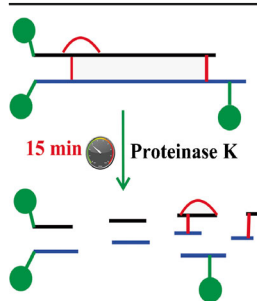


# Rapid Characterization of Insulin Modifications and Sequence Variations by Proteinase K Digestion and UHPLC-ESI-MS

Rong-Sheng Yang,  Weijuan Tang, Huaming Sheng, Fanyu Meng

Analytical Research and Development, Merck and Co., Inc., Rahway, NJ 07065, USA



**Abstract.** Discovery of novel insulin analogs as therapeutics has remained an active area of research. Compared with native human insulin, insulin analog molecules normally incorporate either covalent modifications or amino acid sequence variations. From the drug discovery and development perspective, methods for efficient and detailed characterization of these primary structural changes are very important. In this report, we demonstrate that proteinase K digestion coupled with UPLC-ESI-MS analysis provides a simple and rapid approach to characterize the modifications and sequence variations of insulin molecules. A commercially available proteinase K digestion kit was used to process recombinant human insulin (RHI), insulin glargine, and fluorescein isothiocyanate-labeled recombinant human insulin (FITC-RHI) samples. The LC-MS data clearly showed that RHI and insulin glargine samples can be differentiated, and the FITC modifications in all three amine sites of the RHI molecule are well characterized. The end-to-end experiment and data interpretation was achieved within 60 min. This approach is fast and simple, and can be easily implemented in early drug discovery laboratories to facilitate research on more advanced insulin therapeutics.

**Keywords:** Insulin modifications, Sequence variations, Proteinase K, HR-LCMS

Received: 14 July 2017/Revised: 6 October 2017/Accepted: 16 December 2017/Published Online: 29 January 2018

## Introduction

Insulin is an important therapeutic agent for the treatment of type I and type II diabetes mellitus [1, 2]. While natural insulins from animal or recombinant sources are still widely used today [1], significant advances have been made in the development of insulin analogs with better therapeutic performance (e.g., insulin glargine [3], insulin aspart [4], and numerous others [1]). Two major strategies have been used to produce insulin analogs: change of the primary sequence [3] and addition of covalent modifications [5, 6]. For insulin analogs with

covalent modifications, the lysine residue at the 29th position of B chain (B29Lys) is the most common site of modification [5]; however, the primary amine groups at N-terminus of insulin of the A and B chain (designated as N-A1Gly or N-B1Phe, respectively) are also possible sites for modifications. Therefore, it is critical to definitively confirm the modification sites, which is often determined as one critical quality attribute for modified insulins.

Mass spectrometry (MS) is a well suited approach to determine the amino acid sequence variations and post-translational modifications for proteins or peptides [7, 8]. There has been a recent industry trend of adopting an MS-based approach to gain a deeper understanding of the quality attributes of therapeutic proteins at the molecular level [9]. Although the data quality remains to be the most important consideration, a simple MS-based assay is highly desirable to enable the broad application of such approach, especially in early drug discovery settings. Ideally, no high-end MS instrumentation or extensive data interpretation should be required.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13361-017-1887-5>) contains supplementary material, which is available to authorized users.

Correspondence to: Rong-Sheng Yang; e-mail: rong-sheng.yang2@merck.com, Fanyu Meng; e-mail: fanyu\_meng@merck.com

Top-down mass spectrometry is an effective approach to determine the amino acid sequence variations and post-translational modifications [7, 8]. Generally, intact peptides (or proteins) are subjected to gas-phase fragmentation [10, 11], and the resulting tandem mass spectrometry (MS/MS) data is interrogated to derive the primary structural information. These MS/MS methods may be used to detect mass shifts that arise from modifications and sequence variations. In some special cases, even sequence variations that incur no mass shift may be characterized [12]; however, such methods may have factors that limit their widespread implementation in drug discovery settings. For example, labile protein modifications such as phosphorylation or prenylation [11, 13] may pose challenges to the commonly used collision-induced dissociation (CID) techniques. The utilization of electron, photon, and surface-induced fragmentation techniques such as ECD, ETD may alleviate such issues [11, 13], but specialized equipment is needed. In addition, MS/MS methods often generate complex data, which require expert interpretation to derive meaningful results.

