Published in final edited form as:

Anal Biochem. 2008 August 15; 379(2): 182–191. doi:10.1016/j.ab.2008.04.037.

Detection of Protein Modifications and Counterfeit Protein Pharmaceuticals Using iTRAQ and MALDI TOF/TOF Mass Spectrometry: *Studies with Insulins*

Hongping Ye^{1,*}, John Hill², John Kauffman¹, Connie Gryniewicz¹, and Xianlin Han³

¹US Food and Drug Administration, CDER, DPA, St. Louis, MO 63101

²US Food and Drug Administration, CDER, DPMA I, Silver Spring, MD, 20993

³Division of Bioorganic Chemistry and Molecular Pharmacology, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110

Abstract

iTRAQ (isotope tags for relative and absolute quantification) reagent coupled with MALDI TOF/TOF mass spectrometric analysis has been evaluated as both a qualitative and quantitative method for the detection of modifications to active pharmaceutical ingredients derived from recombinant DNA technologies, and as a method to detect counterfeit drug products. Five types of insulin (human, bovine, porcine, Lispro, Lantus®) were used as model products in the study because of their minor variations in amino acid sequence. Several experiments were conducted in which each insulin variant was separately digested with Glu-C, and the digestate was labeled with one of four different iTRAQ reagents. All digestates were then combined for desalting and MALDI TOF/TOF mass spectrometric analysis. When the digestion procedure was optimized, the insulin sequence coverage was 100%. Five different types of insulin were readily differentiated, including Human insulin (P28K29) and Lispro (K28P29), which only differ by the interchange of two contiguous residues. Moreover, quantitative analyses show that the results obtained from the iTRAQ method agree well with those determined by other conventional methods. Collectively, the iTRAQ method can be used as a qualitative and quantitative technique for the detection of protein modification and counterfeiting.

Keywords

iTRAQ/MALDI; insulins; Glu-C; deamidation; aggregation; generic protein drugs; counterfeit

Introduction

In the next few years, biotechnology-derived products totaling over \$10 billion in sales will be coming off patent protection, making them eligible for multi-source competition. This competition has the potential to have two significant impacts on the continuing supply of these drugs. First, the demand for biotechnology-derived products will encourage

*To whom correspondence should be addressed: Hongping Ye, US Food & Drug Administration, Division of Pharmaceutical Analysis, 1114 Market Street, Room 1002, St. Louis, MO 63101, Telephone: (314)539-2174, Fax: (314)539-2113, hongping.ye@fda.hhs.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

inexperienced Active Pharmaceutical Ingredients (API) and drug product manufacturers to compete for a share of this market. Secondly, high demand and limited availability of biotechnology-derived products may provide incentives for the distribution of counterfeit or adulterated drugs. Recently, there have been several cases of counterfeit or adulterated biotechnology-derived products being sold within the United States [1,2]. Protein drugs are typically much larger and more complex than small-molecule drugs, and are often heterogeneous mixtures of molecules that vary slightly in molecular structure such as in the saccharide portion of glycoproteins. When second generation drugs involve process changes, these products may differ from originally approved products. In addition, small amounts of contaminants that can be difficult to detect may affect the function of the products. Therefore, protein drugs generally cannot be fully characterized using conventional analytical techniques. Historically both extensive analytical and clinical data have been required to ensure the quality, safety and efficacy of the products [3].

The amino acid sequence is the foundation of a protein, and any degradation or modification of a protein's primary structure may affect its biological properties and functions. Numerous conventional analytical techniques have been applied to the determination of the primary structure of therapeutic proteins. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry (MS) are used for molecular weight determination. High performance liquid chromatography (HPLC) is used for protein identification and purity analysis, and HPLC peptide mapping with ultraviolet (UV) or MS detection can be applied to obtain more detailed information on primary structure. N-terminal sequencing can provide more detailed primary structure characterization but is normally only used to analyze the first 10 to 15 amino acids in a protein. Once a protein has been identified, spectroscopic methods such as nuclear magnetic resonance (NMR) can be used to compare similarity between a sample and a standard. In combination, these techniques can provide comprehensive primary structure information [4] but uncertainty still remains. For example, the interchange of two amino acid residues may escape detection. Moreover, none of these methods can provide both primary structural details and quantitative determination, and additional analysis is required for the quantification of the protein.

Stable isotope labeling coupled with MS determination has been widely used for high-precision proteomics [5,6] since Gygi *et al.* first introduced the isotope-coded affinity tag (ICAT) technology in 1999 [7]. Among these methods, isotope tags for relative and absolute quantification (iTRAQ) is advantageous because it can analyze up to four samples (up to eight samples with recently commercialized 8-plex iTRAQ reagents) simultaneously with four isotopically distinguishable reagents [8–10]. The reagents are designed as isobaric tags consisting of a peptide reactive group, a charged reporter group that is isotopically unique to each of the four reagents, and a neutral balance portion to maintain an overall mass of 145 (Scheme I). Proteins from up to four different samples can be separately digested, the resulting peptides can be separately labeled with isotopically distinct iTRAQ reagents, and the uniquely tagged samples can be mixed together for quantitative mass spectrometric comparison of their relative concentrations. The iTRAQ reagents label lysine residues and the N termini of peptides, and with appropriate digestion conditions 100% sequence coverage is possible. Because the peptide remains attached to the isobaric tags until collision-induced dissociation (CID) is conducted, the same peptides from all four samples are simultaneously fragmented to yield strong signature *y*- and *b*-ions for sequence identification. The reporter groups are quantitatively cleaved during CID to yield an isotope series (m/z 114–117), with each isotope representing the quantity of a single peptide of known mass from each of the four samples. By comparing peak areas of reporter groups and signature ions of MS/MS spectra for up to four samples including controls, identification and quantification can be performed in the same experiment. The application of iTRAQ methodology in proteomics has been reviewed by numerous authors [5,6,11–13]. However,

this technology has not been applied to the determination of therapeutic protein primary structure for the purpose of quality control and counterfeit identification. Our purpose is to assess the capability of iTRAQ labeling for simultaneous qualitative and quantitative determination of minor modifications in therapeutic protein primary structure.

