on their preclinical and clinical experience. However, this guidance also recognizes the challenges that exist with HCP assay-specific data and product- and process-specific reagents. Per ICH Q6B (10), “The absolute purity of biotechnological and biological products is difficult to determine and the results are method-dependent.” Along the same line, the EMA (11) has stated that “... standardization of the analytical methods would be problematic since the reagents used are product and production system-related.”

Although FDA guidance describes the need to test for HCPs, it also allows, through validation, removal of testing requirements once clearance has been shown to be uniform and consistent. An appropriately conducted clearance study carried out as part of process validation can be an acceptable alternative for production lot-to-lot testing (12) even in cases where low quantities of HCPs are present in the drug substance. Additionally, the FDA (10) states that “[i]f impurities are qualitatively and quantitatively (i.e., relative amounts and/or concentrations) the same as in the drug substance, testing of the drug product is not necessary.”

As far as the process-related impurities Protein A and insulin are concerned, there is no defined guidance by regulatory agencies (e.g., FDA, EMA). The general rule of thumb that is followed across the biopharmaceutical industry is to develop and optimize processes to bring all process-related impurities (HCPs, Protein A, and insulin) to as low a level as possible (10, 12).

With regard to host cell DNA, FDA requirements in the past stipulated an upper limit of 100 pg per therapeutic patient dose (13). This has now been replaced by World Health Organization (WHO) guidelines that have an upper limit of 10 ng/patient dose (14). This guideline is followed for products produced in mammalian and non-mammalian systems.

## **Immunogenicity Pertaining to HCPs; Quality Issues Related to HCPs and Other Impurities**

### **HCPs**

mAb products are large molecules that can bind to specific targets. mAbs could be considered to have a potential for immunotoxicity, which is reflected in the clinical experience accumulated on mAbs-induced adverse effects related to immunosuppression, immunostimulation and hypersensitivity (immunogenicity). Moreover, it is well known that anti-drug antibodies can be induced when immune systems recognize a material (e.g., mAb) introduced as foreign, especially for parenterally administered products. Biopharmaceuticals can elicit product-specific anti-drug antibodies (ADA), or antibodies against HCPs can develop in either nonclinical or clinical studies.

Anti-HCP antibodies could cause adverse events by inducing immune-mediated clinical sequelae such as injection site reactions, flu-like symptoms, and at worst case, anaphylaxis (15). Immunogenicity also may be related to the state of a patient receiving such therapy, as pre-existing antibodies to HCPs have been identified in individuals with no known exposure to biopharmaceutical products (16). In addition, it has been observed that HCPs may act as adjuvants and enhance an ADA immune response to the product itself (3, 12).

Although the risk of immunogenicity related directly or indirectly to HCPs is not well published, there are some cases, and the general perception is that it is a risk to be considered. Proteases are one of the types of HCPs that may indirectly influence immunogenicity to the product. Proteolytic HCPs have the potential to digest the desired protein product over time, thus altering biological potency, bioavailability, or—through the creation of immunogenic sequences—inducing an ADA response (17). Similar reports of HCP-related ADAs in biosimilars have appeared recently (18).

### **Insulin**

In terms of human safety, high doses of insulin (100–200  $\mu$ IU/kg) is associated with hypoglycemia, coma, convulsions, and even death in very severe cases. There are at least three recombinant human insulin products that have been used or are in use in humans, including Velosulin BR<sup>®</sup>, Humulin<sup>®</sup> R, and Exubera<sup>®</sup> (19). There are no clear reproductive toxicity or carcinogenicity warnings in the labels of these approved products. As no harmful activity has been attributed to insulin so far across products commercialized by the biopharmaceutical industry, it is not expected to have biological or safety effects at the trace levels that may be present

in non-insulin drug products due to residual raw material insulin process-related impurity.

## Protein A

Some published information is available on the toxicity risk attributable to Protein A impurity (20, 21). As Protein A is obtained from the cell wall of the bacterium *Staphylococcus aureus*, a test for bacterial endotoxin is carried out on every biopharmaceutical product before it is released.

## DNA

For host cell DNA, it has been generally shown that DNA with unmethylated cytosine-guanosine (CpG) dinucleotide motifs (i.e., unmethylated dinucleotides of cytosine and guanine) may be immunostimulatory. As mammalian DNA (e.g., NS0 cells) is known to be methylated at 60 to 90% (22), DNA methylated at this level has a very low risk for immunogenicity. In contrast, bacterial DNA (e.g., *E. coli*) possesses immunostimulatory potential as DNA-containing fractions mediate immune modulation. It has been shown that this is due to the relative abundance of unmethylated CpGs in bacterial DNA (23). Additionally, DNA fragments observed in host cell DNA are usually smaller than 200 base pairs (bp) and therefore provide substantial safety margins pertaining to oncogenicity and infectivity for biopharmaceutical products that meet the 10 ng DNA/patient dose limit (24, 25). Reduction of DNA fragment size (< 200 bp) reduces the probability that intact oncogenes or viral infectious sequences would be present.

## Considerations for Manufacturing Consistency and HCP Clearance

A biopharmaceutical product is generally secreted from mammalian cells into the surrounding cell culture medium in a manner similar to secreted native host proteins. When cells die during cell expansion in bioreactors (fermentation), soluble proteins from within the cells may be released into the cell culture media. During cell harvest, cells lyse due to shear stress and may result in increased levels of HCPs.

As clarified cell culture media containing HCPs is taken through the downstream purification process, each purification step (consisting of various types of resins, nanoparticle/virus filters, etc.) affects the clearance of HCPs in different ways. The intrinsic features of the product, the isolation procedure, and the downstream purification process are all factors that influence HCP levels. The particular properties of the HCP may result in its co-purification with the product, in rare cases through direct association (26). Therefore, a well-designed and properly scaled up purification process is imperative to help reduce the HCPs to the lowest levels possible. Once the process is optimized, process characterization and validation studies are necessary to demonstrate removal of HCP by each process purification step. Moreover, it is essential to evaluate

the robustness of each of these steps to show reliable and consistent removal of HCPs during manufacturing. Well-developed and optimized HCP assays play a major role in ensuring consistency of manufacturing runs. In the end, reliable, robust, and reproducible HCP assays are used to quantitate the level of residual HCPs remaining in the final drug product that is to be administered to patients. A downward trend of HCP clearance, beginning with the cell culture medium all the way through the drug substance/drug product using these assays must be established. HCP levels should be well documented for pre-clinical lots used in toxicology studies, in lots used for clinical trials, and in process validation of the final commercial process.

## Methods for Quantitation of HCPs and Other Process-Related Impurities

The methods currently used for the measurement of HCPs and host cell DNA are based on a sandwich ELISA (2, 27) format and real-time quantitative polymerase chain reaction (qPCR), respectively. PicoGreen<sup>®</sup>, a fluorescent nucleic acid stain, is used for host cell DNA quantitation, but this technique has the disadvantage of including the contribution of single-stranded nucleic acids (RNA) and free nucleotides to the measurement signal. In addition, the PicoGreen<sup>®</sup> method is less sensitive than the Threshold<sup>™</sup> or qPCR methods. Therefore, the PicoGreen<sup>®</sup> method is not often used for quantitation unless it is being used for quick, initial screening purposes.

There are both immunological (ELISA) and non-immunological methods (liquid chromatography-tandem mass spectrometry [LC-MS/MS], LC-MS HCP chapter/Volume 3, Chapter 13) to detect and quantitate HCPs (28, 29). Immunological methods are the main methods of choice for measurement of the impurities such as HCP, Protein A, and insulin as they are relatively easy to perform, are quantitative due to their measurement being based on direct antigen (e.g., HCP)-antibody interaction, and provide good sensitivity. The workhorse for HCP quantitation has been the sandwich immunoassay (plate ELISA format, Figure 1), based on polyclonal antibodies raised against a wide range of HCPs. This assay offers a combination of accuracy, precision, reproducibility, sensitivity, relatively high throughput, potential for automation, quantitative value, quick turn-around if properly planned, and a fairly low cost per assay. The polyclonal anti-HCP antibodies are used for both capture and detection, with the detector antibodies being biotin-labelled. Then, a streptavidin-peroxidase conjugate is used followed by the addition of a specific substrate. Absorbance is read in a microplate reader at a specific wavelength. HCP quantities are regressed from the standard curve generated by the software used and corrected for dilution factors. Sample values are reported as the mean concentration of the dilutions yielding spike recovery values (usually 75–125%) in samples containing defined amounts of HCP antigen spiked in them. The  $\pm 25\%$  recovery is generally followed throughout the industry.