Perhaps a more widely adopted approach to probe the primary structure of proteins and peptides is bottom-up mass spectrometry [14, 15]. In this approach, an enzyme (such as trypsin) can be used to digest proteins into smaller peptides before characterization by MS. In the case of insulin, due to limited cleavage sites by trypsin, other enzymes can generate more information-rich digestion products. One such enzyme is proteinase K [16, 17], which has shown broad specificity in cleaving peptide bonds for a variety of amino acid residues [18]. Simple LC-MS analysis of proteinase K digested insulin analogs may provide information useful for the unambiguous characterization of the modifications on the molecule. To validate this hypothesis, we studied fluorescein isothiocyanate-labeled recombinant human insulin (FITC-RHI, Figure 1) and demonstrated the utility of proteinase K in characterizing these modifications. Then, we used the same method for recombinant human insulin (RHI) and insulin glargine, and demonstrated that insulin sequence variations can also be characterized.

## Experimental

### *Chemicals and Reagents*

Recombinant human insulin, FITC-RHI, and insulin glargine samples were acquired from Sigma-Aldrich (St. Louis, MO, USA). The FITC-RHI sample was further purified by reverse-phase chromatography using a KROMASIL C8 column, 50 × 250 mm, 100 Å, 10 μm (AkzoNobel) with 0.05%TFA in water (solvent A) and 0.05%TFA in acetonitrile (solvent B) as mobile phases, and the gradient was 27%–45%B over 25 min. The collected fractions were lyophilized and then stored in a –20 °C freezer until use. The SMART Digest proteinase K was acquired from Thermo Fisher (Waltham, MA, USA). All solvents

were LC-MS grade and were purchased from Fisher Scientific (Waltham, MA, USA).

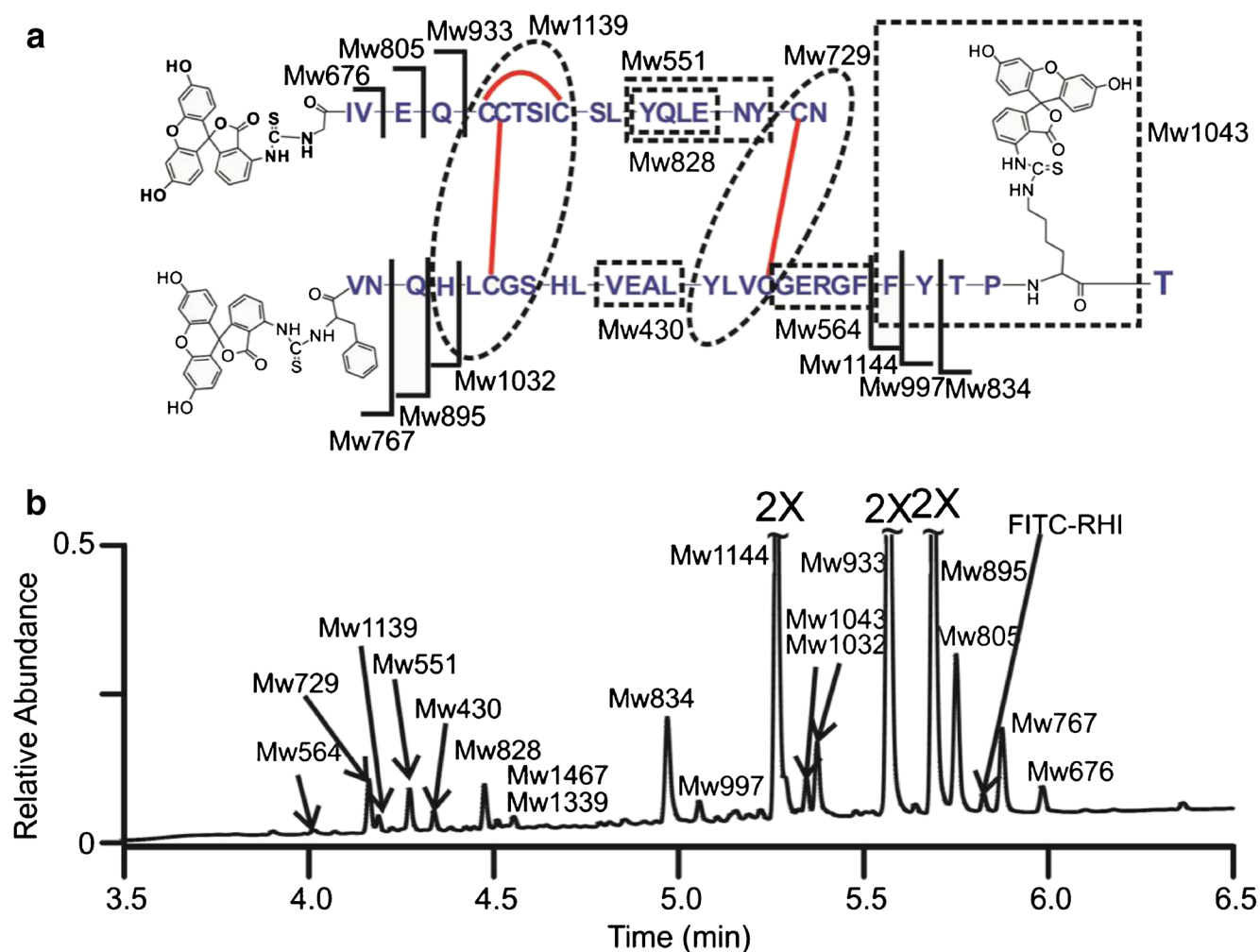
### *Sample Digestion Using Proteinase K*