In the current study, five types of insulin were chosen because of their minor differences in primary structure. Since the isolation of insulin in 1922, this peptide hormone has been the subject of extensive clinical, biochemical and biophysical studies. The most commonly used techniques to characterize insulin have been HPLC [14–17] and LC-MS [18,19]. In this study, the iTRAQ technique coupled with MALDI TOF/TOF is used for the first time to differentiate insulin variants and to quantitatively detect deamidated and aggregated insulin. Our objective is to evaluate this technique as both a qualitative and quantitative method for the detection of modifications to API derived from recombinant DNA technologies and the ability to monitor for counterfeit and adulterated APIs. High sequence coverage is essential for successful identification of single residue modifications; therefore, the first phase of this research is optimization of digestion conditions to maximize sequence coverage.

Materials and Methods

Materials

TRAQ reagent kits were purchased from Applied Biosystems (ABI, Foster City, CA). TPCK treated trypsin from bovine Pancreas was obtained from Sigma-Aldrich (St. Louis, MO). *S. aureus* V8 (endoprotease Glu-C) and TLCK treated chymotrypsin were obtained from Worthington Biochemical (Lakewood, NJ). Insulin drug products were purchased from Washington Wholesale Drug Exchange (Savage, MD). Bovine and porcine insulin drug substances were from Eli Lilly (Indianapolis, IN). MALDI matrix material (α -cyano-4-hydroxycinnamic acid) was purchased from Sigma-Aldrich-Fluka (St. Louis, MO).

Optimization of Digestion Enzymes and Conditions

Endoprotease Glu-C and chymotrypsin were dissolved in milli-Q water at 1 $\mu\text{g}/\mu\text{l}$, and 10 μl aliquots were prepared and stored at -80°C . Trypsin was dissolved in 0.08M CaCl_2 solution at 1 $\mu\text{g}/\mu\text{l}$, and 25 μl aliquots were prepared and stored at -80°C .

Clear insulin drug product solutions were used as received and divided into 20 μl aliquots. For drug product suspensions, 0.1M HCl was added slowly until clear solutions were obtained, and the resulting clear solution was divided into 20 μl aliquots. Insulin crystals (bovine and porcine insulins) were dissolved in milli-Q water with the addition of 0.1 M HCl to make 4 $\mu\text{g}/\mu\text{l}$ solutions, and the clear solutions were divided into 20 μl aliquots. All insulin aliquots were evaporated with a SpeedVac (ThermoSavant, Holbrook, NY) at room temperature for 30 minutes until almost dry and stored at -20°C .

Digestion with trypsin supplied with the iTRAQ reagent kit was performed according to the manufacturer's protocol. Briefly, 1 μl of the denaturant (containing 2% SDS) was added to each sample tube containing 80 μg of insulin and 20 μl of 0.5M triethylammonium bicarbonate buffer (pH 8.5). The samples were vortexed and centrifuged before adding 2 μl reducing reagent (50 mM tris-(2-carboxyethyl)phosphine), and were then mixed and incubated at 60°C for one hour. After reduction, the samples were centrifuged (Spectrafuge Mini Complete, ML MarketLab, Caledonia MI) for 30 seconds to bring the sample to the bottom of the tubes, and cysteine residues were blocked with 1 μl of 200 mM methyl methanethiosulfonate at room temperature for 10 minutes. To each sample tube, 8 μl of trypsin solution was added. Samples were vortexed and incubated at 37°C overnight (12 to 16 hours).

With Glu-C and chymotrypsin, insulin digestion under both non-denaturing (non-reducing) and reducing conditions were compared. Different ratios of protease to insulin from 1:20 to 1:2.5 and different digestion times from 2 to 17 hours were compared to optimize the digestion procedure for maximum sequence coverage and minimum enzyme auto-cleavage. For both Glu-C and chymotrypsin the ratio of 1:10 (protease to insulin) and 4 hours of digestion time were found to be the optimum conditions. These conditions were used throughout the current work. The digestion procedure with reducing conditions was similar to trypsin digestion except 8 μ l of trypsin solution was replaced by 8 μ l Glu-C or chymotrypsin solution. Under non-reducing conditions, the protease solution was added to the insulin solution in the buffer without going through the denaturing and reducing procedures as described above.

Isobaric Labeling and Mass Spectrometric Analysis

The content of one iTRAQ reagent vial (in 70 μ l of ethanol) was transferred to one sample tube, and incubated at room temperature for 1 hour. The samples labeled with different tags were combined, vacuum dried with a SpeedVac and re-dissolved in 0.1% TFA in water. The pH of sample solutions was adjusted to below 4 with 10% TFA in water. A portion of each sample was desalted and concentrated with a C₁₈ ZipTip (Millipore, Billerica, MA), mixed at a 1:2 ratio with the MALDI matrix material (7.5 mg/ml of α -cyano-4-hydroxycinnamic acid) and spotted on a MALDI plate with 12 to 24 spots per sample. The samples on the MALDI plates were analyzed with either an ABI 4700 proteomics analyzer or an ABI 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA). Mass spectra in the range of m/z 450 to 4000 were acquired in positive ion mode with internal mass calibration. The eight most intense ions per spot were selected for subsequent MS/MS analysis by collision induced dissociation (CID) at a collision energy of 1 kV and a collision gas pressure of $\sim 1.5 \times 10^{-6}$ Torr. Each spectrum was averaged over 2000 laser shots. For peptide identification purposes, the MS/MS spectra were acquired with the CID gas off.

