

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/45602555>

# Kim YJ, Doyle ML. Structural mass spectrometry in protein therapeutics discovery. Anal Chem 82: 7083-7089

ARTICLE *in* ANALYTICAL CHEMISTRY · SEPTEMBER 2010

Impact Factor: 5.64 · DOI: 10.1021/ac101575d · Source: PubMed

---

CITATIONS

20

---

READS

12

## 2 AUTHORS:



**Yeoun Jin Kim**

LIH Luxembourg Institute of Health

**13** PUBLICATIONS **261** CITATIONS

SEE PROFILE



**Michael L Doyle**

Bristol-Myers Squibb

**107** PUBLICATIONS **4,199** CITATIONS

SEE PROFILE

# Structural Mass Spectrometry in Protein Therapeutics Discovery

Yeoun Jin Kim\* and Michael L. Doyle

Gene Expression and Protein Biochemistry, Applied Biotechnologies, Bristol-Myers Squibb,  
P.O. Box 4000, Princeton, New Jersey 08543

The protein therapeutics market is one of the highest growing segments of the pharmaceutical industry with an estimated global market value of \$77 billion by 2011 (*Global Protein Therapeutics Market* report by RNCOS: Delhi, India, 2009). This growth has been fueled by several advantages that protein drugs can offer such as higher specificity, reduced side effects, and faster development time compared to small molecule drugs. Major pharmaceutical companies are strategically shifting gears toward protein therapeutics and gradually increasing the biologics portion of their pipelines. Consequently, in the present pharmaceutical industry, there is a rapid growth in the number and types of protein structural mass spectrometry analyses, particularly during the discovery phase where an abundance of new drug candidates are being investigated. This perspective article discusses the role of protein structural mass spectrometry during the discovery of protein therapeutics with focus on recombinant protein production quality control and structural biology applications. The current challenges in technologies associated with this field and the analytical prospects for the future direction will be also discussed.

## PROTEIN PRODUCTION IN PROTEIN THERAPEUTICS DISCOVERY

There are two general classes of proteins that are produced to support the discovery of protein therapeutics: protein targets and the protein therapeutic candidates themselves (Figure 1). The first proteins produced are the set of biological target molecules in order that they may be used to begin the process of obtaining lead drug candidates. There are multiple ways that drug candidates are selected, including traditional immunizations to generate monoclonal antibodies and a variety of molecular display technologies.<sup>1,2</sup> The set of target molecules includes the primary human target molecule itself (often a receptor or soluble ligand). The target molecule may be custom designed with a variety of affinity tags or chemically modified as ways to facilitate the selection of drug lead molecules or the implementation of downstream assays. Other proteins that are closely related to the target molecule include structurally homologous human molecules. Depending on the biological roles of these homologous proteins, it may be desirable to target it with drug (cotarget) or

undesirable to target it with drug (off-target or liability target). Finally, there are a host of other protein reagents that are needed for various downstream secondary screening assays, biophysical studies, and multiple target constructs for structural biology.

Once the target proteins are produced and the identification of potential protein therapeutic candidates is underway, production and characterization of the second class of proteins (the drug candidates themselves) begins. The number of drug candidates produced initially depends on the method used for lead identification (i.e., immunization or different types of molecular display technologies). As these lead molecules progress toward the later stages of discovery, they are frequently formatted in order to equip them with superior biophysical, pharmacokinetic, and pharmacodynamic properties. Formatting includes bioconjugation with polyethylene glycol, molecular fusion with other proteins such as Fc, multivalency, and multispecificity.<sup>2,3</sup> Production of all these reagents, drug candidates, and formatted drug candidates requires downstream biophysical characterization, including protein structural mass spectrometry studies, to assess purity, affinity, solubility, and stability and to verify identity.

## BIOPHYSICAL CHARACTERIZATION OF PROTEIN REAGENTS AND THERAPEUTIC CANDIDATES

During the early phases of protein therapeutic discovery, much of the biochemical and biophysical characterization focus is on the protein targets and reagents. Generally, all proteins produced are evaluated for biochemical purity using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), LC-MS, and analytical size exclusion chromatography. Biophysical behaviors of the proteins are studied using light scattering and differential scanning calorimetry to assess the solution oligomeric state and thermal stability.<sup>4</sup> The drug candidate molecules are evaluated in a similar set of biochemical and biophysical assays but generally are tested under more stringent circumstances to quantify binding affinity, kinetics, solubility, and stability. Those candidates with the best set of properties undergo affinity optimization and again go through a similar set of analytical assays to select the ones with optimized affinity but also retain acceptable biophysical behavior.

Mass spectrometry as a structural characterization methodology provides unparalleled information on the protein structures.