The digestion protocol recommended by the vendor was followed. Briefly, 150 μL of SMART digestion buffer was added into each digestion tube, and the tube was placed in a thermostated-mixer set at 70 °C and 1400 rpm for 5 min. Fifty μL aliquots of FITC-RHI samples (5 mg/mL) were added to each digestion tube. The digestion mixture was removed after 5, 10, 15, 30, and 60 min and immediately acidified by adding 100 μL of methanol containing 1% formic acid. A 300 μL aliquot was then transferred to a filter vial (Thomson SINGLE StEP filter vials 0.45 μm PTFE, Thomson Instrument Company, Oceanside, CA, USA), and the filtered samples were either subjected to LC-MS analysis with no further processing or stored in –20 °C freezer until LC-MS analysis. For RHI and insulin glargine samples, 100 μL aliquots (2.5 mg/mL) were used with a digestion time of 15 min.

### *UPLC-ESI-MS Analysis A*

Waters Acquity UPLC system coupled to a Thermo Fisher LTQ-Orbitrap mass spectrometer was used to analyze proteinase K digested FITC-RHI, RHI, and insulin glargine samples. A CORTECS UPLC C18+ column, 2.1 × 100 mm (Waters, Milford, MA, USA) was used, with 0.05%TFA in water (solvent A) and 0.05%TFA in acetonitrile (solvent B) as mobile phases. Generally, 1–5 μL samples were loaded on the LC column, washed with 5% solvent B (2.5 min), followed by a linear gradient to 80% B in 5, allowing the peptide fragments to be eluted for UV (220–280 nm) and MS analyses. The flow rate was set at 0.3 mL/min, and the column temperature was set at 40 °C. The MS scan range employed was from  $m/z$  300 to 2000. Key MS parameters included the following: spray voltage 4.5 kV; capillary voltage 10 V; tube lens 80 V; microscan 1; the AGC target was set as balanced, and the maximum injection time was 500 ms. Data-dependent MS/MS experiments were also performed for FITC-RHI (15 min digestion), RHI, and insulin glargine samples. For each cycle, one Orbitrap full scan MS followed by two data-dependent MS/MS in LTQ (targeting the top two ions) was used. The isolation window was set as 2  $m/z$  and normalized collisional energy was set at 35. The dynamic exclusion was used so that after five MS/MS were acquired, that targeted mass would be included in the exclusion list for 30 s.

For MS data analysis, the raw data were manually processed, where the mass values of highly abundant peptides were calculated based on the observed monoisotopic  $m/z$  values and their corresponding charge numbers. The theoretical mass values were calculated using GPMW (ver. 10.0). For the time course study, the selected ion chromatogram was generated using Xcalibur with default settings with the exception that the mass option was set at 10 ppm.