Data processing

Peptide identification was performed by searching the MS/MS spectra against the SwissProt database using either a local MASCOT search engine on the GPS (v.3.6, ABI) server or an internet MASCOT search engine from Matrix Science (www.matrixscience.com). The following search parameters were used for search: digestion with one missed cleavage was selected; mass tolerance was 50 ppm for the precursors and 0.3 Da for the MS/MS ions; iTRAQ-labeled N-termini and lysines and MMTS-labeled cysteines were set as fixed modifications; oxidized methionines and iTRAQ-labeled tyrosines were set as variable modifications.

For quantitative analysis, the iTRAQ Reagent signature peak areas for each reporter ion mass (m/z 114.1, 115.1, 116.1, 117.1) were extracted from corresponding MS/MS spectra with a tolerance of $m/z \pm 0.2$ [8]. The obtained data were transferred to Excel, and the isotopic carryover was corrected based on the data provided by the manufacturer in the Certificate of Analysis. Most experiments were performed under non-reducing digestion in which all disulfide bonds should be intact. However, to avoid inconsistent results caused by adventitious disulfide bond cleavage, only peptide B22–30 (B22–32 for Lantus®) was used for quantification unless otherwise noted.

Results

Sequence Coverage with Different Endoproteases

The objective of this study is to determine if the iTRAQ procedure can quantitatively measure all amino acid differences between pairs of similar proteins. Therefore, high

sequence coverage is required. Endoprotease trypsin is included in Applied Biosystems' iTRAQ reagents kit. Trypsin cleaves peptide bonds on the carboxyl side of lysine (unless it is followed by a proline residue) and arginine. Its specificity is appropriate for most protein pharmaceuticals giving the right peptide size for mass spectrometric analysis. Scheme II illustrates the primary structures of the insulin variants examined in this study. In the preliminary experiment, trypsin was used to digest Lispro according to ABI's protocol. Two peptides from chain B were identified (Table I) and confirmed by MASCOT search. Chain A, comprised of 21 amino acids, was not cleaved and was not detectable. To increase the sequence coverage, proteases chymotrypsin and Glu-C were examined as potential digestion enzymes.

Owing to the relatively low specificity of chymotrypsin, digestion of Lispro under non-reducing conditions gave complicated peptide mixtures that were not consistent between experiments. Under reducing conditions, the MS spectra deduced from the peptide mixtures were consistent between experiments and clear enough for identification and quantification. However, the sequence coverage was not improved compared to trypsin, although some peptide fragments from chain A were identified (Table I).

To evaluate the enzyme Glu-C, the digestion of Lispro under reducing and non-reducing conditions was first compared without iTRAQ labeling (Figure 1A). Under reducing conditions, the MS spectrum was very simple, and the only two peptides detected were B22–30 (m/z 1116.58) and B1–13 (m/z 1528.69 with cysteine protected). In very rare cases, the peptide B14–21 (m/z 913.39 with cysteine protected) could be seen at low abundance. In contrast, under non-reducing conditions, several peptide peaks were distributed in different MS ranges with high abundance. Most importantly, all amino acid residues could be found in these peptides, i.e., the sequence coverage was 100%. Therefore, Glu-C under non-reducing conditions was chosen for most experiments in this study. The main peptides are listed in Table I.

The iTRAQ labeling was also found to enhance the MS signals. When Lispro was digested with Glu-C under reducing conditions and labeled with iTRAQ, several additional MS peaks that were not present in the unlabeled spectrum were observed with high abundance. These precursor peaks corresponded to peptides that were missing before iTRAQ labeling (Figure 1B). Therefore, under both reducing and non-reducing conditions, Lispro sequence coverage was 100% with Glu-C and iTRAQ labeling. All amino acid residues could be observed and protein variations could be discerned.

Extensive cleavage occurred between residues A12 and A13 under both reducing and non-reducing conditions. Although this is not the specific cleavage site of Glu-C, preparations of Glu-C have been found previously to cleave the carboxyl side of serine residues of insulin[20] and phospholipase A₂[21]

Identification of Five Different Types of Insulin and Detection of Two Interchanged Amino Acid Residues

As described above, when Glu-C was used to digest insulin, the sequence coverage was 100%. To discover how efficiently the iTRAQ method could identify minor amino acid sequence modifications, four different insulin variants were compared. Equal amounts (calculated on the basis of label claim) of bovine insulin, porcine insulin, human insulin and insulin Lispro were digested separately and each was labeled with one of the four different iTRAQ reagents (see the table in Scheme 2 for labeling scheme.) The variants were then mixed and analyzed by MS. Four separate mass ranges were scanned to cover m/z 400 to 4000, and the six or eight most abundant precursor peaks from each mass range were selected for MS/MS. The peak areas corresponding to the reporter groups for all four

insulins were extracted and sorted by precursor masses. Peak area ratios of all four insulin samples to the human insulin peak area can be rapidly reviewed to identify primary structure differences. Theoretically, precursor peaks corresponding to peptides that are conserved across all four samples should display four identical peak ratios with a value of 1. When ratios of zero or infinity are observed, the corresponding peptide is only present in one sample of the pair used to compute the ratio. Accordingly, when a ratio is between 0 and 1, the amounts of the peptide in two samples are different. One of the proteins is partially modified at some site of this peptide, and the ratio reflects the percentage of the modification.

Figure 2 illustrates the identification of conserved peptide fragments across four samples using the iTRAQ reagents. The MS spectrum of the mixture of four digested protein samples over the 420–1850 m/z range is presented in Figure 2A. Inset *A1* is the reporter mass region (m/z 113–118) of the MS/MS spectrum for the precursor peak at m/z 1374.96, which corresponds to the peptide B28–30 that is conserved in bovine (mass 114) and porcine (mass 116) insulins. Large signals are observed at masses 114 and 116, but not at masses 115 (Lispro) or 117 (human). The reporter region of the MS/MS spectrum for the same peptide (precursor at m/z 1404.97) of Lispro (mass 115) and human insulin (mass 117) is shown in inset *A2*, and strong signals are only observed at masses 115 and 117. Inset *A3* is the reporter region of the MS/MS spectrum for the precursor peak at m/z 1665.97 (peptide A18–21+B14–21), which is conserved across all four variants, and all four reporter masses are observed in the MS/MS spectrum of the reporter region. Figure 2B displays the MS spectrum of the mixture over the 2850–3150 m/z range. Inset *B2* shows the reporter mass region for the m/z 3008.43 precursor, corresponding to peptide A1–12+B1–13, which is conserved for Lispro, porcine and human insulin, and signals from their mass tags are evident in the reporter MS/MS spectrum. Bovine insulin has a modification at A8 and A10, and its reporter mass alone is observed in the MS/MS spectrum of the m/z 2964.40 precursor peak (inset *B1*) for peptide A1–12+B1–13.