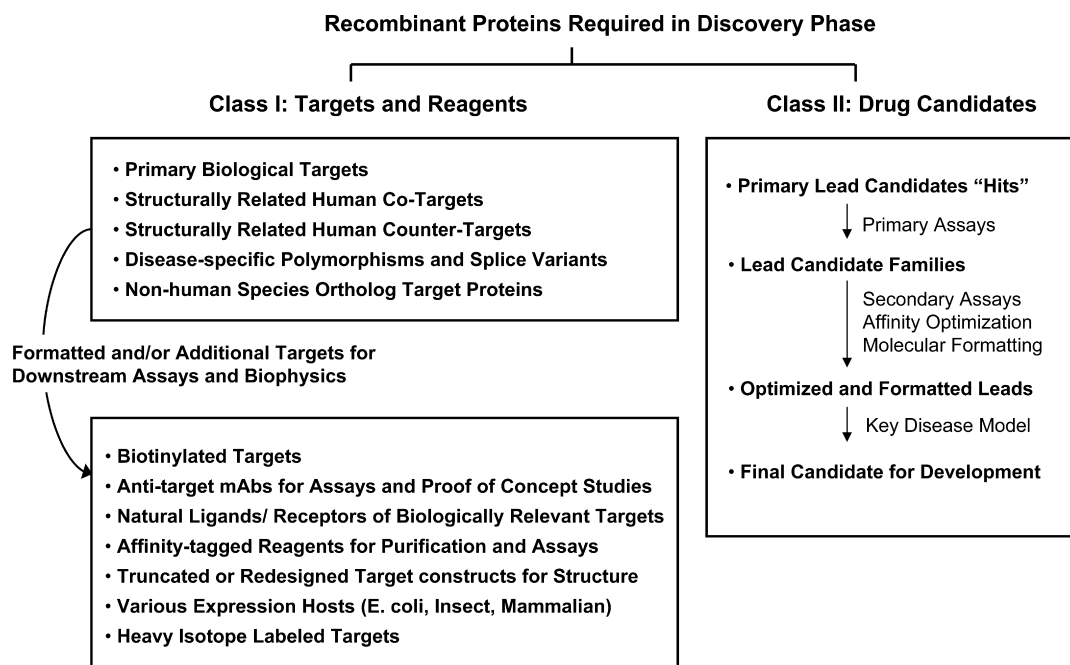
\* Corresponding author. E-mail: yeounjin.kim@bms.com. Phone: (609)-252-5115.

(1) Jackel, C.; Kast, P.; Hilvert, D. *Annu. Rev. Biophys.* **2008**, *37*, 153–173.

(2) Weissner, N. E.; Hall, J. C. *Biotechnol. Adv.* **2009**, *27*, 502–520.

(3) Labrijn, A. F.; Aalberse, R. C.; Schuurman, J. *Curr. Opin. Immunol.* **2008**, *20*, 479–485.

(4) Doyle, M. L.; Hensley, P. In *Advances in Molecular and Cell Biology*; JAI Press Inc.: Greenwich, CT and London, UK, 1997; Vol. 22A, pp 279–337.



**Figure 1.** Representation of the two classes of proteins produced during the course of a typical discovery-phase protein therapeutic project.

It has played a critical role in characterization of recombinant proteins with various formats produced for a variety of applications listed above. The following discussion on the protein mass spectrometry in protein therapeutics discovery will be focused in two main areas: mass spectrometric characterization of recombinant proteins as production quality control (QC) and structural biology support.

### CHARACTERIZATION OF RECOMBINANT PROTEINS AS PRODUCTION QUALITY CONTROL

Figure 2 illustrates a general workflow of the mass spectrometric QC in protein production. Two large categories of this workflow are whole protein analysis and proteolytic peptide analysis.

**Whole Protein Mass Analysis.** Whole protein mass analysis (aka intact mass analysis) refers to measuring the whole molecular weight of protein. The primary goal of the whole protein mass analysis is the identity confirmation of proteins. Molecular weight ( $M_w$ ) fitting to the theoretical  $M_w$  calculated from the originally designed DNA sequence (expression construct) verifies that the targeting protein has been properly translated, expressed, and purified.<sup>5</sup> In this analysis, structural modifications such as internal cleavages (chemical or proteolytic), amino- or carboxy-terminal processing, and post-translational modification (PTM) or postexpression modification (PEM) status can be monitored.<sup>5,6</sup> The structural information acquired from such analyses provides crucial feedback to improve the protein expression and purification strategy or possibly to redesign the protein construct.

As illustrated in the first inserted box of the whole protein analysis panel (Figure 2, left side), whole mass analysis is directly applied to trouble shoot protein degradation that could occur

during protein expression and purification. By correctly identifying the amino acid where the initial cleavage event occurred, this study can diagnose the major cause of the degradation. Depending upon the causes of degradation, the production strategy can be improved by introducing point mutation at the unstable site, by inhibition of protease activity, or by redesign of the purification protocol. High resolving power and mass accuracy are critical to narrow down the initial cleavage site by accurately mapping the degradation products in the original construct sequences to avoid false positive conclusions. High-end quadrupole time-of-flight (Q-TOF) instruments or benchtop Orbitrap would be a suitable instrument selection for this purpose.

The second topic in this category is PTM profiling. Recombinant proteins are purified from a variety of organisms including bacteria, yeast, plant, insect, and mammalian cells. The production systems are chosen on the basis of the expression levels and molecular features (size, PTM level, disulfide bond) of each protein, anticipated applications, and financial considerations.<sup>7</sup> The choice of the expression system is particularly of interest for mass spectrometric PTM analysis, since each expression system uses unique cellular machineries that result in different types of post-translational or cotranslational processing. Table 1 summarizes the frequent mass spectrometric PTM check points in the most common protein expression systems.