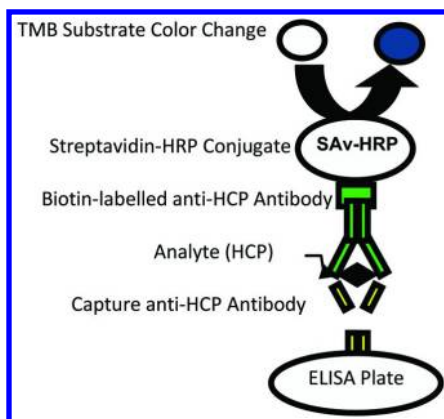


Figure 1. Sandwich enzyme-linked immunosorbent assay (ELISA) format. HCP = host cell protein, HRP = horseradish peroxidase, TMB = 3,3',5,5'-tetramethylbenzidine.

Other immunoassay formats, such as the direct and competition immunoassay methods, are not suitable because they lack the higher sensitivity afforded by the sandwich ELISA. It is pertinent to point out here that the polyclonal anti-HCP reagents used in the sandwich ELISA (plate method) cannot recognize all of the possible HCP species because some HCP species can either be weakly immunogenic or have no immunogenic potential at all. The percent recognition of all possible HCP species present is defined as percent coverage. The methods for coverage determination are challenging and results can vary widely from analysis to analysis, limiting the quantitative power of this technique. It is generally accepted that “good” polyclonal antisera usually react with more than 50% of the HCPs present. Orthogonal methods for assessing product purity may need to be developed. These may involve the use anti-HCP antibodies (in 2-dimensional [2-D] gel western blots) or may be non-antibody based (LC-MS/MS methods, LC-MS HCP chapter/Volume 3, Chapter 13). Non-immunological LC-MS methods (28, 29) can be used to provide increased assurance of product purity and safety. For example, it could be used for identification and/or quantitation of the HCP impurity when present in high abundance, or to identify HCPs for which antibodies are not easily raised for use in the HCP ELISA. Knowledge of the identity of the HCP impurity is gathered and a quantitative method for that particular protein impurity could then be developed if it is deemed immunogenic or risky. Orthogonal LC-MS methods are an emerging technology that shows great promise to provide complementary information to ELISA, and are expected to play an increasing role in biopharmaceutical development as described in the LC-MS HCP chapter/Volume 3, Chapter 13.

Protein A and insulin also are commonly measured using ELISA. With regard to the sandwich ELISA formats for quantitation of Protein A and insulin, the interest is focused on a specific protein rather than the entire proteome of a given cell line. Therefore, a majority of the industry uses commercially available antibodies such as polyclonal anti-Protein A antibody and anti-human

insulin antibody, respectively, for capture and appropriate biotinylated antibodies for detection. These assays are mostly developed in-house using plate ELISA formats.

## Characterization and Development of Reagents with Reference to HCP Critical Reagents

### Critical Reagents

HCPs are complex proteins numbering anywhere from several hundred (cell lysate from bacteria) to more than a thousand (mammalian cell culture supernatant) species of proteins (4, 5), varying in molecular weight, pI, hydrophobicity, and potential for being immunogenic. The HCP antigen generated will be used as the reference standard in HCP assays, and most importantly, this HCP antigen will be used as the immunogen to raise anti-HCP polyclonal antibodies in various animals (rabbits, sheep, or goats). The antigen is therefore paramount to producing an anti-HCP panel having good reactivity to as many HCP antigen species as possible. The HCP antigen and the anti-HCP polyclonal antibodies are the most critical reagents for the sandwich ELISA used for quantitating HCPs.

### Preparation of the HCP Antigen

The HCP antigen preparation is typically made from null cells that do not contain the gene for the target or the product of interest (30). Two methods may be used. Method 1 consists of utilizing the non-transfected parental cell line that was expanded to create the production cell line (4). A non-transfected cell line does not express selection markers or other genes coded for by the expression plasmid. In method 2, mock-transfected cells can be used as null cells to produce HCP antigen (5). These cells are created by transfecting the parental cell line with an empty or blank plasmid that was used to create the production cell line but missing the gene for the actual product. Once the null cells have been identified, the HCP antigen is prepared by a mock production run. The prevalent approach is to use a platform cell culture process that is used for several products, such as monoclonal antibodies, from the same cell line. When a platform cell culture process is used, it is very minimally processed in order to ensure that the material contains a broad range of HCPs. For example, the conditioned media from the bioreactor fermentation run containing all the secreted or released HCPs is used as the HCP antigen source as this contains all the HCPs possibly originating from that particular cell culture process. Therefore, this allows for the use of the antigen for HCP monitoring of multiple products generated from the same platform upstream process using mammalian cell lines (e.g., CHO cells).

### Preparation of Anti-HCP Polyclonal Antibodies

A wide range of animal species are used to generate anti-HCP antisera. The HCP antigen (immunogen), derived from either bacterial or mammalian cells, is combined with an adjuvant (complete or incomplete Freund's adjuvant) before



immunization of the animals (most commonly rabbits, goats, or sheep) (2). During the immunization, each animal receives a priming immunization followed by several (4 to 6) booster shots to produce high titer, high affinity/avidity polyclonal antibodies. Antiserum is collected from each animal prior to the first priming immunization (control bleed) and at certain intervals (~2 weeks) subsequent to the booster shots (sample bleeds). Normally, four to six bleeds are collected from each animal. The test samples obtained from the animal antisera are assessed by a titer immunoassay and western blot to evaluate the antiserum with regard to antibody affinity, specificity, degree of immunoreactivity, and titer. The titer is usually identified as the dilution of antibody yielding a signal at least three- to four-fold higher than the non-specific binding signal seen in the titer immunoassay. High titer responses are typically in the 1:1000 or 1:2000 dilution ranges, whereas low titers are below the 1:500 dilution range. It is critical to assess the level of immunoreactivity of each test bleed. Test bleeds are combined to ensure that the antisera included in the final pool display immunoreactivity to most of the HCP proteins present during the product manufacturing process and do not contain antibodies that react with the product. It is important to perform titer and western blot analysis of the each test bleed prior to combining high titer antisera, and this allows one to also remove sera samples with low titer. Once fully characterized, appropriate sera samples are combined and the total antibody fraction purified. Antibody purification is usually carried out by Protein A or Protein G and/or HCP column affinity chromatography.

The choice of use of Protein A or Protein G affinity ligand depends on the animal species from which the antibodies were generated. Anti-HCP antibodies produced in rabbits can be purified on either Protein A or Protein G affinity columns. In general, goat or sheep antibodies are purified using a Protein G resin column. Some use HCP affinity columns to specifically purify the anti-HCP antibody fraction. Both purification methods (Protein A/G, HCP affinity) are in common usage, sometimes in combination.

### **Characterization of Anti-HCP Polyclonal Antibodies**

The affinity-purified antibodies are used as both capture and detector antibodies in the HCP sandwich ELISA. If prepared and stored properly as a large batch (frozen at  $-70^{\circ}\text{C}$ ) and used efficiently, these antibodies should last for many years (5–10 years). Initially, the protein concentration of the batch is measured by  $A_{280}$  or bicinchoninic acid assay (BCA) or with any other suitable method. The correct antibody pair (coating or capture antibody and biotin-labelled detector antibody) for the HCP sandwich ELISA is finalized by testing HCP-containing samples obtained from the unit operations of the appropriate purification process. The sensitivity and specificity are assessed by ensuring non-reactivity with the product of interest.

One of the methods for characterization of anti-HCP polyclonal antisera is to carry out 1-dimensional (1-D) gel western blot analysis initially to get a gross picture of HCP binding efficiency, followed by a confirmatory 2-D gel western blot analysis using the HCP immunogen used to raise the anti-HCP polyclonal antibodies (Figure 2). In the first dimension, the HCP immunogen is separated

into various protein species based on its pI, whereas in the second dimension, it is lined up based on the molecular weight of each protein component. More than 1200 HCP species are present, as shown by silver staining, which is a comprehensive method that labels all protein species present (Figure 2A). Western blot analysis is then done using the anti-HCP polyclonal antibodies raised (Figure 2B). Comparison of the two gels allows one to estimate the coverage (percent coverage) of the raised antibodies. In this particular case, the percent coverage estimated was ~75%. The dark spots (black) in Figure 2B show the HCP species that are recognized and bound by the anti-HCP polyclonal antibodies. Some HCP species of higher molecular weight (100–150 kD) and higher pI values (8–10) seen in Figure 2A are not recognized by the anti-HCP antisera and are therefore not bound (Figure 2B). The percent coverage of HCPs helps to ascertain the quality of the antibodies for the HCP immunogen used. As these coverage analyses are quite complex and require a high level of skill to execute, it is imperative to do this on large-format gels at least a few times and derive an average value. The large-format gels (18 cm or longer) have higher resolution and are therefore more meaningful than small format gels (< 10 cm). Although there is no upper limit to this percent coverage value, the higher the number, the better the coverage of HCPs by that particular batch of antibodies generated. The percent coverage analysis is an important measure of the quality of the anti-HCP polyclonal antibodies. The 2-D gel procedure is prone to artifacts (e.g., HCPs are denatured and may not reflect those bound in solution in ELISA) and highly variable. For this reason, they see their main application as a qualitative comparison of antibody batches as part of an overall evaluation. It is important to emphasize that the coverage estimated by 2-D gel western blots is only approximate because the exact value is very technique-dependent (i.e., having low reproducibility) and is influenced by the amount of protein loaded on the gel and the exact way the western blotting is carried out.