The peak intensities of reporter groups of four conserved precursors (m/z 809, 1011, 1627, and 1666 as shown in Table I) from the same experiments were used to determine the relative amount of these samples. The data from two separate labeling experiments and three MS determinations were used for quantification. For each MS determination, MS/MS data were acquired on at least five sample spots. All data were combined and it was found that the ratios of bovine, porcine and Lispro to human insulin were 0.90 (± 0.10), 0.94 (± 0.11), 0.98 (± 0.09), respectively.

Lantus® is a recombinant human insulin analog that is a long-acting, parenteral blood-glucose-lowering agent. Its primary structure is similar to human insulin except Asn is replaced by Gly at position A21 and two more Arg are added to the C-terminus of chain B (Scheme II). When a mixture of Glu-C digested human insulin and insulin Lantus® (labeled with tag 114) was desalted and analyzed using MALDI TOF/TOF MS, the differences in their primary structures were easily identified by the MS and MS/MS spectra and reporter signal intensity (spectra not shown). The difference in chain A was determined by performing MS/MS on the precursor peptide A18–21+B14–21. The Lantus® precursor (m/z 1609.00) only reported a mass 114 signal, and the human precursor (m/z 1666.02) only reported a mass 117. The chain B modification was determined with peptide B22–30 for human insulin at m/z 1405.02, and peptide B22–32 for insulin Lantus® at m/z 1717.24.

Because the target protein (insulin in this study) is digested before isobaric labeling and each peptide is labeled at least once (each lysine side chain and N-terminal group of the peptide), as long as the sequence coverage is 100%, any primary structure modification between two proteins that result in non-conserved peptide masses can be narrowed down to a peptide

from the MS/MS reporter masses. Further analysis of the full MS/MS spectra for the suspect peptide can be conducted to elucidate the exact modification site on the basis of the informative *y*- and *b*-ions. This procedure can also be automated using a search engine such as MASCOT. These methods do not require iTRAQ labeling, but the simultaneous analysis of two proteins facilitates the identification and quantitation of the suspected modified peptide. Subsequent selection of the distinct precursor peaks of the intact and modified peptide for MS/MS analysis simplifies the identification of the modified residue. Figure 3 provides an example in which two proteins (Lispro, K28P29 and Human insulin, P28K29) that differ only by the interchange of two amino acids were determined by MASCOT search. The search results included the peptide sequence and matched *y*-, *b*-, *b**-, *a*, *a** and several other ions as labeled in the figure. Figure 3 demonstrates that the MS/MS spectra of precursor B22–30 for insulin Lispro (Figure 3A) and human insulin (Figure 3B) are comparable, but there are several distinct *yb* ions for insulin Lispro: FYTK with *m/z* 684.38, KP with *m/z* 370.26, TK with *m/z* 374.25, YTK with *m/z* 537.32, and one *y*₂ ion PT with *m/z* 217.17. In addition, a group of peaks (top right inset) which appear in the MS/MS spectrum of Lispro are not observed in the human insulin spectrum. This may be caused by fragmentation differences between the KP and PK sequences. This example demonstrates that any primary structure modification including two reversed amino acid residues can be identified qualitatively by examining their MS/MS spectra. The modification can usually be quantified from the reporter group intensities, except in cases where the modification does not alter the fragment mass, such as the interchange of two amino acid residues within a peptide.

Identification and Quantification of Deamidated Human Insulin

Deamidation of Asn residues is one of the most common protein degradation pathways, and has been used as a marker for the stability of protein drugs. Deamidation of Asn and to a lesser extent Gln residues is a well known reaction that proceeds more or less rapidly depending on the structural and sequence context [22], and is catalyzed both at basic and acidic pH. Human insulin has been reported to deamidate at B3 and A21 at elevated temperature with shaking [23]. Deamidation has been most often detected by analytical methods that respond to changes in surface charge such as isoelectric focusing and ion-exchange HPLC [9,24]. To determine if the iTRAQ technique can be used to identify and quantify deamidation, a sample was stressed by preparing a 1.6 mg/ml solution from crystalline human insulin in 0.01 M HCl and holding it at room temperature for five days. The sample was analyzed by HPLC according to the USP monograph for the assay of insulin [25]. A small peak that eluted behind the main insulin peak was assigned as desamido-A21 and was determined to be 6.3% of the total insulin in the sample using the chromatogram's peak area. The same sample and a control were separately digested with Glu-C under non-reducing conditions and labeled with different iTRAQ tags as described in the Materials and Methods section. The sample was labeled with iTRAQ tag 115 and the control was labeled with tag 116. They were then mixed together and analyzed by MALDI TOF/TOF MS. As discussed above, all acquired reporter peak areas were compared for each precursor peak, and they were all statistically indistinguishable except the precursor peak for peptide A18–21+B14–21, which includes the deamidated Asn-21 residue. Figure 4A shows the MS spectrum of the intact (control) and desamido forms of the A18–21+B14–21 peptide, and iTRAQ tag reporter regions for their MS/MS spectra. Figures 4B and C compare the full precursor MS/MS spectra of the intact and desamido-A21 peptides, and careful analysis of these spectra indicates that all *b* ions, which do not bear A21, are comparable between the intact (control) and desamido-A21 peptides. However, all *y* ions, which bear the A21 residue, are different by 17 units because of the deamidation of A21 residue.