Structural analysis of the glycosylation has drawn strong attention in biopharmaceutical development<sup>8</sup> mainly because of the increasing awareness of their pharmacokinetic, immunological, and therapeutic importances.<sup>9–13</sup> Currently, engineering sugar or metabolism of the expression systems to optimize the glycan patterns are hot research areas for the biopharmaceutical industry

(5) Andersen, J. S.; Svensson, B.; Roepstorff, P. *Nat. Biotechnol.* **1996**, *14*, 449–457.

(6) Brady, L. J.; Valliere-Douglass, J.; Martinez, T.; Balland, A. J. *Am. Soc. Mass Spectrom.* **2008**, *19*, 502–509.

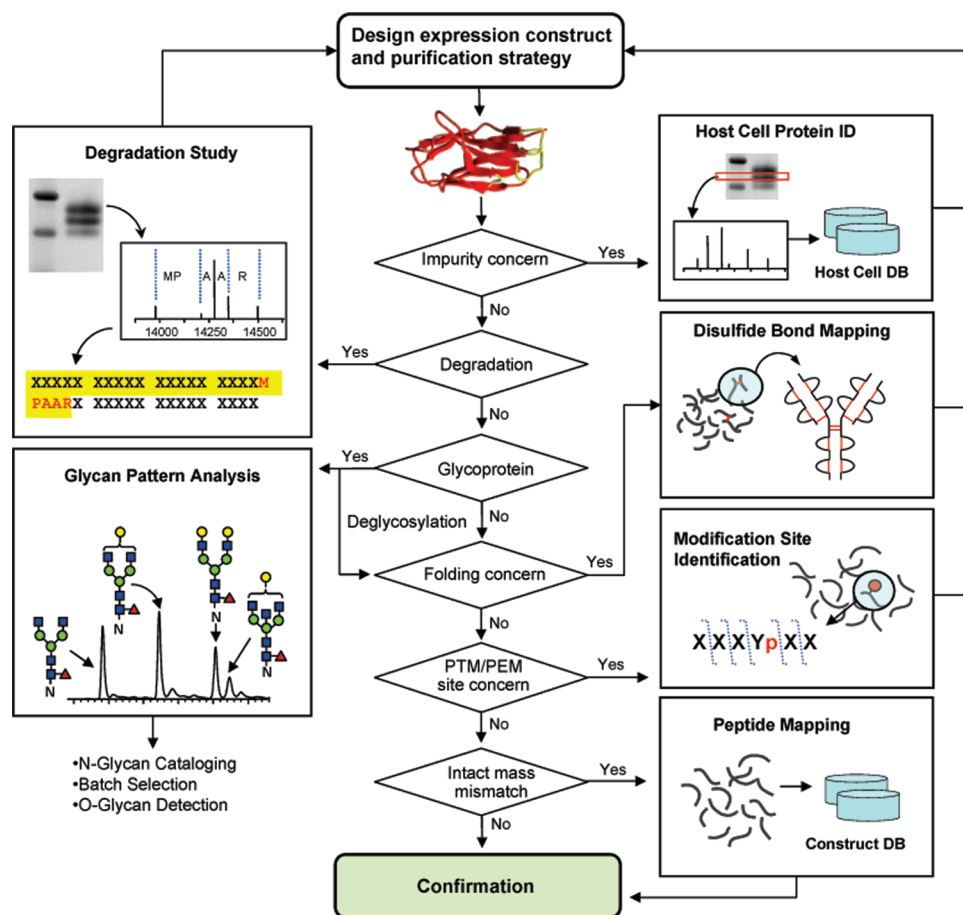
(7) Verma, R.; Boleti, E.; George, A. J. *J. Immunol. Methods* **1998**, *216*, 165–181.

(8) Burton, D. R.; Dwek, R. A. *Science* **2006**, *313*, 627–628.

(9) Jefferis, R. *Nat. Rev. Drug Discovery* **2009**, *8*, 226–234.

(10) Jefferis, R. *Trends Pharmacol. Sci.* **2009**, *30*, 356–362.

(11) Li, H.; d'Anjou, M. *Curr. Opin. Biotechnol.* **2009**, *20*, 678–684.



**Figure 2.** Mass spectrometric QC workflow in the recombinant protein production. Common problems encountered during protein production processes are shown in this diagram. Whole protein analyses are featured in the left side boxes. Internal cleavages and PTM levels can be investigated in this path. Peptide level analyses are shown in the right side boxes. The highlights in this path are identification of unknown proteins from the host cells, mapping of disulfide bonds, identification of PTM/PEM modified amino acid, and the sequence confirmation. The structural information generated from these analyses provides the feedback to improve purification strategy or expression construct design.

as emphasized by the recent acquisitions of glyco-engineering biotech by large pharmaceutical companies.<sup>14</sup>

In early discovery research, cataloging and controlling glycan patterns of recombinant proteins are important in many decision points of expression and assay development.<sup>15</sup> Glycan patterns can be analyzed by mass spectrometry as glycoprotein forms or solely carbohydrate forms after deglycosylation. Figure 3 displays whole mass analyses of glycoproteins with representative glycan patterns often found in analytical laboratories. Figure 3A shows the typical core fucosylated biantennary complex type of N-glycans attached to the IgG heavy chain expressed in mammalian cells (HEK293); Figure 3B shows the high mannose type of N-glycan attached to the glycoprotein expressed in insect cells (SF9). (See the glycosylation pattern difference summarized in Table 1.) Figure 3C is a glycoprotein with mucin type O-glycan acquired by engineering of Cys to Ser point mutation.