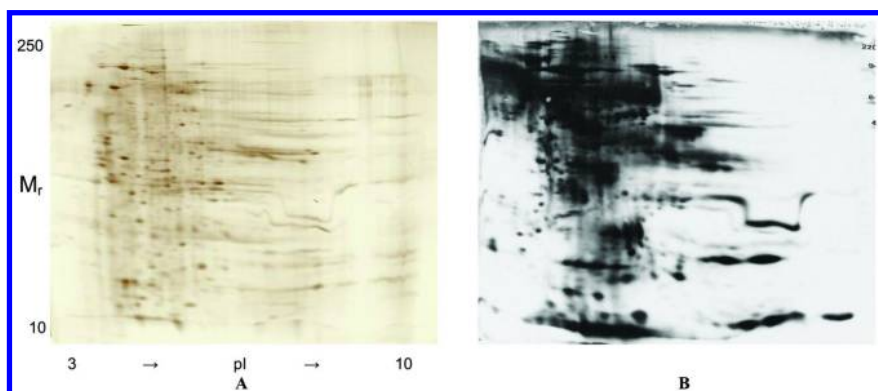


Figure 2. Two-dimensional gel western blot analysis. (A) Host cell proteins (HCPs) seen after silver staining. (B) HCPs recognized and bound by anti-HCP polyclonal antibodies.

## Good Manufacturing Practice (GMP) Testing

As previously stated in this chapter, all of the HCPs present in the HCP immunogen used to raise polyclonal antibodies are not necessarily immunogenic because some could be weakly immunogenic and some not immunogenic at all. Therefore, during the assessment of product purity, the sandwich HCP ELISA may not detect and accurately quantify all of the potential HCPs present at the end of the product purification process. The ELISA is therefore used as only one part of a system of evaluation of product purity. There are other orthogonal methods, such as sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels, reversed phase-high-pressure liquid chromatography (RP-HPLC), and LC-MS/MS, which could be used to identify the hard-to-quantitate HCPs. Efforts should be made to identify those unique HCPs that have a predilection to be co-purified with product. As a result, HCP ELISAs used for product purity assessment are usually part of the GMP control system. It is generally carried out as an in-process test, but some companies across the biopharmaceutical industry also perform a Certificate of Analysis (C of A) test. The HCP ELISA needs to have appropriate alert and reject limits in place. Orthogonal methods are most useful as characterization tools when demonstrated to be scientifically sound.

### Management of HCP Assay Critical Reagents through the Life Cycle of the Product

The most critical reagents for the HCP assay are the HCP antigen and anti-HCP polyclonal antibodies. When stored frozen and properly validated for bench stability, HCP reagents are generally stable for many years. However, as stocks are depleted, it is essential to have a process and plan for their replacement. The general rule-of-thumb is to prepare a high volume of anti-HCP antibodies soon after the HCP immunogen/calibration standard is prepared. This is the standard that is prepared initially from the null or mock transfected cells. When the supply is exhausted, a new pair of reagents (HCP antigen and anti-HCP antibodies) should be generated following the same procedure that was used for the previous set of reagents in order to maintain consistent performance of the new and previous HCP assays. Often across the industry, the manufacturing process is modified to improve the yield of the biopharmaceutical product. In such a situation, the original HCP assay reagents might have been produced according to the previous or old process that is not similar to the newly developed process. The old process reagents may not yield HCP values similar to those obtained using the new pair of reagents (HCP antigen and anti-HCP antibodies). If the upstream cell culture process has changed significantly, a new reagent may be required. The new reagent performance should be evaluated for accuracy again in the HCP assay using spike recovery, precision/reproducibility, sensitivity, range of detection, lack of interference from various sample matrices, and linearity in all in-process samples and the final drug substance. Some laboratories in the industry also use a technique called 2D-difference gel electrophoresis (DIGE), as it allows for a direct comparison of HCP populations between new and old processes (31, 32). In 2D-DIGE, different samples (up to 5) can be simultaneously analyzed and

compared on the same gel by labeling with CyDyes. This helps in evaluating if the changes between the two processes are large enough to merit going through the development of a new HCP assay. It is important to demonstrate that the new HCP assay is appropriate by showing that HCP clearance (fold clearance from step-to-step), beginning with the conditioned media (step 1, harvest) all the way down to the drug substance (final step), is similar or better than that observed with the old HCP assay.

In order to achieve this goal, the new HCP antigen should be prepared using a method similar to that used for making the old or previous version of the HCP antigen. After measuring the protein concentration of the new HCP antigen, a comparison of the new and old HCP standards should be carried out by SDS-PAGE gel characterization to demonstrate similarity between the two.

With regard to the anti-HCP polyclonal antibody preparation, it is better to use the same animal species that had been used for raising the old antibody reagent. The new anti-HCP polyclonal antibodies need to be characterized by 2-D western blot coverage analysis, as described above. In particular, a side-by-side comparison of the old and new anti-HCP polyclonal antibodies using the same sandwich HCP ELISA format is essential to show the suitability of the new reagents in the new HCP assay. Assay qualification using appropriate parameters (accuracy; precision, including repeatability and reproducibility; range; linearity; and sensitivity) should be performed. Bridging studies also should be performed using several samples from step 1 (harvest), all in-process steps, and the final drug substance. Identical samples should be tested in the old and new HCP sandwich ELISAs, and data obtained should be compared. The anti-HCP polyclonal antibodies generally are very stable when stored frozen ( $-80^{\circ}\text{C}$ ), and maintaining storage integrity throughout the life cycle of the product it is intended for is critical.

## **NIST Sample: Assay Protocols and Summary of Results**

The NISTmAb described throughout the current series is intended as an analytical reference material that mimics a typical biopharmaceutical drug substance. Therefore, to serve as a representative material, it should be free of process-related impurities to a level expected for a typical drug substance. For this reason, a subset of process-related impurity methods were applied to the NISTmAb sample to evaluate HCPs, Protein A, insulin, and residual DNA levels.

### **Sample**

The NISTmAb sample, received from NIST, was described to be a purified, humanized IgG1 $\kappa$  mAb produced in murine myeloma suspension cells. The antibody was at a concentration of 100 mg/mL in 12.5 mM L-His, 12.5 mM L-His HCl (pH 6.0).

## Assay Protocols

The protocols used for quantitation of HCP and the other process-related impurities (Protein A, insulin, and host cell DNA) in the sample (IgG1 $\kappa$  mAb, # 203972) sent by NIST are shown in brief below:

### *HCP Assay*

HCP antigen standard was derived from the HCP antigen (immunogen) from an in-house murine cell line in order to mimic as closely as possible the profile one may expect from the NISTmAb. HCP antigen (ranging from 125 ng/mL to 5 ng/mL), the NISTmAb sample (in serial dilutions, 1:10, 1:50, 1:100, etc.) and controls (samples with no NISTmAb and positive controls) were added to an ELISA plate coated with anti-HCP polyclonal antibodies (capture antibody). Following incubation, biotinylated anti-HCP antibody (detector antibody) was added and incubated. Next, a streptavidin-peroxidase conjugate was added and incubated. Finally, the substrate ABTS (2,2'-azinobis[3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt) was added and incubated before stopping the reaction with the addition of ABTS stop solution. The absorbance was measured at 405 nm using a microplate reader. HCP level present in the NIST sample was regressed from the standard curve generated by the software and corrected for the dilution factor. The sample HCP concentration was reported as the mean concentration of the dilutions yielding spike recovery values of 75 to 125%. This has been discussed under "Methods for quantitation of HCPs and other process-related impurities" earlier in this chapter. The HCP level in the sample, after correction based on a mAb product concentration of 100 mg/mL, was calculated to be 1.8 ng/mg (Table 1). System suitability requirements of the standard operating procedure (SOP) used for this assay were successfully met. In general, system suitability requirements pertain to obtaining a correlation coefficient (R) of  $\geq 0.990$  with the HCP antigen standard curve, negative control showing a value below the lower limit of quantitation (LLOQ), and the positive control (1000 ng/mL) to fall within 750 to 1250 ng/mL in the HCP assay.