The difference in the amount of desamido-A21 for sample and control can be seen visually from the reporter region of the MS/MS spectra of the deamidated precursor peptide (m/z 1648.75, inset A1), and the intact precursor peptide (m/z 1665.76, inset A2). To determine the amount of desamido-A21, two separate labeling experiments and three MS determination were performed. For each MS determination, MS/MS data for the precursor m/z 1665.76 on at least 12 sample spots were acquired, and the reporter peak intensities for the sample and control were extracted. The ratio of sample (m/z 115) to control (m/z 116) averaged over all MS measurements was 0.94 (± 0.05). Therefore, the sample contains 6% more desamido-A21 than the control.

The above data demonstrate that the iTRAQ method can give quantitative results similar to the USP's HPLC method. HPLC is appropriate for quantitative analysis when all analytes are well separated. The accuracy and precision of the HPLC method is normally better than quantitative analysis by MS. However, with iTRAQ technology, all samples are combined for desalting and MS analysis, which eliminates the analytical error caused by sample processing and unpredictable changes in MS instrument conditions between runs. Therefore, the quantitative results obtained with iTRAQ reagents are much more reliable than traditional MS methods. MS analysis by the iTRAQ method identifies and quantifies protein modifications in a single experiment based on comparison of peptide masses and reporter group intensities; whereas, protein identification and quantification by HPLC relies on comparison of residue and peptide retention time and peak area with standards. HPLC analysis is particularly troublesome when two or more compounds in a mixture have the same retention time.

Determination of Aggregated Human Insulin

One goal of the current study is to determine the range of structural modifications that can be elucidated with iTRAQ methods. Aggregation is a common protein degradation pathway that has been addressed with several analytical techniques. Size exclusion chromatography with multiangle laser light scattering detection is one of the most advanced techniques in the characterization of aggregates [26]. The iTRAQ methods often begin by denaturing, reducing and cysteine protecting the proteins under study in order to obtain as much primary structural and quantitative information as possible. Denaturation changes the non-covalent interactions between molecules, and reduction decomposes aggregates formed through covalent disulfide bridges; therefore, these processes will hinder the identification of protein aggregates. To evaluate the applicability of iTRAQ reagents for determination of aggregation in therapeutic protein products, a stressed insulin sample was compared with a control. The stressed sample was shaken at 60°C to form aggregates according to the literature procedure [23]. After shaking for 152 hours, the sample solution became viscous, but remained clear. Aggregate content was determined using size exclusion chromatography, following the procedure for the limit of high molecular weight proteins listed in the USP monograph for Insulin Human Injections. The sample was found to be 81% aggregated[25]. The control sample had 0.3% aggregation. The same stressed and control samples were digested with Glu-C under non-reducing conditions as described in the Materials and Methods section, labeled with separate iTRAQ reagents, mixed and desalted for MALDI TOF/TOF MS analysis. Two reporter masses were always observed in the reporter region during MS/MS analysis of the iTRAQ labeled precursor peptides, indicating that no detectable primary structural changes occurred after the stress was applied. However, the signal intensity of the reporter group corresponding to the stressed sample was much lower than the signal intensity of the control. To determine the amount of aggregates and further investigate the types of aggregation (covalent or non-covalent), several labeling experiments under different conditions were conducted and repeated. For each experiment, the stressed sample was paired with a control sample for MS and MS/MS analysis. The MS/

MS data was acquired on more than 12 sample spots. The signal intensities of reporter peaks were extracted and compared. Under non-reducing conditions, the signal intensity of the stressed sample was only $16.4 \pm 0.5\%$ of the signal intensity of the control, indicating that 84% of the stressed sample was not digested and labeled. This result is consistent with the aggregate content of the sample measured with the SEC reference method, and indicates that iTRAQ methods can be developed to probe protein aggregation. When the sample and control were denatured and incubated with reducing reagent at 60°C for one hour, the signal strength of reporter tag for the sample was increased to $53.4 \pm 1.7\%$ of the control signal. When the incubation time at 60°C was extended to two hours, the signal strength of the sample was increased to $67.1 \pm 1.4\%$ of the control sample. It has been reported that the insulin polymer formation is mainly due to a chain reaction involving disulfide interactions [27]. Incubating with a reducing reagent at 60°C decomposes some covalent bonds between molecules and frees some insulin molecules. Longer incubation times (3 and 4 hours) did not increase the signal strength any further compared to control. However, some unexpected peptide precursors were observed and may be the result of unrestricted protein decomposition under higher temperature with longer incubation time. The results indicate that the iTRAQ technique can be used to identify and quantify protein aggregation by comparison of the reporter mass signal strengths of test and control samples. If non-denaturing, non-reducing conditions are used for normal analysis, one sample and one control which are treated under denaturing and reducing conditions can be included for MS analysis. The reporter group intensity ratios of sample to control from two digestion conditions are compared. If the sample/control ratio from reducing conditions is significantly higher than the sample/control ratio from non-reducing conditions, then the sample/control ratio from non-reducing conditions represents the un-aggregated portion of the analyzed protein. In some cases, especially when sample and control are from different sources, the ratio only represents how much “good protein” is present compared to control. The percentage of aggregates can not be determined when the total protein in sample and control are different.

Comparison of Novolin to Humulin

Recombinant human insulin is sold in three common formulations: regular, NPH and a 70/30 mix of NPH:regular commonly referred to as 70/30. The primary structure of these products should be independent of the manufacturer, and their potencies should agree with the label claim. The results described above suggest that iTRAQ methods can be developed to qualitatively compare the primary structure and quantitatively compare the potencies of insulins prepared by different manufacturers. Novolin NPH, Novolin R and Novolin 70/30 manufactured by Novo Nordisk (Princeton, NJ) were compared with Humulin NPH, Humulin R and Humulin 70/30 manufactured by Eli Lilly (Indianapolis, IN). Three pairs of samples were prepared for comparative analysis: Novolin NPH versus Humulin NPH, Novolin R versus Humulin R, and Novolin 70/30 versus Humulin 70/30. Each sample in a pair was individually labeled as described previously, and both samples for each pair were then combined for desalting and MALDI TOF/TOF MS analysis. The MS/MS spectra of the iTRAQ reporter regions for all precursor masses appeared as pairs of peaks, indicating that samples from both manufacturers were composed of the same peptide fragments. The analysis revealed that no precursor peak originated from only one sample, verifying the qualitative similarity of the primary structures of products from both manufacturers.