The biological importance of the glycosylation in the biopharmaceutical industry together with the increasing regulatory

consideration of carbohydrate contents of glycoproteins means the analytical demand in structural mass spectrometry on glycoproteins will only increase. In this aspect, an efficient data managing system that allows promptly comparing sample to sample variation in different batches and enables extracting the series of expression and purification information associated to the sample via database will be critical to fully integrate the mass spectrometric analysis into development.

Another important area for discovery QC using whole protein mass analysis is postexpression modification (PEM). Biochemical modification after protein purification is often required in discovery research for various reasons, including protein formatting for bioassay development (e.g., biotinylation), activity modification (e.g., in vitro phosphorylation), stoichiometry modulation (e.g., cross-linking), and affinity tag cleavages<sup>16</sup> to name a few. One bioconjugation approach of high current interest for therapeutic proteins is PEGylation.<sup>17</sup> PEGylation is the process of attaching poly ethylene glycol polymer (typically >20K Da) covalently to the therapeutic protein in order to improve its pharmacokinetic, pharmacodynamic, and/or immunological profile. A total of eight PEGylated proteins have been marketed after the initial success of Adagen (PEGylated form of adenosine deaminase for the

- (12) Elliott, S.; Lorenzini, T.; Asher, S.; Aoki, K.; Brankow, D.; Buck, L.; Busse, L.; Chang, D.; Fuller, J.; Grant, J.; Hernday, N.; Hokum, M.; Hu, S.; Knudsen, A.; Levin, N.; Komorowski, R.; Martin, F.; Navarro, R.; Osslund, T.; Rogers, G.; Rogers, N.; Trail, G.; Egrie, J. *Nat. Biotechnol.* **2003**, *21*, 414–421.
- (13) Egrie, J. C.; Browne, J. K. *Nephrol. Dial. Transplant* **2001**, *16* (Suppl 3), 3–13.
- (14) Sheridan, C. *Nat. Biotechnol.* **2007**, *25*, 145–146.
- (15) Hossler, P.; Khattak, S. F.; Li, Z. J. *Glycobiology* **2009**, *19*, 936–949.

(16) Malhotra, A. *Methods Enzymol.* **2009**, *463*, 239–258.

(17) Kang, J. S.; Deluca, P. P.; Lee, K. C. *Expert Opin. Emerging Drugs* **2009**, *14*, 363–380.

**Table 1. Most Common Mass Spectrometric Check Points of Protein Expression Systems**

	expression systems (most commonly used systems)	common post-translational modifications in recombinant proteins					
		glycosylation		phosphorylation acetylation acylation	N-terminal Met excision	gluconoylation <sup>b</sup> N-terminal His tag <sup>49,50</sup>	signal peptide release
		N-linked	O-linked				
bacteria	<i>E. coli</i>	no	no	no <sup>a</sup>			disulfide bond mostly incorrect from refolding
yeast	<i>S. cerevisiae</i> <i>P. pastoris</i>	high mannose no terminal $\alpha$ -1,3 mannose					
insect cell	<i>H. polymorpha</i> Sf9, Sf20, HiFive	high mannose, paucimannose, no sialic acid <sup>52</sup>	yes	yes	Met followed by small amino acids <sup>48</sup>	unlikely	may need mapping
mammalian cell	CHO, NS0, BHK, HEK	biantennary complex type with a core hepta saccharide <sup>53</sup>					mostly proper pairing, mapping may be required when the construct significantly altered the native form

<sup>a</sup> Exceptions in expression of PTM-promoting enzymes like kinases. <sup>b</sup> Phospho-gluconoylation copresent.