### *Protein A Assay*

Standards (ranging from 12.5 to 0.1 ng/mL to generate a standard curve), the NISTmAb sample, and controls were added to an ELISA plate coated with commercially available polyclonal anti-Protein A antibody (capture antibody). Following incubation, biotin-labelled anti-Protein A antibody (detector antibody) was added and incubated. Next a streptavidin-peroxidase conjugate was added and incubated. Finally, the colorimetric substrate ABTS was added and incubated before stopping the reaction with the addition of ABTS stop solution. The absorbance was measured at 405 nm using a microplate reader. The Protein A level in the NIST sample was regressed from the standard curve generated by the

software and corrected for the dilution factor. Sample Protein A concentration was reported as the mean concentration of the dilutions yielding spike recovery values of 75 to 125%. The Protein A level in the sample, after correction based on a mAb product concentration of 100 mg/mL, was calculated to be < 0.01 ng/mg (Table 1). System suitability requirements of the SOP used for this assay were successfully met. In general, system suitability requirements pertain to obtaining an R value of  $\geq 0.995$  with the Protein A standard curve, negative control showing a < LLOQ value, and the positive control (2.5 ng/mL) to fall within 1.875 to 3.125 ng/mL in the Protein A assay.

**Table 1. NIST Sample Results in Process-Related Impurity Assays**

<i>Process-Related Impurity Method</i>	<i>Assay System Suitability</i>	<i>LLOQ<sup>a</sup></i>	<i>Result Obtained<sup>b</sup></i>
Host Cell Protein (HCP) ELISA <sup>c</sup>	Met requirements	5 ng/mL	1.8 ng/mg
Protein A ELISA	Met requirements	0.1 ng/mL	< 0.01 ng/mg (No signal observed)
Insulin ELISA	Met requirements	4 $\mu$ IU/mL	< 0.4 $\mu$ IU/mg (No signal observed)
Host Cell DNA qPCR <sup>d</sup>	Met requirements	0.3 pg/mL	< $2.00 \times 10^{-5}$ ng/mg (No signal observed)

<sup>a</sup> LLOQ is the lower limit of quantitation of each impurity assay. <sup>b</sup> Results obtained in all the assays generally calculated using the following formula:

Method LLOQ  $\times$  Dilution Factor  $\div$  Product (monoclonal antibody [mAb]) Concentration.

<sup>c</sup> ELISA = enzyme-linked immunosorbent assay. <sup>d</sup> qPCR = quantitative polymerase chain reaction.

### *Insulin Assay*

Standards (ranging from 250  $\mu$ IU/mL to 1.95  $\mu$ IU/mL to generate a standard curve), the NISTmAb sample, and controls were added to an ELISA plate coated with commercially available anti-human insulin antibody (capture antibody). Following incubation, biotinylated anti-human insulin antibody (detector antibody) was added and incubated. Next, HRP-labeled streptavidin conjugate was added and incubated. The colorimetric substrate TMB (3,3',5,5'-tetramethylbenzidine) was then added and incubated before stopping the reaction by adding sulfuric acid. The absorbance was measured at 450 nm using a microplate reader. The insulin level in the NIST sample was regressed from the standard curve generated by the software and corrected for the dilution factor. The sample insulin concentration was reported as the mean concentration

of the dilutions yielding spike recovery values of 75 to 125%. The insulin level in the sample, after correction based on a mAb product concentration of 100 mg/mL, was calculated to be  $< 0.4 \mu\text{IU} / \text{mg}$  (Table 1). System suitability requirements of the SOP used for this assay were successfully met. In general, system suitability requirements pertain to obtaining an R value of  $\geq 0.990$  with the insulin standard curve, negative control showing a  $< \text{LLOQ}$  value, and the positive control to fall within 80 to 120  $\mu\text{IU}/\text{mL}$  in the insulin assay.

### *Residual Host Cell DNA Assay*

The assay was performed in a 96-well format using the Applied Biosystems 7500 Real-Time PCR System and Sequence Detection System (SDS) software using Power SYBR Green chemistry, with forward and reverse primers specific for short interspersed repetitive elements (SINE) found within the mouse genome. The NISTmAb sample, NS0 genomic DNA calibrators, and controls were added to a reagent master mix containing primers and SYBR Green. Quantitation of DNA in the NISTmAb sample was regressed from an assay-specific standard curve generated by the software. System suitability requirements of the SOP used for this assay were successfully met. In general, system suitability requirements pertain to obtaining an R value of  $\geq 0.990$  with the NS0 DNA standard curve, and no template control showing a  $C_t > \text{Mean } C_t \text{ of Standard Calibrator } 0.001 \text{ pg}/10 \mu\text{L}$  or undetermined quantity in the host cell DNA assay.

After running all of the process-related impurity assays described above, the results we obtained on the NIST sample mAb at MedImmune are shown below in Table 1.

## **Conclusions**

The current chapter has described the various process-related impurities one would encounter in a biopharmaceutical product (mAb). It is pertinent to mention here that the HCP values are critically evaluated from the point of view of the type of disease these products are clinically used for (acute or chronic) as well as the route of administration and the total dose used in patients.

As described in the section on regulatory guidance, the numbers shown for process-related impurities in Table 1 for the NIST sample are within the levels generally seen in FDA-reviewed products for HCP (3, 7) (ELISA-based HCP levels of 1–100 ng/mg) and the host cell DNA levels that need to be met per WHO guidelines (14) (upper limit of 10 ng/patient dose for monoclonal antibodies). There are no clear guidelines on the limits (upper levels) allowed for other process-related impurities such as Protein A and insulin in biopharmaceutical products, although the aim is to have them as low as possible. The NISTmAb also showed very low levels of these potential process impurities. In summary, the results obtained on the NISTmAb sample for all four process-related impurities (HCP, Protein A, insulin, and host cell DNA), as shown in Table 1, are (a) extremely low for HCP and (b) below the LLOQ for Protein A, insulin, and host cell DNA.

With regard to these process impurities, the NISTmAb material has at or near comparable levels expected for a regulatory submission. Therefore, although not intended for clinical use, the material can serve as a representative material typical of a purified drug substance.

## References

1. Gronemeyer, P.; Ditz, R.; Strube, J. Trends in upstream and downstream process development for antibody manufacturing. *Bioengineering* **2014**, *1*, 188–212.
2. Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A. Host cell protein analysis in therapeutic protein bioprocessing – Methods and applications. *Biotechnol. J.* **2013**, *8*, 655–670.
3. Champion, K.; Madden, H.; Dougherty, J.; Shacter, E. Defining your product profile and maintaining control over it, Part 2. *Bioprocess Int.* **2005**, *3* (8), 52–57.
4. Champion, K.; Nishihara, J. C.; Aldor, I. S.; Moreno, G. T.; Andersen, D.; Stults, K. L.; Vanderlaan, M. Comparison of *Escherichia coli* proteomes of recombinant human growth hormone producing and nonproducing fermentations. *Proteomics* **2003**, *3*, 1365–1373.
5. Krawitz, D. C.; Forrest, W.; Moreno, G. T.; Kittleson, J.; Champion, K. Proteomic studies support the use of multi-product immunoassays to monitor host cell protein impurities. *Proteomics* **2006**, *6*, 94–110.
6. *Implementation of Mustang® Q Membrane Chromatography as a Polishing Step (Residual DNA Removal) in Monoclonal IgG1 Production from CHO Cell Culture*; Application Note USTR 2827; PALL: Port Washington, NY, 2012.
7. Eaton, L. C. Host cell contaminant protein assay development for recombinant biopharmaceuticals. *J. Chromatogr. A* **1995**, *705*, 105–114.
8. Wang, X.; Hunter, A.; Mozier, N. Host cell proteins in biologics development. *Biotechnol. Bioeng.* **2009**, *103*, 446–458.
9. General Biological Products Standards. *Code of Federal Regulations*, Part 610, Title 21, 2014.
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. *CPMP Position Statement on DNA and Host Cell Proteins (HCP) Impurities, Routine Testing versus Validation Studies*; European Medicines Agency: London, 1997.
12. Shukla, A. A.; Jiang, C.; Ma, J.; Rubacha, M.; Flansburg, L.; Lee, S. S. Demonstration of robust host cell protein clearance in biopharmaceutical downstream processes. *Biotechnol. Prog.* **2008**, *24*, 615–622.
13. *Points To Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*; U.S. Food and Drug Administration: Washington, DC, 1997. <http://www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/>



OtherRecommendationsforManufacturers/UCM153182.pdf (accessed June 2015).