Quantitative studies were performed by preparing two duplicate labeling runs for each sample, and then performing two separate MS measurements on each duplicate pair of samples (i.e., four measurements of each sample were performed). For all three pairs of human insulin evaluated in this study, the amount of insulin in Humulin was found to be higher than the amount of insulin found in Novolin as shown in Table II. These results were

independent of the reporter groups employed. When the same lots of Humulin 70/30 and Novolin 70/30 were assayed by HPLC, the ratio of Humulin 70/30 to Novolin 70/30 was found to be 1.03 (± 0.01). As discussed previously, quantitative results by HPLC are normally more accurate than results from traditional MS method when the chromatographic peaks are well resolved. Comparative analysis by iTRAQ appears to have accuracy similar to HPLC, though the precision is not as high. However, mass spectrometric analysis with iTRAQ labeling has the important advantage that the mass spectrum can be used for qualitative identification based on fragment mass, even in the absence of standards; whereas, HPLC analysis requires standards for both qualitative and quantitative analysis.

Discussion

Unlike clinical tissue or cell culture samples which may contain more than hundreds of proteins, recombinant therapeutic protein products are normally purified and contain only one or few proteins. Therefore, these products are often suitable for direct mass spectrometric analysis of labeled peptide mixtures, without the need for multi-dimensional liquid chromatographic separations. In addition, because the expected primary sequence and mass of these products are known, it is possible to predict peptide fragments that can be used for identification and quantitative analysis based on digestion conditions. As a result, any missing or extra peptides in the target sample can be identified relatively easily by mass spectrometric comparison with control. The iTRAQ reagents simplify this process, because tagged samples and controls are mixed before performing the processing steps that typically lead to quantitative variations in sample composition, and are then simultaneously measured in the mass spectrometer.

Methods based on iTRAQ reagents appear to be suitable for regulatory analysis and counterfeit detection, but these tasks will often require 100% sequence coverage, because modifications to peptides that are not detected will not be observed. Furthermore, natural proteins can often be identified by database search engines based on a small number of peptides, but biotechnology-derived therapeutic proteins are, in most cases, natural protein analogs with several amino acid substitutions. Therefore, database searches cannot be used to verify the protein primary structure for regulatory purposes unless they are based on a large number of fragments, which requires high sequence coverage. In mass spectrometric analysis, sequence coverage of less than 50% is not uncommon, and the work presented here indicates that optimization of digestion conditions may be critical to the successful identification of minor protein modifications. This work also demonstrated that labeling of peptides with iTRAQ reagents can enhance the MS signal, either by improving fragmentation and/or improving ionization efficiency resulting in an increase in sequence coverage as demonstrated with the labeling of the reduced Lispro sample. This observation is consistent with previous reports of improved sequence coverage with iTRAQ labeling [28]. In addition, the iTRAQ method has inherently higher signal levels when more than one sample is labeled and combined for analysis because of the additivity of precursor ion signals in the mass spectrometer allowing for the enhancement of individual peptide signals that may be in low abundance in any given sample. Whereas, the reporter group ions appear as distinct peaks between m/z 114–117 following CID in the MS/MS signal, the informative y - and b - ions used for sequence identification remain as additive isobaric signals, which increase the confidence of sample identification.

One of the primary disadvantages of iTRAQ labeling for protein identification is the large number of experiments that are required to develop confidence in the conclusions. With iTRAQ reagents, all peptides of a target protein are labeled at least once, and protein identification is typically based on the accumulated MS/MS results from many precursor ions. Furthermore, numerous MALDI runs (typically 10 spots per sample in this work) and

multiple labeling experiments are required to verify the qualitative and quantitative results from the experiments described herein. Because differences in peptide levels can only be determined after tandem MS when using methods based on iTRAQ, each and every major peptide must be subjected to tandem MS analysis, making iTRAQ time-consuming. This disadvantage can be partly overcome by automating acquisition, but the time between receiving a sample and recording a result is still significant. Quantitative analysis requires even more data collection, and the quantitative results reported here are the averages of at least 12 data points. More complete quantitative evaluation of the iTRAQ method including impurity detection limits and statistical analysis of the data will be published separately.

The results presented here demonstrate that the iTRAQ technology can be used to qualitatively and quantitatively characterize modifications to protein API and detect protein counterfeits in which one variant of a therapeutic protein is substituted for another. iTRAQ technology may also provide an alternative analytical method in support of pharmaceutical manufacturing such as the characterization of protein post-translational modifications that can alter the biological potency of biotechnology-derived products. Detection of such changes might indicate a manufacturing process that is operating outside its design space and may be an early indicator that further manufacturing process development is required.

Conclusions

The iTRAQ technology has the potential to provide much more information than analytical methods currently used by the pharmaceutical industry to characterize and quantify protein pharmaceuticals. It can be used to qualitatively and quantitatively characterize up to four samples in a single experiment. Any primary structure modifications including post-translational modifications can be identified as long as the sequence is fully covered. Changes in the peptide profiles are indicative of changes in the API due to either manufacturing process irregularities, or counterfeit or adulterated API. When coupled with multi-dimensional analysis, iTRAQ chemistry has the potential to monitor the efficiency and control of the fermentation process. iTRAQ may also offer a unique method to compare multiple crude drug substance lots derived from extended cell cultures and evaluate them for consistency of manufacture. Because iTRAQ tags offer the ability to perform differential profiling of multiple test samples, test samples and reference materials, or test samples and calibration controls, this technique is also a useful method to demonstrate similarity between approved and biosimilar products. Finally, the application of the iTRAQ chemistry may provide an alternative method for monitoring protein changes or product-related impurities introduced by the purification process.