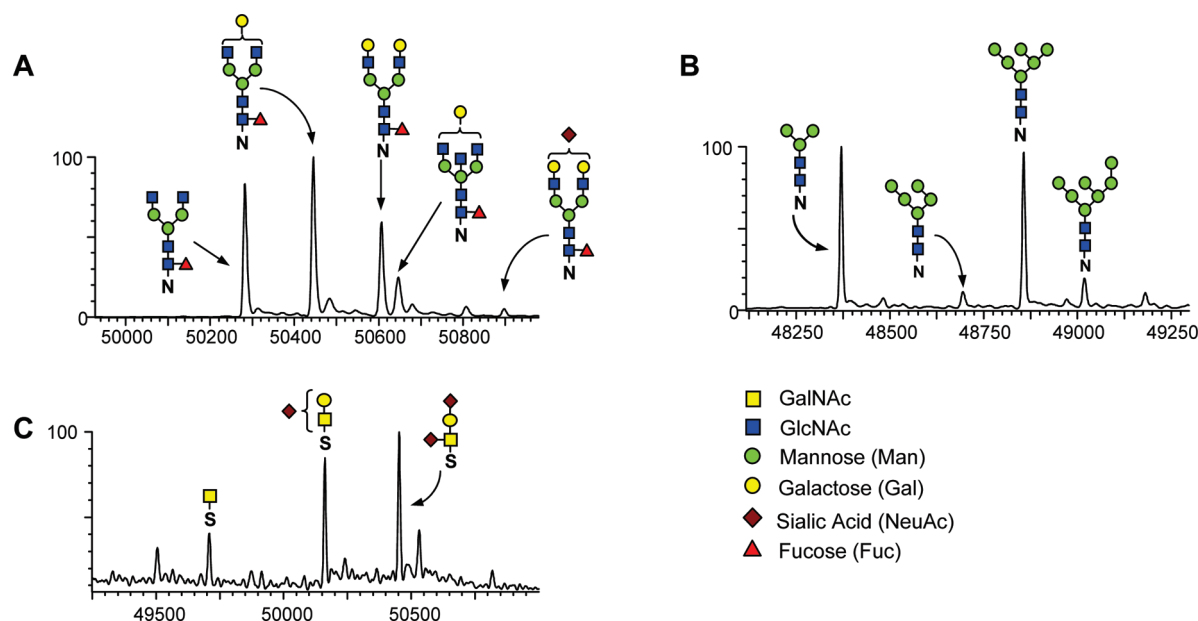
treatment of severe combined immunodeficiency disease, Enzon Pharmaceuticals) in 1990.<sup>18</sup> Although mass spectrometry has played an important role in the characterization of PEGylated proteins,<sup>19,20</sup> measuring the whole mass of PEGylated proteins is still very challenging. The complexity and polydispersity of the polymeric structural distribution and extensive charge states generated from the polyethylene glycol (PEG) molecule and protein make electrospray ionization (ESI)-MS analysis nearly impossible, even with higher resolution instrumentation. Matrix-assisted laser desorption ionization (MALDI), on the other hand, can generate a simpler form of the  $M_w$  distribution data. However, the large size of the singly charged PEG–protein conjugate limits the use of reflectron mode application, and the biased ionization and transition efficiency toward smaller molecules makes detection of un-PEGylated protein copresent in the sample more dominant. These factors impede successful analysis of PEGylated proteins.<sup>21</sup> Consequently, accurate average MW can be obtained only for relatively small polymers.<sup>22</sup>

Faced with this analytical challenge for PEGylated proteins, mass spectrometric method development to achieve good quality (distinction of polymeric ladder) with reasonable sensitivity is a hot research goal in many pharmaceutical and biotech industries. Bruker Daltonics with the scientists of Amgen showed ISD (in-source decay)-based MALDI-TOF analysis of PEGylated protein with top-down amino-terminal sequencing using 20K PEGylated protein.<sup>23</sup> Another group from Amgen and Waters Corp. reported the use of an IMS (ion mobility spectrometer) in PEGylated protein analysis with the help of a charge-stripping method and the high resolving power of ESI-IMS-QTOF instrument using 20K PEG.<sup>24,25</sup> Scientists of Eli Lilly reported an innovative method of postcolumn amine addition to reduce the charge state using an ESI-TOF instrument and showed QC application on free 40K PEGs and structural analyses of 20K PEGylated proteins.<sup>26</sup> The size of the PEG used for pharmacokinetic (PK) enhancement can be 40K Da or even larger, and the larger sizes of PEG are expected to show longer half-lives for therapeutic proteins.<sup>27</sup> Therefore, the method allowing higher  $M_w$  PEG (>40K) detection would be more practically valuable. Considering the expanding applications of large PEG and increasing regulatory requirement of the characterization of PEGylated proteins for drug approval, instrumental development of simplified arrangement including IMS and a gas-phase chemistry chamber dedicated to PEGylated protein analysis would be desirable.

**Peptide Level Analysis.** In cases when the whole protein analysis was unable to confirm the identity of protein (ID) or revealed structural issues that require microstructure information,

- (18) Jevsevar, S.; Kunstelj, M.; Porekar, V. G. *Biotechnol. J.*, **5**, 113–128.
- (19) Vestling, M. M.; Murphy, C. M.; Keller, D. A.; Fenselau, C.; Dedinas, J.; Ladd, D. L.; Olsen, M. A. *Drug Metab. Dispos.* **1993**, *21*, 911–917.
- (20) Veronese, F. M. *Biomaterials* **2001**, *22*, 405–417.
- (21) Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A. *Bioconjugate Chem.* **2009**, *20*, 384–389.
- (22) Veronese, F. M.; Mero, A. *BioDrugs* **2008**, *22*, 315–329.
- (23) Yoo, C.; Suckau, D.; Sauerland, V.; Ronk, M.; Ma, M. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 326–333.
- (24) Bagal, D.; Zhang, H.; Schnier, P. D. *Anal. Chem.* **2008**, *80*, 2408–2418.
- (25) Chakraborty, A. B.; Chen, W.; Gebler, J. In *Proceedings 57th ASMS Conference*; Proceedings 57th ASMS Conference: Philadelphia, PA, 2009.
- (26) Huang, L.; Gough, P. C.; Defelippis, M. R. *Anal. Chem.* **2009**, *81*, 567–577.
- (27) Kaminskis, L. M.; Boyd, B. J.; Karellas, P.; Krippner, G. Y.; Lessene, R.; Kelly, B.; Porter, C. J. *Mol. Pharmaceutics* **2008**, *5*, 449–463.