**Figure 1.** (a) Structure of FITC-RHI, a recombinant human insulin molecule that was labeled with fluorescein isothiocyanates at the N-termini (A1, B10, and B29 sites). Various proteinase K digested products observed in the LC-MS analysis have been labeled. (b) UV total scan (220–280nm) for the LC-MS analysis of the proteinase K digested FITC-RHI sample at 15 min

## Results and Discussion

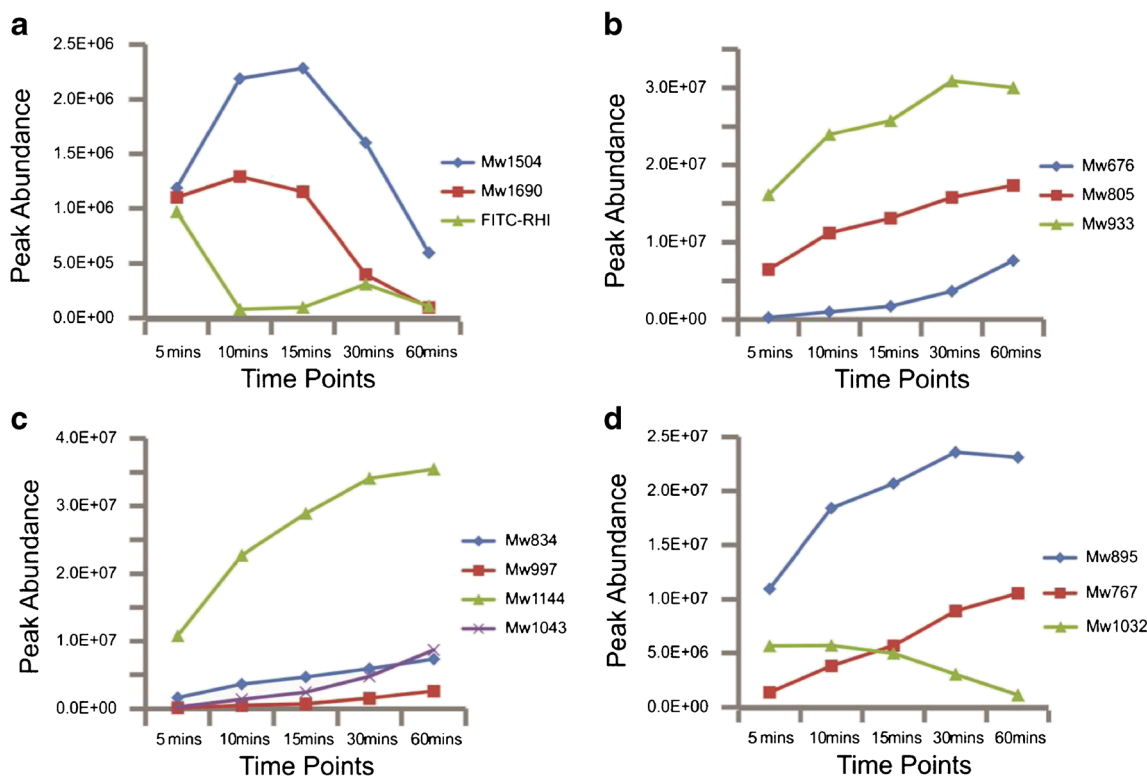
### Characterization of FITC-Modified RHI

As shown in Figure 1a, the FITC-RHI molecule contains three modified amino acid residues: N-terminus of the A1 glycine (N-A1Gly), N-terminus of the B1 phenylalanine (N-B1Phe), and B29 lysine (B29Lys). The total UV chromatogram (220–280 nm) of the 15 min digested FITC-RHI sample is shown in Figure 1b. Twenty-three major observed ions are listed in Supplementary Table S1 (in Supporting Information). The observed mass values were consistent with those of the expected digestion products, with less than 2 ppm mass error. Overall, our data showed that peptide mapping analysis utilizing proteinase K resulted in high sequence coverage for FITC-RHI (Figure 1a). In particular, a series of abundant ion species with observed nominal mass values of 834, 997, 1144, and 1043 Da (designated as Mw834, Mw997, Mw1144, and Mw1043, respectively) were eluted within 4.9–5.4 min. These mass values were consistent with the theoretical masses of digested peptide