Acknowledgments

This work was supported through the CDER Regulatory Science and Review (RSR) Enhancement Program, RSR #06-03.

Abbreviations

iTRAQ	isotope tags for relative and absolute quantification
API	active pharmaceutical ingredients
MALDI	matrix-assisted laser desorption ionization
TOF	time of flight
CID	collision induced dissociation

References

1. <http://www.fda.gov/medwatch/SAFETY/2001/serostim.htm>. 1-22-2001.
2. <http://www.fda.gov/bbs/topics/NEWS/2003/NEW00877.html>. 3-11-2003.
3. Woodcock J, Griffin J, Behrman R, Cherney B, Crescenzi T, Fraser B, Hixon D, Joneckis C, Kozlowski S, Rosenberg A, Schrager L, Shacter E, Temple R, Webber K, Winkle H. The FDA's assessment of follow-on protein products: a historical perspective. *Nat Rev Drug Discov*. 2007; 6:437–442. [PubMed: 17633790]
4. Jiskoot, W.; Crommelin, DJA, III. *Biotechnology: Pharmaceutical Aspects*. Arlington: American Association of Pharmaceutical Scientists; 2005. *Methods for Structural Analysis of Protein Pharmaceuticals*.
5. Putz S, Reinders J, Reinders Y, Sickmann A. Mass spectrometry-based peptide quantification: applications and limitations. *Expert Rev Proteomics*. 2005; 2:381–392. [PubMed: 16000084]
6. Schneider LV, Hall MP. Stable isotope methods for high-precision proteomics. *Drug Discov Today*. 2005; 10:353–363. [PubMed: 15749284]
7. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 1999; 17:994–999. [PubMed: 10504701]
8. iTRAQ Reagent Quantification using 4700 Proteomics Analyzer and GPS Explorer v2.0 Software. Technical Note. Jun 12. 2007 www.appliedbiosystems.com
9. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004; 3:1154–1169. [PubMed: 15385600]
10. Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics*. 2007; 7:1004–1014.
11. Aggarwal K, Choe LH, Lee KH. Shotgun proteomics using the iTRAQ isobaric tags. *Briefings in Functional Genomics and Proteomics*. 2006; 5:112–120. [PubMed: 16772272]
12. Gafken PR, Lampe PD. Methodologies for Characterizing Phosphoproteins by Mass Spectrometry. *Cell Commun Adhesion*. 2006; 13:249–262.
13. Zieske LR. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot*. 2006; 57:1501–1508. [PubMed: 16574745]
14. Moslemi P, Najafabadi AR, Tajerzadeh H. A rapid and sensitive method for simultaneous determination of insulin and A21-desamido insulin by high-performance liquid chromatography. *J Pharm Biomed Anal*. 2003; 33:45–51. [PubMed: 12946530]
15. Oliva A, Farina J, Llabres M. Fingerprint Analysis of Insulin: Application in Stability Studies of Pharmaceutical Preparations. *Drug Dev Ind Pharm*. 1997; 23:127–132.
16. Oliva A, Farina J, Llabres M. Development of two high-performance liquid chromatographic methods for the analysis and characterization of insulin and its degradation products in pharmaceutical preparations. *J Chromatogr B*. 2000; 749:25–34.
17. Purcell AW, Aguilar MI, Hearn MTW. Conformational effects in reversed-phase high-performance liquid chromatography of polypeptides I. Resolution of insulin variants. *J Chromatogr A*. 1995; 711:61–70. [PubMed: 7496495]
18. Fierens C, Stockl D, Thienpont LM, De Leenheer AP. Strategies for determination of insulin with tandem electrospray mass spectrometry: implications for other analyte proteins? *Rapid Commun Mass Spectrom*. 2001; 15:1433–1441. [PubMed: 11507756]
19. Thevis M, Thomas A, Delahaut P, Bosseloir A, Schanzer W. Doping Control Analysis of Intact Rapid-Acting Insulin Analogues in Human Urine by Liquid Chromatography-Tandem Mass Spectrometry. *Anal Chem*. 2006; 78:1897–1903. [PubMed: 16536426]
20. Rose K, Savoy L-A, Simona MG, Offord RE, Wingfield P. C-Terminal peptide identification by fast atom bombardment mass spectrometry. *Biochem J*. 1988; 250:253–259. [PubMed: 3281660]
21. Evenberg A, Meyer H, Gaastra W, Verheij HM, De Haas GH. Amino Acid Sequence of Phospholipase A2 from Horse Pancreas. *J Biol Chem*. 1977; 252:1189–1196. [PubMed: 838712]

22. Bischoff R, Kolbe HV. Deamidation of Asparagine and Glutamine Residues in Proteins and Peptides: Structural Determinants and Analytical Methodology. *J Chromatogr Biomed Appl.* 1994; 662:261–278.
23. Oliva A, Farina JB, Llabres M. Influence of temperature and shaking on stability of insulin preparations: degradation kinetics. *Int J Pharm.* 1996; 143:163–170.
24. Bischoff, R.; Barroso, B. Liquid Chromatography. In: Jiskoot, W.; Crommelin, DJA., editors. *Methods for Structural Analysis of Protein Pharmaceuticals.* American Association of Pharmaceutical Scientists; Arlington: 2005. p. 277-329.
25. United States Pharmacopeia. National Formulary. 24. Rockville, MD: United States Pharmacopeial Convention; 2006. p. 113629th rev
26. Ye H. Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography-multiangle laser light scattering. *Anal Biochem.* 2006; 356:76–85. [PubMed: 16839514]
27. Brange, J.; Langkjar, L. Insulin Structure and Stability. In: Wang, YJ.; Peariman, R., editors. *Stability and Characterization of Protein and Peptide Drugs: Case Histories.* Plenum Press; New York: 1993. p. 315-350.
28. iTRAQ Reagents: Amine-specific Labeling Reagents for Multiplexed Relative and Absolute Protein Quantitation. Aug 8. 2007 www.appliedbiosystems.com

Figure 1. MALDI-TOF mass spectra of Lispro peptides following digestion with Glu-C under reducing conditions. Glu-C was added at a ratio of 1:10 (w/w, Glu-C/Lispro), and the mixture was incubated at 37°C for 4 h. The resultant peptides were desalted for MALDI-TOF analysis (A), or labeled with iTRAQ reagent for 1 h and then desalted for MALDI-TOF analysis (B). MS analysis was performed in the positive ion reflector mode utilizing an ABI 4700 proteomics analyser.