**Figure 3.** Three major glycosylation patterns in recombinant glycoprotein analysis. (A) The complex type N-glycans of proteins expressed in mammalian cells. (B) The high mannose type N-glycans of proteins expressed in insect cells. (C) Mucin type O-glycans of protein expressed in mammalian cells.

peptide level analysis is the next. Proteins are hydrolyzed with adequate proteolytic enzymes to generate a manageable size of constituent peptides for further LC-MS and LC-MS/MS analyses. Four major areas in this category are described in the inserted boxes at the right side of the flowchart in Figure 2. The first one depicts MS/MS-based protein identification of impurities involved in purification. These protein impurities are most likely originated from the host genome of the expression system and copurified with the target protein due to their similar physicochemical properties. Knowing the identity of the host proteins that contaminate the target protein production provides useful information to improve expression and purification strategies.<sup>28</sup> To facilitate identification of the contaminating host cell protein, it is important to have the host cell proteome databases incorporated in the search engine where MS/MS data are submitted.

The second topic in the peptide level analysis is disulfide bond mapping. Proteins made in prokaryote or lower eukaryotic cells may encounter incorrect disulfide bond pairing, leading to the production of improperly folded proteins. Even in higher organisms, mis-folding or unexpected aggregation can occur when the expression constructs have been designed with significant alteration from the natural DNA sequence. If the folding issue is believed to be due to mismatched disulfide bonding, mass spectrometric peptide mapping is the choice method for disulfide bond pairing analysis. There are several challenges, however: poor fragmentation efficiency of bridged peptides, promotion of disulfide bond scrambling during proteolysis, and complicated data analysis for cysteine abundant proteins. Although the use of prealkylating free cysteines and conducting proteolysis under lower pH condition ( $\sim$ pH 6.5) minimizes the extent of scrambling, the inability of fragmenting the S–S bond with CID (collision induced dissociation) remains as an inherent problem. Promisingly, Wu et. al demonstrated the usefulness of ETD (electron

transfer dissociation) to overcome this limitation with minimal instrumental modification.<sup>29</sup> In their report, S–S bond dissociation enabled by ETD assisted disulfide bond mapping in multicysteine containing peptides. Implementing a top-down sequencing strategy for simplistic analysis in disulfide bond mapping is also encouraging. Moreover, development of software that generates a comprehensive peptide pair  $m/z$  list from the amino acid sequence and automates SIC (selected ion chromatograph) construction from the raw data is anticipated to make disulfide mapping analysis more robust and efficient.

PTM/PEM site identification through MS/MS analysis is depicted as the third topic in the flowchart in Figure 2. This method is mature in the mass spectrometry field and has proven powerful in the broad range of applications. Both protein ID and PTM/PEM characterization often require in-gel digestion followed by LC-MS/MS to analyze separated proteins in SDS-PAGE. Although this technology has been established over the past decade and routinely applied in various studies, recovering proteolytic peptides from the gel still relies on diffusion-driven extraction which is slow, less reproducible, and not favorable to larger pieces of peptides. Along with attempts to automate the process,<sup>30</sup> multiple instrumentation companies recently introduced electrophoretic devices that allow proteins being eluted from the gel for mass spec analysis in a faster and more reliable manner. However, this area still has room for improvement, including attaining reasonable throughput by automation and minimizing artifactual oxidations during the electrophoresis.<sup>31</sup>

Lastly, peptide mapping solely depending on the peptide mass composition is used to verify protein sequence when Western blot or intact mass analyses were inconclusive. High resolving power and robust separation systems enable this method to confirm the

(28) Liu, Z.; Bartlow, P.; Varakala, R.; Beitle, R.; Koepsel, R.; Atai, M. M. *J. Chromatogr., A* **2009**, *1216*, 2433–2438.

(29) Wu, S. L.; Jiang, H.; Lu, Q.; Dai, S.; Hancock, W. S.; Karger, B. L. *Anal. Chem.* **2009**, *81*, 112–122.

(30) Visser, N. F. C. L. H.; Li, K. W.; Irth, H. *Chromatographia* **2005**, *61*, 433–442.

(31) Sun, G.; Anderson, V. E. *Electrophoresis* **2004**, *25*, 959–965.

protein sequences by mapping the  $M_w$  values of proteolytic peptides. Since the expression constructs developed for pharmaceutical research differ from the native DNA sequences in publicly available databases, building a custom construct database would be very helpful to streamline the analysis.