14. WHO Study Group: *Acceptability of Cell Substrates for Production of Biologicals*; Technical Report Series 747; World Health Organization: Geneva, Switzerland, 1986.
15. DuMontelle, J.; Zolodz, M.; Deora, A.; Mozier, M.; Golding, B. Use of toll-like receptor assays to detect and identify microbial contaminants in biological products. *J. Clin. Microbiol.* **2009**, *47*, 3427–3434.
16. Xue, L.; Johnson, R.; Gorovits, B. Prevalence and isotypic complexity of the anti-chinese hamster ovary host cell protein antibodies in normal human serum. *AAPS J.* **2010**, *12*, 98–106.
17. Robert, F.; Bierau, H.; Agugiaro, D.; Soranzo, T.; Broly, H.; Mitchell-Logean, C. Degradation of an Fc-fusion recombinant protein by host cell proteases. Identification of a CHO cathepsin D protease. *Biotechnol. Bioeng.* **2009**, *104*, 1132–1141.
18. Schellekens, H. Biosimilar therapeutics—What do we need to consider? *Nephrol. Dial. Transplant. Plus* **2009**, *2* (Suppl 1), i27–i36.
19. Drugs@FDA. <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm> (accessed March 2010).
20. Sato, S.; Nishida, N.; Abe, T.; Suzuki, T.; Miyajima, H.; Mayahara, H. Four week intraperitoneal toxicity study of protein A in rats. *Pharmacometrics* **1989**, *37*, 395–405.
21. Yanagimoto, Y.; Yamamoto, H.; Nishida, M.; Matsuo, A.; Matsuoka, K.; Sato, H. Four week intravenous toxicity study of staphylococcal protein A in monkeys. *Pharmacometrics* **1989**, *37*, 517–527.
22. Ehrlich, M.; Gama Sosa, M. A.; Huang, L.-H.; Midgett, R. M.; Kuo, K. C.; McCune, R. A.; Gehrke, C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.* **1982**, *374*, 2709–2721.
23. Krieg, A. M.; Yi, A. K.; Matson, S.; Waldschmidt, T. J.; Bishop, G. A.; Teasdale, R.; Koretzky, G. A.; Klinman, D. M. CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature.* **1995**, *374*, 546–549.
24. *Meeting Report: WHO Study Group on Cell Substrates for Production of Biologicals*; World Health Organization: Geneva, Switzerland, 2007.
25. Yang, H.; Zhang, L.; Galinski, M. A probabilistic model for risk assessment of residual host cell DNA in biological products. *Vaccine* **2010**, *28*, 3308–3311.
26. Nogal, B.; Chhiba, B.; Emery, J. C. Select host cell proteins coelute with monoclonal antibodies in protein a chromatography. *Biotechnol. Prog.* **2012**, *24*, 454–458.
27. Flatman, S.; Alam, I.; Gerrard, J.; Mussa, N. Process analytics for purification of monoclonal antibodies. *J. Chromatogr. B* **2007**, *848*, 79–87.
28. Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Sapels, M.; Chen, W. Analysis of HCPs in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *mAbs* **2012**, *4*, 24–44.

29. Bomans, K.; Lang, A.; Roedl, V.; Adolf, L.; Kyriosoglou, K.; Diepold, K.; Eberl, G.; Mølhøj, M.; Strauss, U.; Schmalz, C.; Vogel, R.; Reusch, D.; Wegele, H.; Wiedmann, M.; Bulau, P. Identification and monitoring of host cell proteins by mass spectrometry combined with high performance immunochemistry testing. *PLoS One* **2013**, *8*, 110–121.
30. Savino, E.; Hu, B.; Sellers, J.; Sobjak, A. Development of an in-house, process specific ELISA for detecting HCP in a therapeutic antibody. *Bioprocess Int.* **2011**, *9*, 68–75.
31. Jin, M.; Szapiel, M.; Zhang, J.; Hickey, J.; Ghose, S. Profiling of host cell proteins by two-dimensional gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnol. Bioeng.* **2009**, *105*, 306–316.
32. Grzeskowiak, J. K.; Tscheliessnig, A.; Toh, P. C.; Chusainow, J. 2-D DIGE to expedite downstream process development for human monoclonal antibody purification. *Protein Expression Purif.* **2009**, *66*, 58–65.

# Appendix

**Table 1. Acronyms**

1-D	1-dimensional
2-AA	2-aminobenzoic acid
2-AB	2-aminobenzamide
2-D	2-dimensional
3-OH-Kyn	3-hydroxykynurenine
5-OH-Trp	5-hydroxyl-tryptophan
AAA	amino acid analysis
AAPH	2,2'-azobis(2-amidinopropane) dihydrochloride
AARS	aminoacyl-tRNA synthetases
ABS	<i>Athrobacter ureafaciens</i> sialidase
ABTS	2,2'-azinobis[3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt
ADA	anti-drug antibodies
ADC	antibody–drug conjugate
ADCC	antibody-dependent cellular cytotoxicity
ADCC	antibody-dependent, cell-mediated cytotoxicity
AFFFF	asymmetric flow field-flow fractionation
APTS	9-aminopyrene-1,4,6-trisulfonic acid
Asu	succinimide
ATR	attenuated total reflection
ATR-FTIR	attenuated total reflection-Fourier transform infrared
AUC	area under the curve
AUC	analytical ultracentrifugation
BCA	bicinchoninic acid assay
BKF	bovine kidney fucosidase
BLA	Biologics License Application
bp	base pairs

*Continued on next page.*

**Table 1. (Continued). Acronyms**

BSA	bovine serum albumin
BTG	bovine testes $\beta$ -galactosidase
C of A	Certificate of Analysis
CaCl <sub>2</sub>	calcium chloride
CBG	green coffee bean $\alpha$ -galactosidase
CCD	charge-coupled device
CD	circular dichroism
CDC	complement-dependent cytotoxicity
cDNA	complementary DNA
CDR	complementarity-determining regions
CE	capillary electrophoresis
CEX	cation exchange
CEX	cation exchange chromatography
CFG	Consortium for Functional Glycomics
CG-MALS	composition gradient-multi-angle light scattering
C <sub>H1</sub>	first constant domain of the heavy chain
C <sub>H2</sub>	second constant domains of the heavy chain
C <sub>H3</sub>	third constant domains of the heavy chain
CHO	Chinese hamster ovary
CIC	cross-interaction chromatography
CID	collision-induced dissociation
cIEF	capillary isoelectric focusing
CpB	carboxypeptidase B
CpG	cytosine-guanosine dinucleotide
CQA	critical quality attribute
cSDS	capillary sodium dodecylsulfate electrophoresis
cyno	cynomolgus
CZE	capillary zone electrophoresis
DAD	diode array detector
DDA	data-dependent acquisition
DHFR	dihydrofolate reductase
DIA	Data Independent Analysis
DIGE	difference gel electrophoresis

*Continued on next page.*

**Table 1. (Continued). Acronyms**

diH <sub>2</sub> O	deionized water
DiOia	dioxindolylalanine diastereomers
DLS	dynamic light scattering
DMSO	dimethylsulfoxide
DOE	design of experiment
DPBS	14190-DPBS obtained from Life Technologies
DPBS	Dulbecco's phosphate-buffered saline
DSC	differential scanning calorimetry
DSF	differential scanning fluorimetry
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EI	electron impact
EIC	extracted ion chromatograms
EI-GC-MS	electron impact-gas chromatography-mass spectrometry
ELISA	enzyme-linked immunosorbent assay
ELSD	evaporating light scattering detector
EMA	European Medicines Agency
ESI	electrospray ionization
ESZ	electrical sensing zone
ETD	electron transfer dissociation
ETS	error tolerant search
FA	formic acid
FAB	fast atom bombardment
FcRn	neonatal Fc receptor
FDA	Food and Drug Administration
FDA	U.S. Food and Drug Administration
FFF	field-flow fractionation
FI	flow imaging
FLD	fluorescence detection
FP	fusion protein
FQ	3-(2-furoyl)quinoline-2-carboxaldehyde
FR	fraction rate

*Continued on next page.*

**Table 1. (Continued). Acronyms**

FRET	fluorescence resonance energy transfer
FTICR	Fourier transform ion cyclotron resonance
FTIR	Fourier-transform infrared spectroscopy
FTIR	Fourier transform infrared
FTN	flow through needle
GlcNAc	N-acetyl-glucosamine
GMP	good manufacturing practice
GS	glutamine synthetase
GU	glucose unit
GUH	<i>Streptococcus pneumoniae</i> hexosaminidase
H	heavy
H	heavy chain
HC	heavy chain
HCD	higher energy collision dissociation
HCD	higher energy collisional dissociation
HCID	higher energy collision-induced dissociation
HCl	hydrochloric acid
HCP	host cell protein
HDX	hydrogen-deuterium exchange
HESI	heated electrospray ionization source
hG-CSF	human granulocyte colony stimulating factor
HIC	hydrophobic interaction chromatography
HILIC	hydrophilic interaction liquid chromatography
HMWS	higher molecular weight species
HOS	higher order structure
HPLC	high-performance liquid chromatography
HPLC	high-pressure liquid chromatography
HRP	horseradish peroxidase
HRP II	histidine-rich protein II
HsTIM	homodimeric triosephosphate isomerase
HTP	High Throughput
IAA	iodoacetic acid
IAM	iodoacetamide