Figure 2.

MALDI-TOF mass spectrum and product ion mass spectra for mixture of four iTRAQ labeled insulins. Glu-C was added at a ratio of 1:10 (w/w, Glu-C/insulin) to each insulin sample, and the mixture was incubated at 37°C for 4 h. The resultant peptides for each insulin sample were labeled with a unique iTRAQ reagent for 1 h. Then four samples were mixed for desalting and MS analysis. MS analysis was performed in the positive ion reflector mode utilizing an ABI 4700 proteomics analyser. MS/MS spectra were acquired with a CID gas pressure of $\sim 1.5 \times 10^{-6}$ torr. (A) iTRAQ evaluation of the non-conserved B22–30 peptide and the conserved A18–21+B14–21 peptide. (B) iTRAQ evaluation of the non-

conserved A1–12+B1–13 peptide. The insets in dotted squares represent the reporter region of the MS/MS spectra of the selected precursor ions.

Figure 3. MS/MS spectra of precursor peptide B22–30 for insulin Lispro (A) and human insulin (B) labeled with iTRAQ reagents. The MS/MS analysis was performed with an ABI 4800 MALDI TOF/TOF Analyzer with the CID gas off. Most product ions are comparable between the two spectra, but several distinct y_b ions for insulin Lispro are evident: FYTK with m/z 684.38, KP with m/z 370.26, TK with m/z 374.25, YTK with m/z 537.32, and one y_2 ion PT with m/z 217.17.

Figure 4.

MS and MS/MS spectra of intact and desamido-A21 peptide A18–21+B14–21 for human insulin. The sample and control were labeled separately and mixed for MS analysis. The deamidated sample was labeled with iTRAQ tag 115, and the control was labeled with tag 116. The analysis was performed with an ABI 4800 MALDI TOF/TOF Analyzer. The MS/MS spectra in (A) were acquired with a CID gas pressure of $\sim 1.5 \times 10^{-6}$ torr for quantitative analysis. The MS/MS spectra in (B) and (C) were acquired with the CID gas off for identification analysis. (A) MS spectrum of the mixture and MS/MS spectra in the iTRAQ tag reporter region. (B) MS/MS spectrum of intact peptide. (C) MS/MS spectrum of desamido-A21 peptide.

Scheme I.

The general structure of iTRAQ reagents.

Species	A8	A10	A21	B28	B29	B30	B31	B32
Bovine (tag 114)	Ala	Val	Asn	Pro	Lys	Ala		
Lispro (tag 115)	Thr	Ile	Asn	Lys	Pro	Thr		
Porcine (tag 116)	Thr	Ile	Asn	Pro	Lys	Ala		
Human (tag 117)	Thr	Ile	Asn	Pro	Lys	Thr		
Lantus (tag 114)	Thr	Ile	Gly	Pro	Lys	Thr	Arg	Arg

Scheme II.
The primary structure of human insulin. Gray circles represent variations between insulins which are illustrated in the table. The tag listed next to each insulin represents the specific iTRAQ reagent.

Table I

Sequence Coverage and Peptide masses of Insulin Lispro (iTRAQ labeled)

Lispro sequence: Chain A GIVEQCCTSI CSLYQLENYC N Chain B FVNQHLCGSH LVEALYLVCGERGFFYTKPT				
Protease	Peptide	Observed <i>m/z</i>	Calculated <i>m/z</i>	Sequence
Trypsin	B1–22	2723.62	2723.62	FVNQHLCGSHLVEALYLVCGER
	B23–30	1248.87	1248.87	GFFYTKPT
Chymotrypsin	A15–19	810.58	810.50	QLENY
	A15–21	1073.66	1073.66	QLENYCN
	B17–25	1217.80	1217.80	LVCGERGFF
	B26–30	897.71	897.71	YTKPT
	B27–30	734.62	734.65	TKPT
Glu-C	A13–17	809.58	809.54	LYQLE
	A18–21	657.37	657.37	NYCN
	B1–13	1627.05	1626.91	FVNQHLCGSH LVE
	B14–21	1011.69	1011.62	ALYLVCGE
	B22–30	1405.00	1404.97	RGFFYTKPT
	A18–21+B14–21	1666.01	1665.97	NYCN + ALYLVCGE
	A1–12+B1–13	3008.54	3008.43	GIVEQCCTSICS + FVNQHLCGSH LVE

Table II

Quantitative Comparison of Humulin to Novolin

	NPH (Humulin/Novolin)			70/30 (Humulin/Novolin)			Regular (Humulin/Novolin)		
	Mean intensity ratio	SD	N	Mean intensity ratio	SD	N	Mean intensity ratio	SD	N
Individual experiment	1.204	0.024	24	1.034	0.021	24	1.126	0.009	24
	1.096	0.024	25	1.112	0.015	25	1.126	0.013	25
	1.211	0.080	12	1.027	0.036	12	1.133	0.043	12
	1.160	0.063	12	1.122	0.080	12	1.164	0.062	12
Combination of four experiments	1.16	0.04	73	1.07	0.04	73	1.13	0.03	73