fragments containing FITC modified B29Lys. Similarly, a group of three ions, Mw676, Mw805, and Mw933, were consistent with the peptides containing FITC-modified N-A1Gly, and a series of ions, Mw767, Mw895, and Mw1032, were consistent with the peptides containing FITC-modified N-B1Phe. The data indicated that the three modified amino acid residues in the FITC, N-A1Gly, N-B1Phe, and B29Ly could all be identified accurately and effectively.

### Kinetic Study of Proteinase K Towards Tri-FITC-RHI

The kinetic profile of proteinase K digestion of FITC-RHI was investigated in a 60-min time course study. Digested samples were quenched at 5, 10, 15, 30, and 60 min and subsequently analyzed. The total UV chromatograms (220–280 nm) are shown in Figure S1 (Supporting information). Selected groups of ions were monitored to gain insight into the digestion kinetics. We found that the abundance of intact FITC-RHI



**Figure 2.** Selected ion chromatograms showing the abundance for the FITC-RHI peptide and its digestion products from the proteinase K treatment at different digestion time points. **(a)** FITC-RHI peptide and Mw1504 and Mw1690 peptides, **(b)** Mw676, Mw805, and Mw933, **(c)** Mw1144, Mw834, Mw997, and Mw1043 peptides, and **(d)** Mw767, Mw895, and Mw1032 peptides. Assigned sequence of each peptide can be found in Supplementary Table S1 in Supporting Information

dropped rapidly within the first 10 min of digestion (Figure 2a), which indicated a fast rate of initial enzymatic digestion. The relative abundance of larger peptides such as Mw1504 and Mw1690 were detected within 5 min of incubation. These peptides reached their peak abundances in 10–15 min, and then decreased within the 60 min survey time (Figure 2a). The data suggested that these larger peptides were first formed during proteinase K digestion, and were further digested into smaller ones over time.

The ion series corresponding to major abundant peptides containing FITC modifications were monitored. The abundance of Mw676, Mw805, and Mw933 were plotted against the incubation time (Figure 2b). The peptide Mw933 corresponding to G\*IVEQ (\* indicates the FITC modification) grew rapidly, and remained abundant after 30 min. Similar behavior was observed with Mw805. Interestingly, the abundance of the Mw676 ion species continued to grow in the 60 min surveyed. Similar behavior was observed for the Mw1144, Mw834, Mw997, and Mw1043 ion species (Figure 2c). The Mw767, Mw895, and Mw1032 species are an interesting series of ions that contain FITC labeled N-B1Phe. We found that Mw767 and Mw895 ions continue to grow within the 60 min. However, the Mw1032 ion seemed to plateau in the 5–10 min time frame, and then showed decreased abundance from 15 to 60 min. This is likely is an indication of further digestion of the Mw1032 into smaller peptides, including Mw767 and Mw895.