*Continued on next page.*

**Table 1. (Continued). Acronyms**

ICH	International Conference on Harmonisation
ICIEF	imaged capillary isoelectric focusing
ICP-MS	inductively coupled plasma-mass spectrometry
ID	interior diameter
IEC	ion exchange chromatography
IEF	isoelectric focusing
IEX	ion exchange chromatography
IM-MS	ion mobility mass spectrometry
IRMPD	infrared multiphoton dissociation
iso-Asp	iso-aspartic acid
IUPAC	International Union of Pure and Applied Chemistry
KCN	potassium cyanide
Kyn	kynurenine
L	light
L	light chain
LC	liquid chromatography
LC	light chain
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LC-UV-MS	liquid chromatography-UV-mass spectrometry
LED	light-emitting diode
LIF	laser-induced fluorescence
LLOQ	lower limit of quantitation
LO	light obscuration
LS	light scattering
LTQ	linear trap quadrupole
Lys-C	lysyl endopeptidase
mAb	monoclonal antibody
MALDI	matrix-assisted laser desorption ionization
MALDI-TOF	matrix assisted laser desorption/ionization-time of flight
MALS	multi-angle light scattering
MC	methylcellulose
MCAE	membrane-confined analytical electrophoresis

*Continued on next page.*

**Table 1. (Continued). Acronyms**

MCE	microchip electrophoresis
MCE	membrane-confined electrophoresis
MCE-SSE	membrane-confined electrophoresis-steady-state electrophoresis
MCT	mercury cadmium telluride
MPA	mobile phase A
MPB	mobile phase B
MRE	mean residue ellipticity
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS <sup>2</sup>	tandem mass spectrometry
MS <sup>n</sup>	sequential mass spectrometry
MSX	methionine sulfoximine
MTX	methotrexate
MW	molecular weight
MWCO	molecular weight cutoff
MWD	multiple wavelength detector
NA	not available
NaCNBH <sub>3</sub>	sodium cyanoborohydride
NaOH	sodium hydroxide
NCE	normalized collision energy
ND	not detected
NEM	N-ethylmaleimide
Neu5GC	N-glycolylneuraminic acid
NFK	<i>N</i> -formylkynurenine
NGH	non-glycosylated heavy chain
NGS	next-generation sequencing
NIBRT	National Institute for Bioprocessing Research and Training
NIST	National Institute of Standards and Technology
Nle	norleucine
NMR	nuclear magnetic resonance spectroscopy
NR	nonreducing
NS0	Murine Myeloma

*Continued on next page.*



**Table 1. (Continued). Acronyms**

NSI-MS	MS nano-infusion mode
NTA	nanoparticle tracking analysis
Nva	norvaline
Oia	oxindolylalanine diastereomomers
OP	operational parameter
p/v	peak-to-valley
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCQA	potential critical quality attribute
PCR	polymerase chain reaction
PCS	photon correlation spectroscopy
Pd	polydispersity
PDA	photodiode array
PENNYK peptide	GFYPSDIAVEWESNGQPENNYK
pI	isoelectric point
PIMT	protein L-isoaspartyl methyltransferase
PNGase F	peptide N-glycosidase F
poly IgG	polyclonal IgG
PPG 2700	poly(propylene glycol) 2700
pQ	pyroglutamic acid
pre-mRNA	primary transcript messenger RNA
PRI	process-related impurity
PSL	polystyrene latex
PTM	post-translational modification
pyro-Glu	N-terminal pyroglutamate
QbD	quality by design
QC	quality control
QELS	quasi-elastic light scattering
qPCR	quantitative polymerase chain reaction
QTOF	quadrupole time-of-flight
qTOF	quantitative time-of-flight
R	correlation coefficient

*Continued on next page.*

**Table 1. (Continued). Acronyms**

R	reducing
RA	relative abundance
RM	reference material
RMM	resonance mass measurement
RMSD	root-mean-square deviation
RP	reversed phase
RP-HPLC	reversed-phase-high-performance liquid chromatography
RPM	revolutions per minute
RSD	relative standard deviation
S/N	signal to noise
scFc	single chain Fc
SD	standard deviation
SDS	Sequence Detection System
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE-AUC	sedimentation equilibrium-analytical ultracentrifugation
SEC	size exclusion chromatography
SEC-MALS	size exclusion chromatography-multi-angle light scattering
SIC	self-interaction chromatography
SINE	short interspersed repetitive elements
SLS	static light scattering
SNP	single-nucleotide polymorphism
SOP	standard operating procedure
SRDP	size-related degradation product
Sv	svedberg unit
SV	sequence variant
SVA	sequence variant analysis
SV-AUC	sedimentation velocity-analytical ultracentrifugation
TAMRA.SE	5-carboxytetra-methylrhodamine succinimidyl ester
t-BHP	tert-butyl hydroperoxide
TCEP	tris(2-carboxyethyl)phosphine
TETA	triethylenetetramine
TFA	trifluoroacetic acid

*Continued on next page.*

**Table 1. (Continued). Acronyms**

TMB	3,3',5,5'-tetramethylbenzidine
TOF	time-of-flight
TPP	target product profile
TRamp	temperature ramp
Tris	tris(hydroxymethyl)aminomethane
Tris	tris(hydroxymethyl)aminomethane/tris(hydroxymethyl)aminomethane HCl
tRNA	transfer RNA
TUV	tunable ultraviolet
UHPLC	ultrahigh-pressure liquid chromatography
UHPLC	ultra-high-performance liquid chromatography
UHR-ESI-QTOF MS	ultrahigh-resolution-electrospray ionization-quadrupole time-of-flight mass spectrometry
UPLC	ultra performance liquid chromatography
USP	United States Pharmacopeia
V <sub>H</sub>	variable heavy
VWD	variable wavelength detector
WAX	weak anion exchange
WHO	World Health Organization
XIC	extracted ion chromatogram

# Subject Index

## B

Biotherapeutic proteins, 63  
cell culture, 110  
molecular biology cell culture, 111  
introduction, 65  
amino acid analysis (AAA), 87  
analytical chemistry, focus, 90  
biological characterization assays, 83  
biosynthesis, fidelity, 67  
biosynthesis of proteins, processes, 68*f*  
biosynthetic errors, 71  
biotherapeutic proteins, 89  
capillary isoelectric focusing (cIEF), 86  
cell clones, 80  
CHO-produced anti-Her2 antibody, 79  
codon-anticodon base pairing, 70  
DNA mutations, 79  
DNA recombination, 89  
high-pressure liquid chromatography (RP-HPLC), 85  
human homodimeric triosephosphate isomerase (HsTIM), 71  
ion exchange chromatography (IEC), 86  
Leu biosynthesis, 73  
liquid chromatography-mass spectrometry, role of peptide mapping, 88  
living cells, protein synthesis, 66  
mammalian cell lines, errors in protein expression, 74  
mRNA translation, 69  
multiple orthogonal methods, advantages, 84  
peptide mapping methods, evolution, 87  
quality mammalian recombinant proteins, 73  
quality of recombinant proteins, expression errors, 81  
recombinant IgG1 antibody, sequence variants, 91  
recombinant proteins in mammalian cell lines, sequence variants, 75*t*  
single amino acid change, effects, 82  
single-nucleotide polymorphism (SNP), 68  
translational errors, 89  
tRNA mischarging, 72

methods overview, 92  
hydrophobic residues, frequency, 93  
LC-MS analysis and database searching, 94*t*  
modern mass spectrometers, 95  
post-Mascot search criteria, 96  
sample preparation and digestion details, 92*t*  
results, 97  
accuracy, 109  
amino acid substitutions, 98*t*  
annotated tandem mass spectrometry (MS/MS) spectra, lab 1, 103*f*  
annotated tandem mass spectrometry (MS/MS) spectra, lab 2, 102*f*  
estimation, 107*f*  
extracted ion chromatograms, lab 1, 100*f*  
extracted ion chromatograms, lab 2, 101*f*  
mass spectrometry (MS), log<sub>10</sub>-log<sub>10</sub> dependence, 106*f*  
merit, analytical figures, 105  
method detection limit, 106  
precision, 108  
regression lines, characteristics, 110*t*  
sequence variant analysis, Venn diagram, 99*f*  
sequence variants, correlation, 109*f*  
spiking experiments, low-level variants, 104