## STRUCTURAL BIOLOGY SUPPORT IN PROTEIN THERAPEUTIC DISCOVERY

Understanding of high-resolution 3D structures of proteins can facilitate lead selection and optimization by structure-based drug design. For this reason, structural studies by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been core activities in pharmaceutical research. Mass spectrometry has been employed in many parts of structural biology studies, both indirectly by supporting these two main technologies and directly by contributing several innovative mass spec-based methods. In X-ray crystallography, proteins that are homogeneous and minimally disordered are generally believed to yield diffraction quality crystals. Mass spectrometry supports X-ray structural studies by determining the PTM heterogeneity and purity profiles and is used to guide the expression and purification strategy to select the crystallizable population. Additionally, identifying the disordered regions of the proteins by limited proteolysis with mass spectrometry helps to optimize crystallizable constructs by removing such regions. Limited proteolysis has proven to be a powerful tool to define structured domain and is routinely used for finding expression constructs that exhibit higher solubility and greater propensity to form high quality crystals.<sup>32,33</sup> High resolution protein NMR studies require heavy isotope labeling of recombinant proteins. Heavy isotope incorporation can be done in the entire sequence using N15 and/or C13 labeled cell supplements<sup>34</sup> or by targeting a specific amino acid using a heavy isotope labeled amino acid with the medium lacking corresponding amino acid.<sup>35</sup> In both cases, mass spectrometry can monitor the heavy isotope incorporation rate and amino acid specific incorporation behavior to optimize the expression condition.<sup>35</sup>

The central goal of structural biology in the pharmaceutical industry is to understand the molecular interactions between drug candidates and their target proteins by defined 3D structures of complexes. Although X-ray crystallography and NMR studies can provide high resolution analysis of molecular interactions, they also require a period of time and effort to produce. As an alternative, mass spectrometry-driven epitope mapping strategies are promising because they can often be done with a shorter time frame with lower amount of proteins, albeit with lower resolution. The most recognized technologies in this field are hydrogen/deuterium exchange (HDX) measurement<sup>36,37</sup> and limited proteolysis.<sup>38,39</sup> Both technologies apply the concept of "footprinting" that differentiates the protected amino acid residues involved in the interaction from other residues by calculation of the kinetic profiles of HDX exchange or proteolysis rate, respectively. As successfully demonstrated in recent studies, these technologies allow one to characterize detailed epitopes and

paratopes and conformational changes upon protein-protein interaction with protein complexes that are not amenable to other structural techniques.<sup>37,40,41</sup> Mass spectrometry as a structural method does face some challenges in protein therapeutics applications. Both HDX and limited proteolysis-based MS technologies require multiple experimental points (multiple enzymes and varying incubation times and concentrations) since the addition of the experimental variance in multiple angles provides the higher structural resolution. This generates a substantial amount of data to analyze. In protein therapeutics where protein-protein interaction is involved, the data size increases even more. The key point of the data analysis in these complicated experiments is to correlate peptide signals from a multiple LC-MS data set for robust kinetic studies. Software development to tackle this challenging analysis is inevitable to make the mass spectrometry-based epitope mapping analysis more popular in biopharmaceutical research.

## FUTURE DIRECTION

The rapidly growing interest on therapeutic protein analysis in pharmaceutical industry and the instrumental advances in modern mass spectrometry are opening endless opportunities of developing novel mass spectrometric applications for protein characterization. To meet these needs, there are several areas where major changes are anticipated.

The first is changing a "mindset" in handling structural MS data. Traditionally, structural protein MS study focuses on an individual protein of interest. There has been little need for developing software or databases in this arena. However, in demands of efficient structural QC of therapeutic proteins and tackling a vast amount of LC-MS data in structural biology support, the trend needs to be changed. As pointed out in multiple subjects throughout this article, building a custom expression construct database and host cell proteome database certainly help to streamline the QC processes; building a data depository of characterization analyses such as glycan profiles of therapeutic proteins will allow efficient data maintenance, and designing software to handle data set from comprehensive epitope mapping or disulfide bond mapping studies benefit systematic structural biology support. In this regard, data processing methods developed in the MS-based proteomics field could be effectively integrated in structural MS efforts.

Another area is the instrumental and strategic development tailored to therapeutic protein characterization. One of the great potentials is employing IMS-MS to biophysical characterization of proteins. IMS separates proteins based on their shapes (collision cross sections) in addition to their size and charge states.<sup>42</sup> With subsequent MS analysis, the folding status of purified proteins can be measured quantitatively. In addition, the thermal stability of the protein can be assessed on the basis of the mobility change

(32) Tao, L.; Kiefer, S. E.; Xie, D.; Bryson, J. W.; Hefta, S. A.; Doyle, M. L. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 841–854.

(33) Gao, X.; Bain, K.; Bonanno, J. B.; Buchanan, M.; Henderson, D.; Lorimer, D.; Marsh, C.; Reynes, J. A.; Sauder, J. M.; Schwinn, K.; Thai, C.; Burley, S. K. *J. Struct. Funct. Genomics* **2005**, *6*, 129–134.

(34) Marley, J.; Lu, M.; Bracken, C. *J. Biomol. NMR* **2001**, *20*, 71–75.