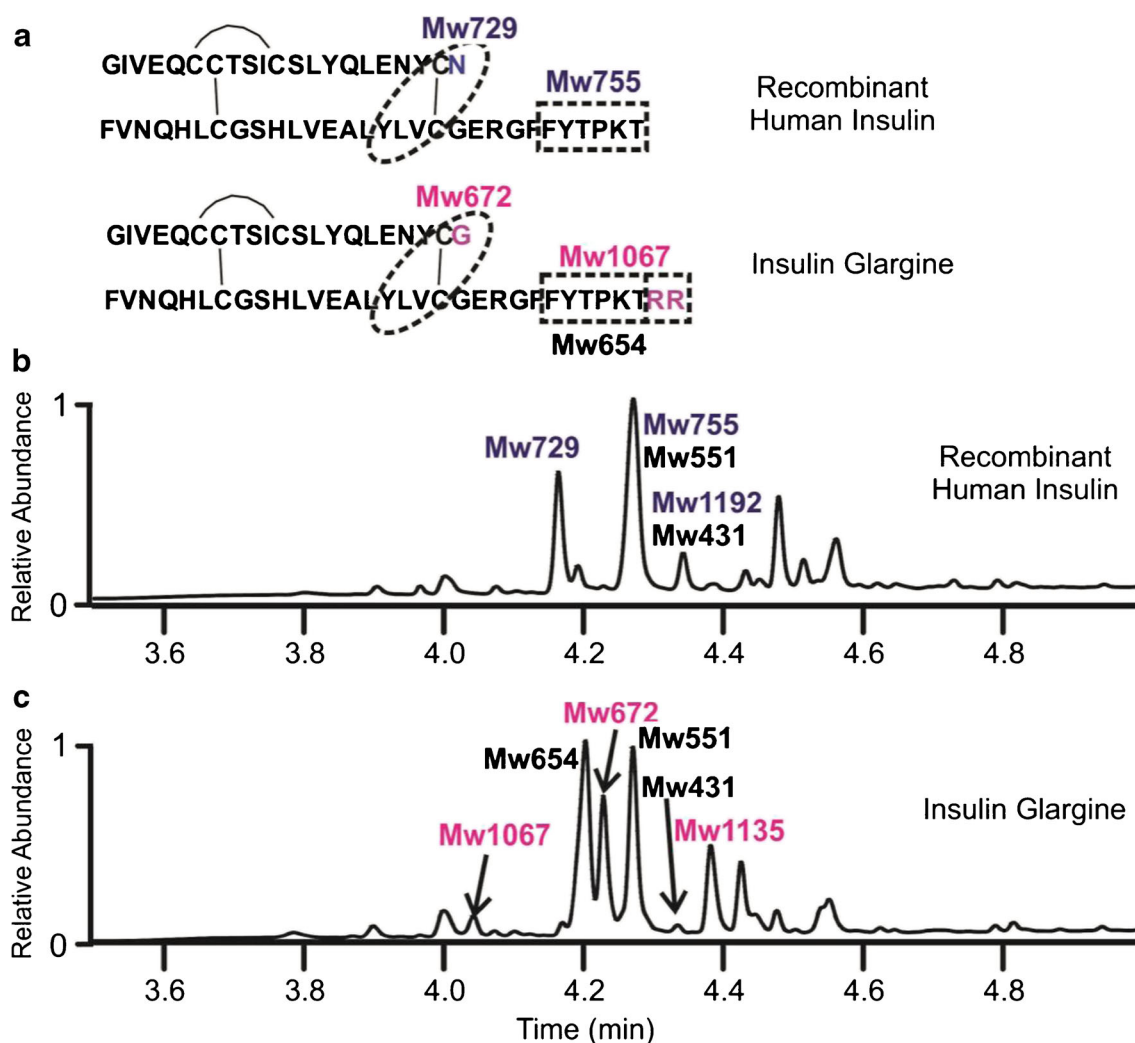
In summary, we found that digestion of the FITC-RHI using proteinase K is a fast process. Some of the desired digestion products began to appear within 1 min (data not shown), and the majority of intact FITC-RHI was consumed within 10 min of digestion. Within the time period surveyed (15–60 min), the major digestion products with FITC modified amino acid residues, such as Mw933, Mw895, or Mw1144, remained as abundant ions. Consequently, these digestion products can be used to unambiguously characterize the modifications of N-A1Gly, N-B1Phe, and B29Lys sites on recombinant human insulin.

### Insulin Sequence Variation Characterization

In addition to post-translational modifications, the workflow described in this report is also useful in detecting sequence variations of human insulin and insulin analogs. Digestion of RHI and insulin glargine samples was performed to demonstrate this capability. As shown, the primary sequence for insulin glargine is slightly different from that of RHI. The A21Asn of RHI is replaced with an A21Gly residue, and the insulin glargine contains two extra Arg residues (B31Arg and B32Arg) at the C-terminus of B chain (Figure 3a).