## H

Host cell proteins and other process-related impurities, 387  
assay protocols and summary, NIST sample, 398  
process-related impurity assays, NIST sample results, 400*t*  
general introduction, 388  
good manufacturing practice (GMP) testing, 397  
HCP assay critical reagents, management, 397  
HCP clearance and consistency, 391  
HCP critical reagents, characterization and development, 394  
two-dimensional gel western blot analysis, 396*f*

HCPs, immunogenicity, 390  
quantitation methods, 392  
  enzyme-linked immunosorbent assay  
  (ELISA), 393*f*  
regulatory guidance, 389

## M

Monoclonal antibodies, higher order  
structure and interactions, 285

analytical ultracentrifugation  
  Beckman XL-I analytical  
  ultracentrifuge, 294  
  curve fitting with SEDFIT, parameters  
  used, 295*t*

equilibrium concentration gradient,  
289

NISTmAb, sedimentation  
equilibrium, 292*f*

sedimentation equilibrium  
experiments, 291

sedimentation equilibrium-analytical  
ultracentrifugation (SE-AUC), 289

sedimentation velocity-analytical  
ultracentrifugation, 296

sedimentation velocity-analytical  
ultracentrifugation (SV-AUC), 293

SV-AUC, fitted distribution, 298*f*

three orthogonal techniques,  
comparison of molecular weights,  
293*t*

UV light-degraded NISTmAb  
mixtures, 299*t*

global methods, 289

light scattering

  Brownian motion, 304

  collective diffusivity, 306*f*

  diffusivity, 305

  DLS best fit parameters, 306*t*

  dynamic light scattering, 303

  experimental static light scattering,  
  302*t*

  fitted parameters, 303*t*

  NIST IgG1 mAb, static light scattering  
  data, 303*f*

  static light scattering, 300

membrane-confined analytical

  electrophoresis, 307

  free-boundary electrophoretic  
  mobility, 308

  membrane-confined electrophoresis-  
  steady-state electrophoresis  
  (MCE-SSE), 311

  NISTmAb, charge (zDHH)  
  determination, 310*f*  
  steady-state electrophoresis, 312*f*  
secondary and tertiary structure,  
signatures  
  amide I frequency peaks, assignment,  
  320*t*  
  circular dichroism, 323  
  differential scanning calorimetry, 313  
  differential scanning fluorimetry, 316  
  DSC results, 315*t*  
  fourier transform infrared  
  spectroscopy, 318  
  infrared radiation, absorption, 319  
  molecular ellipticity, 324*f*  
  NISTmAb, differential scanning  
  calorimetry (DSC) profiles, 316*f*  
  NISTmAb, DSF profiles, 318*f*  
  NISTmAb analysis, experimental  
  DSC parameters, 314*t*  
  NISTmAb hydrophobic exposure  
  temperatures, 317*t*  
  NISTmAb samples, absorbance  
  spectra with no salt, 321*f*  
  NISTmAb samples, absorbance  
  spectra with salt, 322*f*

Monoclonal antibodies, post-translational  
modifications, 119

  common PTMs, 124

  carboxyl tail, loss of lysine, 128

  collision-induced dissociation (CID)  
  spectra, 137*f*

  cysteines, variants, 137

  deamidation, 130

  deamidation and iso-aspartate  
  formation, 129

  Edman sequencing, 130

  enzymatic digestion, 131

  GFYPSDIAVEWESNGQPENNYK  
  peptide, deamidation, 132*f*

  glutamine and glutamate, cyclization,  
  128

  glycation, 135

  glycosylation, 127

  hydroxylation, 139

  intact mAb, fragmentation, 138

  isobaric species, 134

  oxidation, 132

  oxidized peptides, 133

  post-translational modifications  
  (PTMs) in mAbs, 125*t*

  PTM characterization

    artifactual oxidation, 157

    chemical oxidation, 169

    closing remarks, 172

C-Terminal K and N-Terminal Pyro-Glu, 152  
 C-terminal Lys changes, presence, 149*f*  
 deamidation, 153  
 degradation conditions, 165*t*  
 des-lysine peptide, relative abundance, 149*t*  
 forced degradation analysis, importance, 163  
 glycation, 159  
 glycosylation, 150  
 interlaboratory study, conclusions, 171  
 liquid chromatography-mass spectrometry (LC-MS) analysis, 142*t*  
 met oxidized peptide H, quantitation, 146*f*  
 NIST reference mAb, 150  
 NISTmAb, identification of C-terminal lysine loss and N-terminal Pyro-Glu, 153*t*  
 NISTmAb control sample, tryptic digestion, 155*t*  
 NISTmAb glycation, identification and quantification, 160*t*  
 NISTmAb methionine oxidation, 157*t*  
 NISTmAb PTMs, overview, 161  
 NISTmAb stress results, 167  
 NISTmAb tryptophan oxidation, 159*t*  
 oxidation, 156  
 oxidative stress, 166  
 peptides, chemical modification, 168*f*  
 pH stress, 166  
 post-translational modifications (PTMs), list, 162*f*  
 putative glycans, 151*t*  
 quantification strategies, 144  
 reconstructed ion chromatograms, 154  
 relative abundance, 157*f*  
 relative post-translational modification (PTM) levels, comparison, 170*f*  
 sample preparation, 140*t*  
 stressed NISTmAb samples, PTM level changes, 164  
 study, purpose, 139  
 thermal stress, 165  
 XIC, software coding, 148  
 PTM characterization, methods, 121  
 bottom-up approaches, 123  
 intact mass analysis, 122

## N

NISTmAb N-glycan structure, elucidation and quantitation, 185  
 glycan characterization workflow, 197  
 charged 2-aminobenzamide (2-AB)-labeled glycans, structural assignments, 217*f*  
 co-eluting species, 211*t*  
 collision-induced dissociation (CID), 207  
 deglycosylated and aglycosylated TH(296-304) peptide, 199*t*  
 deglycosylated and aglycosylated TH(296-304) peptide, extracted ion chromatogram, 200*f*  
 FA2G1 structure, lactosamine extension, 214*f*  
 glycan species, permethylated fragments, 212*t*  
 N-Glycolylneuraminic acid (Neu5Gc), identification and quantitation, 216*t*  
 NIST 2-aminobenzoic acid (2-AA) glycan, overlay, 218*f*  
 NIST IgG1, representative zoomed glycan map, 203*f*  
 NIST IgG1 glycan grouping, 219*t*  
 NIST monoclonal antibody, overlay, 209*f*  
 N-linked glycan compositions, identification, 204*t*  
 non-human immunogenic glycan species, evaluation, 214  
 NSI-MSn precursor-product pathway, 212*s*  
 product characterization studies, use of glycan characterization workflow, 202*f*  
 putative structure, tandem mass spectrometry (MS/MS), 208*f*  
 second dimension ultra-performance liquid chromatography reversed phase (UPLC RP) separation, 210*f*  
 sequential mass spectrometry (MSn) spectrum, 213*f*  
 site determination and occupancy, 198  
 species identification, 201  
 tandem mass spectrometry (MS/MS) higher energy collisional dissociation (HCD) spectrum, 199*f*  
 weak anion exchange (WAX) fractionation, 216*f*  
 zoomed NIST monoclonal antibody (mAb) glycan profiles, overlay, 215*f*