(35) Strauss, A.; Bitsch, F.; Cutting, B.; Fendrich, G.; Graff, P.; Liebetanz, J.; Zurini, M.; Jahnke, W. *J. Biomol. NMR* **2003**, *26*, 367–372.

(36) Tsutsui, Y.; Wintrode, P. L. *Curr. Med. Chem.* **2007**, *14*, 2344–2358.

(37) Coales, S. J.; Tuske, S. J.; Tomasso, J. C.; Hamuro, Y. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 639–647.

(38) Dhungana, S.; Williams, J. G.; Fessler, M. B.; Tomer, K. B. *Methods Mol. Biol.* **2009**, *524*, 87–101.

(39) Zhao, Y.; Muir, T. W.; Kent, S. B.; Tischer, E.; Scardina, J. M.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4020–4024.

(40) Lu, X.; DeFelippis, M. R.; Huang, L. *Anal. Biochem.* **2009**, *395*, 100–107.

(41) Yi, J.; Skalka, A. M. *Biopolymers* **2000**, *55*, 308–318.

(42) Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Jr. *J. Mass Spectrom.* **2008**, *43*, 1–22.

with varying parameters. These applications are extremely attractive in terms of that the biophysical properties of proteins are measured at the molecular level with a very targeted manner instead of averaging the values of copresenting proteins.<sup>43</sup> A lower sample amount required for the experiment is an additional benefit. Presently, traditional biophysical technologies based on calorimetry and light scattering still hold practical advantages over IMS-MS. Improvement of the IMS-MS performance that makes this technology available for routine biophysics such as folding status and thermal stability assessment is highly encouraged. Another notable technology is top-down sequencing. In many cases shown in the section of peptide analysis, the reason of preparing proteolytic peptides is simply because of the technical limitation to handle a larger size of polypeptides in MS/MS analysis. MS manufacturers are constantly expanding the capability of top-down analysis in their instrumentations, especially focusing on improving resolution, fragmentation, and gas-phase separation.<sup>44,45</sup> With this trend, top-down sequencing will soon replace the majority of the peptide analyses discussed above.

Finally, investment needs to be encouraged in open access protein mass spectrometry platforms. As implied in Figure 1, the scope of recombinant proteins required in therapeutic development is immense. The number of proteins grows exponentially in the field of therapeutic protein discovery and, thus, so does the mass spectrometric demand. Among the various MS based characterization, straightforward whole protein mass measurement of purified proteins does not necessarily require trained mass

spectrometrists. Open access mass spectrometers in small molecule analysis for chemists have been routinely practiced throughout the pharmaceutical development processes.<sup>46</sup> Applying the same concept to protein analysis requires more attention in standardizing sample preparation, column choice, and data processing for deconvolution, yet it is a desirable solution for the current environment.<sup>47</sup>

## ACKNOWLEDGMENT

Y.J.K. acknowledges the members of GEPB department for useful discussions throughout the topics in this article.

## NOTE ADDED AFTER ASAP PUBLICATION

This paper was published on the Web on August 10, 2010 with a typographical error in the abstract and an error in reference 4. The corrected version was reposted on August 13, 2010.

Received for review June 14, 2010. Accepted July 22, 2010.

AC101575D

- (43) Smith, D. P.; Giles, K.; Bateman, R. H.; Radford, S. E.; Ashcroft, A. E. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2180–2190.
- (44) Suckau, D.; Resemann, A. *J. Biomol. Tech.* **2009**, *20*, 258–262.
- (45) Second, T. P.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Anal. Chem.* **2009**, *81*, 7757–7765.

- (46) Mallis, L. M.; Sarkahian, A. B.; Kulishoff, J. M., Jr.; Watts, W. L., Jr. *J. Mass Spectrom.* **2002**, *37*, 889–896.
- (47) White, W. L.; Wagner, C. D.; Hall, J. T.; Chaney, E. E.; George, B.; Hofmann, K.; Miller, L. A.; Williams, J. D. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 241–249.
- (48) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T. *Mol. Cell Proteomics* **2006**, *5*, 2336–2349.
- (49) Aon, J. C.; Caimi, R. J.; Taylor, A. H.; Lu, Q.; Oluboyede, F.; Dally, J.; Kessler, M. D.; Kerrigan, J. J.; Lewis, T. S.; Wysocki, L. A.; Patel, P. S. *Appl. Environ. Microbiol.* **2008**, *74*, 950–958.
- (50) Wang, Y.; Wu, S. L.; Hancock, W. S.; Trala, R.; Kessler, M.; Taylor, A. H.; Patel, P. S.; Aon, J. C. *Biotechnol. Prog.* **2005**, *21*, 1401–1411.
- (51) von Heijne, G. *Nucleic Acids Res.* **1986**, *14*, 4683–4690.
- (52) Watanabe, S.; Kokuho, T.; Takahashi, H.; Takahashi, M.; Kubota, T.; Inumaru, S. *J. Biol. Chem.* **2002**, *277*, 5090–5093.
- (53) Takahashi, M.; Kuroki, Y.; Ohtsubo, K.; Taniguchi, N. *Carbohydr. Res.* **2009**, *344*, 1387–1390.