The total UV chromatograms of RHI and insulin glargine are shown in Figure 3b. Most abundant ion species observed in the proteinase K digestion mixture for RHI and insulin glargine were similar, but a few apparent differences were observed





**Figure 3.** (a) The sequences of RHI and insulin glargine molecules with selected major digestion products. The UV total scan chromatogram for the LC-MS analyses of Proteinase K digested: (b) RHI sample and (c) insulin glargine

(Supplementary Table S2, in Supporting Information). An abundant Mw672 ion was observed in the insulin glargine digest (Figure 3c), but not in the RHI digest (Figure 3b). This ion could be assigned as either YLVC(CG) or LVC(YCG), which contained the A21Gly residue and a disulfide bond between the two cysteine residues. The corresponding Mw729 peak was observed as an abundant ion in the RHI digest, which can be assigned as YLVC(CN) or LVC(YCN). These data showed that proteinase K digestion produced peptides that contain unique A21Asn (for RHI) or A21Gly (for insulin glargine) residues, which can be readily identified by LC-MS.

Another interesting pair of peptides are Mw1135 (insulin glargine digest) and Mw1192 (RHI digest) peptides. The Mw1192 could be assigned as either YLVC(NYCN)GE or YLVC(ENYCN)G, and based on MS/MS (data not shown), YLVC(NYCN)GE was the most abundant form. In either case, the peptide contains the unique A21Asn residual for RHI, and consistent with LC-MS data, this ion is only observed in the RHI digest, but not in the insulin glargine digest. The

corresponding peptide for glargine would be the Mw1135 peptide, which could be assigned as peptide YLVC(NYCG)GE (more likely based on MS/MS) or YLVC(ENYCG)G. Both of these peptides contain the A21Gly that is unique to insulin glargine, and indeed, the Mw1135 ion was observed with high signal abundance for insulin glargine. Surprisingly, we also observed an Mw1135 ion in the RHI digest with the exact same mass, although the MS signal was much lower (~100-fold lower based on selected ion chromatography). Further inspection of the protein sequences suggested that this was likely the isobaric peptide YLVC(ENYCN), which is shared by both RHI and insulin glargine. The MS/MS data of the Mw1135 ions in both RHI and insulin glargine samples confirmed our assignments (data not shown). It is likely that the YLVC(ENYCN) also existed in the insulin glargine digest as a minor component, which overlaps with the major YLVC(NYCG)GE peptide.

Detection of the Mw1067 peptide (FYTPKTRR) uniquely in insulin glargine digests showed that the B31Arg and B32Arg residues can be characterized using this approach. In addition,

we also found an interesting pair of Mw654 peptide (FYTPK) and Mw755 peptide (FYTPKT). Both of these peptide sequences are shared by RHI and insulin glargine molecules. For Mw654 peptide, its abundance was much higher in insulin glargine than RHI. In contrast, the Mw755 showed high abundance in RHI, but was not detected in insulin glargine. The pair of Mw1014 (RGFFYPTK) and Mw1115 (RGFFYPTKT) peptides showed similar trends. A possible explanation is that the amide bond between the lysine and threonine residues is not a preferred proteolytic site for proteinase K as in the case of RHI, whereas the addition of B31Arg and B32Arg made that amide bond more likely to be cleaved as in the case of insulin glargine.

## Conclusion

In this report, we described the use of proteinase K digestion and LC-MS as a rapid and simple approach for the characterization of covalent modifications and sequence variations for insulin and related molecules. With simple sample preparation, proteins can be efficiently digested within 15 min at elevated temperatures, and LC-MS analysis of the digest mixture provided sufficient data to pinpoint modifications or sequence variation sites on insulin molecules. The end-to-end experiment (including data analysis) can be completed in less than 1 h, rendering the method amenable to labs that are equipped with simple mass spectrometers. As such, we envision that this fast and simple method can be very useful in early drug discovery labs, which will greatly facilitate the efforts to search for more advanced insulin therapeutics.

## Acknowledgments

The authors specially acknowledge Pei Huo for the FITC-RHI purification and Gary Martin for his helpful comments on this manuscript. The authors thank Bing Mao, R. Thomas Williamson, and Caroline McGregor for their support and advice.