- High-Throughput N-linked glycan profiling, 189
- 2-aminobenzamide (2-AB)-labeled NIST monoclonal antibody (mAb) glycans, sequencing, 194*f*
- glycoanalysis platform, 190
- GlycoBase web interface, 191*f*
- glycomics platform, National Institute for Bioprocessing Research and Training (NIBRT), 189*f*
- hydrophilic interaction liquid chromatography (HILIC) fluorescence chromatogram, 191*f*
- hydrophilic interaction liquid chromatography (HILIC) glycan, 192*t*
- NIST monoclonal antibody (mAb), performance characteristics, 196*t*
- process optimization support, qualification of glycan assays, 195
- routine sample testing, 197*t*
- interlaboratory comparison, 224
- glycoprofiling, compositional identifications, 225*f*
- N-linked glycans, chemical structure, 188*f*
- ultra-performance liquid chromatography (UPLC) systems, 187
- LC-FLD glycosylation mapping, 219
- compositional glycoprofiling, 223
- fluorescence and mass spectrometry (MS) signals, overlay, 221*f*
- fluorescence signal, 222*f*
- mass spectrometric detection, 220
- NISTmAb primary structure, determination, 1
- MS/MS fragmentation, 4
- proteomics analysis, 3
- peptide mapping, technique, 33
- automated data analysis, 39
- chymotrypsin, 51
- current peptide mapping, 35
- data acquisition modes, 37
- discussion, 56
- disulfide-linked peptides, 55*t*
- enzyme digestion, 41
- experimental materials, 39
- history, 34
- identity testing, 54
- lab 1, 39
- lab 2, 41
- liquid chromatography and mass spectrometry instruments, types, 36
- MS analysis, 40
- NISTmAb, Lys-C peptide mapping, 43*f*
- NISTmAb peptide coverage, 52*f*
- NISTmAb peptide mapping, summary, 49*t*
- non-reduced peptide map, 41
- non-reduced peptide mapping, 53
- orthogonal assays, 48
- peptide identification, 41
- peptide identification, using Byonic software package, 40
- peptide mapping, 41
- peptide mapping, enzyme digestion, 39
- peptide-level coverage, 38
- results, 42
- 1960s, 34
- 1990s, 35
- 1970s and 1980s, 34
- theoretical and observed peptides, 44*t*
- primary structure, 5
- chromatographic separation, 8
- C-terminal lysine residues, 15
- De-N-glycosylated NISTmAb, average masses, 16*t*
- de-N-glycosylated NISTmAb, intact mass measurements, 16*f*
- direct infusion, 18*f*
- experimental materials, 9
- glycated proteoforms, 15
- intact mAb, molecular mass analysis, 8
- intact NISTmAb, theoretical mass calculations, 12*t*
- intact protein analysis, 9
- liquid chromatography-mass spectrometry (LC-MS), intact mass measurements of the NISTmAb, 13*f*
- NISTmAb, average masses, 14*t*
- NISTmAb, schematic representation, 6*f*
- NISTmAb amino acid sequence, 7*f*
- orbitrap fusion tribrid mass spectrometer, 10
- results, 11
- top-down analysis, sequence coverage of the NISTmAb, 19*f*
- top-down fragmentation, 10
- top-down sequencing results, 17
- reference material complexity, 56
- subunit and IdeS fragment analysis, 21
- accurate mass assignments, 26*t*
- chromatographic resolution, 33
- De-N-glycosylated IdeS fragments, accurate mass assignments, 30*t*

direct infusion, middle-down mass measurements, 31*t*  
experimental materials, 23  
fragment maps, 32*f*  
IdeS digestion, 22*f*  
IdeS fragment mass analysis, 24  
light (L) chain subunit mass spectra, experimental and theoretical analysis, 29*f*  
liquid chromatography-mass spectrometry (LC-MS) fragment analysis, 25*f*  
middle-down sequencing, 30  
NISTmAb fragments, deconvoluted mass spectra, 28*f*  
raw mass spectrum, 31*f*  
results, 24

## O

Orthogonal and separation methods, 237  
RP-HPLC analysis, 276  
methods  
  capillary electrophoresis with sodium dodecyl sulfate, 241  
  capillary isoelectric focusing, 246  
  capillary zone electrophoresis, 245  
  cation exchange HPLC, 244  
  hydrophobic interaction chromatography, 244  
  imaged capillary isoelectric focusing, 246  
  microchip electrophoresis with sodium dodecyl sulfate, 242  
  reversed-phase HPLC, 243  
  size exclusion chromatography with multi-angle light scattering, 240  
  size exclusion chromatography-HPLC, 239  
  size exclusion chromatography-ultrahigh-pressure liquid chromatography, 240  
  sodium dodecyl sulfate-polyacrylamide gel electrophoresis, 241  
results and discussion  
  apparent isoelectric point (pI), experimental determination, 275*f*  
  capillary electrophoresis with sodium dodecyl sulfate, 253  
  capillary isoelectric focusing, 271  
  carboxypeptidase B-Treated NIST monoclonal antibody standard, peak area, 274*t*

charge variant analysis, 267  
hydrophobic interaction chromatography, 265  
hydrophobic variant analyses, summary, 267*t*  
hydrophobic variant analysis, 261  
imaged capillary isoelectric focusing, 273  
LIF detection, advantages, 255  
light chain and glycosylated light chain, 257*f*  
microchip electrophoresis with sodium dodecyl sulfate, 258  
multi-angle light scattering (MALS) analysis, 251*f*  
multi-angle light scattering, size exclusion chromatography, 250  
native (black) and carboxypeptidase B (CpB)-treated (grey) sample, overlay, 270*f*  
native NISTmAb, effect of varied focusing times, 272*f*  
NISTmAb, expanded-view size heterogeneity profiles, 254*f*  
NISTmAb, reversed-phase chromatograms, 263*f*  
NISTmAb, size exclusion chromatography (SEC)-HPLC analysis, 249*f*  
NISTmAb, size heterogeneity, 259*f*  
NISTmAb, zoomed view, 264*f*  
NISTmAb before carboxypeptidase B treatment, cation exchange (CEX)-HPLC chromatograms, 268*f*  
NISTmAb following incubation, cation exchange (CEX)-HPLC chromatograms, 269*f*  
NISTmAb size heterogeneity, 256*f*  
NISTmAb with UV detection, 250*f*  
NIST monoclonal antibody size variant analyses, summary, 260*t*  
NIST monoclonal antibody standard, peak area, 274*t*  
NISTmAb, hydrophobic interaction chromatography (HIC)-HPLC analysis, 266*f*  
reduced and non-reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis, 252*f*  
reduced NIST monoclonal antibody standard, 261*t*  
reversed-phase HPLC, 262  
size exclusion chromatography, 248  
size variant analysis, 247



sodium dodecyl sulfate-  
polyacrylamide gel electrophoresis,  
251

## P

Proposed NIST monoclonal antibody,  
developability assessment, 329  
developability analysis, 353  
materials and methods

capillary isoelectric focusing, 333  
cross-interaction chromatography, 335  
differential scanning calorimetry, 334  
dynamic light scattering, 334  
microfluidic electrophoresis, 333  
phase separation in serum, 335  
post-translational modification risk  
assessment, 336  
self-interaction chromatography, 335  
size-exclusion chromatogra-  
phy–multi-angle light scattering,  
334  
solubility and stability, 332  
UV and DLS analysis, 337

results and discussion

capillary isoelectric focusing, 339  
cIEF, profiles, 349*f*  
cIEF profiles, summary, 349*t*  
concentratability, 337  
cross-interaction chromatography, 344  
cross-interaction chromatography  
(CIC) profile, 345*f*  
cross-interaction chromatography  
data, 345*t*  
differential scanning calorimetry, 342  
DSC Data, summary, 343*t*  
dynamic light scattering, 341  
forced degradation, results, 346  
microfluidic electrophoresis, 338  
microfluidic electrophoresis analysis,  
347  
microfluidic electrophoresis profiles,  
339*f*  
microfluidic electrophoresis  
quantification after photo-oxidation  
stress exposure, summary, 348*t*  
microfluidic electrophoresis  
quantification at high concentration,  
summary, 339*t*  
NISTmAb, A350 nm (turbidity)  
values, 352*f*  
NISTmAb, capillary isoelectric  
focusing (cIEF) profile, 340*f*

NISTmAb, microfluidic  
electrophoresis, 348*f*  
NISTmAb, summary of protein  
concentration, 338*t*  
NISTmAb after incubation, DLS data,  
342*t*  
NISTmAb after incubation,  
SEC-MALS data, 341*t*  
NISTmAb stability samples, 350*t*  
normalized intensity values, DLS,  
352*f*  
phase separation testing, 344*f*  
preformulation (Buffer pH), 351  
PTM NISTmAb samples, DLS data,  
351*t*  
self-interaction chromatography, 346  
self-interaction chromatography (SIC)  
profile, 346*f*  
serum, phase separation, 343  
SIC data, summary, 346*t*  
size-exclusion chromatogra-  
phy–multi-angle light scattering,  
340  
size-exclusion chromatogra-  
phy–multi-angle light scattering  
(SEC-MALS) profiles, 341*f*  
stressed samples, NISTmAb, 350*f*  
thermal unfolding of NISTmAb,  
thermograms, 343*f*

Protein particles, 357  
advanced techniques, 379  
analytical techniques  
diameter error and count error,  
relationship, 368*f*  
filtration, 370  
light obscuration and flow imaging  
particle counters, schematics, 365*f*  
overview, 363  
particle detection methods,  
effective-diameter range, 364*f*  
particle detection methods, summary,  
364*t*  
particle orientation, 368  
quantification, 367  
reference materials, 366  
sampling, 369

flow imaging, metrology  
overview, 376  
quality assurance, 378  
quantitation, repeatability and  
accuracy, 377  
sample characteristics, 376

introduction  
optical images, 361*f*  
protein aggregates, classification, 361

protein aggregates formation,  
  schematic, 359*f*  
protein particles formation, factors  
  that influence, 360*f*  
protein subvisible particles, sources,  
  358  
  subvisible particles, 362  
light obscuration, metrology  
  overview, 373  
  quantitation, limit and linearity, 374  
  sample handling, 373  
  troubleshooting, 375  
sample characteristics and handling  
  degassing, 371  
  handling, 370  
  interferences, 372  
  sample stability, 371