## References

- Sheldon, B., Russell-Jones, D., Wright, J.: Insulin analogues: an example of applied medical science. *Diabetes Obes. Metab.* **11**, 5–19 (2009)
- Valla, V.: Therapeutics of diabetes mellitus: Focus on insulin analogues and insulin pumps. *Exp. Diabetes Res.* 2010, Article ID 178372 (2010)
- Elrishi, M.A., Jarvis, J., Khunti, K., Davies, M.J.: Insulin glargine and its role in glycaemic management of Type 2 diabetes. *Expert Opin. Drug Metab. Toxicol.* **4**, 1099–1110 (2008)
- Gammeltoft, S., Hansen, B.F., Dideriksen, L., Lindholm, A., Schäffer, L., Trüb, T., Dayan, A., Kurtzhals, P.: Insulin aspart: a novel rapid-acting human insulin analogue. *Expert Opin. Investig. Drugs.* **8**, 1431–1442 (1999)
- Chou, D.H.-C., Webber, M.J., Tang, B.C., Lin, A.B., Thapa, L.S., Deng, D., Truong, J.V., Cortinas, A.B., Langer, R., Anderson, D.G.: Glucose-responsive insulin activity by covalent modification with aliphatic phenylboronic acid conjugates. *Proc. Natl. Acad. Sci. USA.* **112**, 2401–2406 (2015)
- Webber, M.J., Anderson, D.G., Langer, R.: Engineering synthetically modified insulin for glucose-responsive diabetes therapy. *Expert Rev. Endocrinol. Metab.* **10**, 483–489 (2015)
- Kelleher, N.L., Thomas, P.M., Ntai, I., Compton, P.D., LeDuc, R.D.: Deep and quantitative top-down proteomics in clinical and translational research. *Expert Rev. Proteom.* **11**, 649–651 (2014)
- Malaker, S.A., Ferracane, M.J., Depontieu, F.R., Zarling, A.L., Shabanowitz, J., Bai, D.L., Topalian, S.L., Engelhard, V.H., Hunt, D.F.: Identification and characterization of complex glycosylated peptides presented by the MHC Class II processing pathway in melanoma. *J. Proteome Res.* **16**, 228–237 (2017)
- Rogers, R.S., Nightlinger, N.S., Livingston, B., Campbell, P., Bailey, R., Balland, A.: Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics. *mAbs.* **7**, 881–890 (2015)
- Brodbeck, J.S.: Ion activation methods for peptides and proteins. *Anal. Chem.* **88**, 30–51 (2016)
- Coon, J.J.: Collisions or electrons? Protein sequence analysis in the 21st century. *Anal. Chem.* **81**, 3208–3215 (2009)
- Yu, X., Zhong, W.: Differentiation of norvaline and valine in peptides by hot electron capture dissociation. *Anal. Chem.* **88**, 5914–5919 (2016)
- Bhawal, R.P., Shahinuzzaman, A.D., Chowdhury, S.M.: Gas-phase fragmentation behavior of oxidized prenyl peptides by CID and ETD tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 704–707 (2017)
- Furlong, M.T., Ouyang, Z., Wu, S., Tamura, J., Olah, T., Tymiak, A., Jemal, M.: A universal surrogate peptide to enable LC-MS/MS bioanalysis of a diversity of human monoclonal antibody and human Fc-fusion protein drug candidates in preclinical animal studies. *Biomed. Chromatogr.* **26**, 1024–1032 (2012)
- Zhang, Q., Spellman, D.S., Song, Y., Choi, B., Hatcher, N.G., Tomazela, D., Beaumont, M., Tabrizifard, M., Prabhavalkar, D., Seghezzi, W., Harrelson, J., Bateman, K.P.: Generic automated method for liquid chromatography-multiple reaction monitoring mass spectrometry based monoclonal antibody quantitation for preclinical pharmacokinetic studies. *Anal. Chem.* **86**, 8776–8784 (2014)
- Makepeace, K.A., Serpa, J.J., Petrotchenko, E.V., Borchers, C.H.: Comprehensive identification of disulfide bonds using non-specific proteinase K digestion and CID-cleavable crosslinking analysis methodology for Orbitrap LC/ESI-MS/MS data. *Methods.* **89**, 74–78 (2015)
- Fränzel, B., Fischer, F., Steegborn, C., Wolters, D.A.: Proteinase K improves quantitative acylation studies. *Proteomics.* **15**, 44–47 (2015)
- Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H.D., Lang, H.: Proteinase K from *Tritirachium album Limber.* *Eur. J. Biochem.* **47**, 91–97 (1974)