Proteomics Approach for Identification of Hemoglobin Variants and Post-Translational Modifications

Hua Huang[1], Mark E. McComb[1], David H. Perlman[1,2], Bogdan A. Budnik[1,2], Parminder Kaur[1,2], Timothy P. Skelton[3], Martin H. Steinberg[3], David H. K. Chui[3], Peter B. O'Connor[1,2], Catherine E. Costello[1,2]

1. Cardiovascular Proteomics Center; 2. Mass Spectrometry Resource, 3. Boston Comprehensive Sickle Cell Center, Boston University School of Medicine, Boston, MA 02118.

Introduction: Among the more than 1200 recognized hemoglobin variants, many mutations have been found to be related to human diseases. Additionally, post-translational modifications of hemoglobin, to a large extent, may also contribute to biological function and act as a disease marker or contribute to disease pathology. In this study, a MS-based technology platform using MALDI-TOF MS, VC-MALDI-FT MS, ESI-qQq-FT MS/MS and LC-MS/MS have been explored to analyze a large number of human blood samples with diverse variants and post-translational modifications in the hemoglobin chains. While most variants have been identified consistent to their gene-based DNA sequence, PTMs and some variant types have been detected and located only by the current integrated methodology.

Methods: Whole blood was diluted and cleaned up to remove cellular debris and salts. Trypsin digestion and AspN digestion of the intact globin chains were performed for peptide mass mapping and MS/MS. Intact hemoglobin chains were analyzed and top-down sequenced via ESI-qQq-FT MS/MS (quadrupole-FT hybrid constructed in-house), and the peptide mapping was performed on the same instrument. Additionally, digests were analyzed by MALDI-TOF MS (Bruker Reflex IV), MALDI-FT MS/MS (vibrational-cooling MALDI-FT constructed in-house), and online LC-MS/MS (QTOF-API-US, Waters Corporation). Data were processed and searched against SwissProt and custom programmed Hemoglobin/PTM databases using commercially available software (ProteinLynx Global Server 2.1, Waters) and software written in-house.

Results: Minimal requirements for purification, derivatization or separation of the blood samples considerably simplified the sample preparation and reduced artifacts associated with sample purification which may perturb PTMs. MALDI-TOF MS was carried out as first-pass for peptide mapping. More accurate mass mapping was achieved by VC-MALDI-FT MS. Nanospray ESI-qQq-FT MS was applied to the measurement of the intact hemoglobin chains. Calculation of the charge state and identification of the *m/z* of the monoisotopic mass was performed using software written in-house and yielded accurate mass measurement within a few ppm. Mutations and PTMs were observed at the intact protein level. Localization of the mutation(s) and PTMs was achieved using a combination of top-down sequencing, peptide mapping and MS/MS peptide sequencing. For online LC-MS/MS of the hemoglobin digests, data analysis was fully automated. By using an iterative approach to peptide sequencing, pre-programmed hemoglobin database and pre-programmed PTM database, data analysis time was dramatically reduced, and more accurate assignments were obtained. For example: an alpha chain C-terminal truncation from a clinically relevant sample was unambiguously characterized by ESI-qQq-FT MS for intact protein mass, top-down sequence analysis, AspN peptide mapping, isolated single peptide MS/MS sequencing, and LC-MS and MS/MS sequencing (Fig. 1). A sickle beta chain identification was achieved as well via the mass of the intact protein, tryptic peptide mapping, and LC-MS/MS sequencing (data not shown). Over 75 clinically interesting samples, including diverse hemoglobin variants, have been identified using this MS-based proteomics approach. The results were consistent to their DNA sequencing results, and for some samples, showed new results that DNA analysis could not address. Additionally, post-translational modifications have also been revealed by this method. One example is a methionine oxidation identified for an alpha peptide exclusively by LC-MS/MS though false matching proposed by MALDI-TOF-MS data and automated databank search (Fig. 2).

Conclusions: This MS-based proteomics technology platform allows reliable and robust detection and localization of variants and PTMs simultaneously. This method demonstrates the potential for clinical use. The strategies for the application of different kinds of MS and data analysis has been established and validated with clinically relevant samples.

Acknowledgement: This project was funded by NIH grant P41-RR10888, S10-RR15942 and NHLBI contract N01-HV-28178. We are grateful for technological assistance from Waters Corporation.

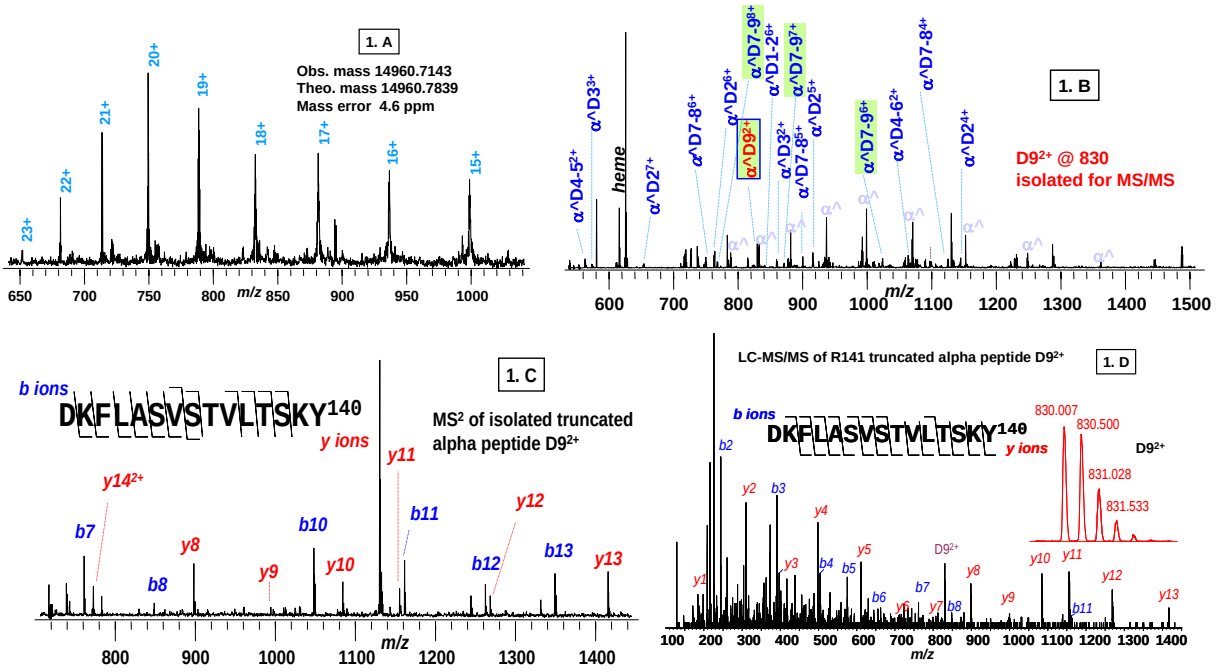

Fig. 1 MS approach identified the C-terminal R141 truncation of hemoglobin alpha chain in a clinical blood sample. A: observed intact protein mass matching R141 truncated alpha chain by ESI-FT-MS. B: AspN peptide mapping by ESI-FT-MS detected peptides generated from the R141 truncated alpha chain (D9 and D7-9). $D9^{2+}$ @ 830 was isolated for MS/MS; C: SORI CAD $MS^2$of the Q1 isolated $D9^{2+}$ by ESI-FT-MS found the sequence matched the truncated peptide; D: LC-MS/MS sequencing confirmed the R141 truncation.

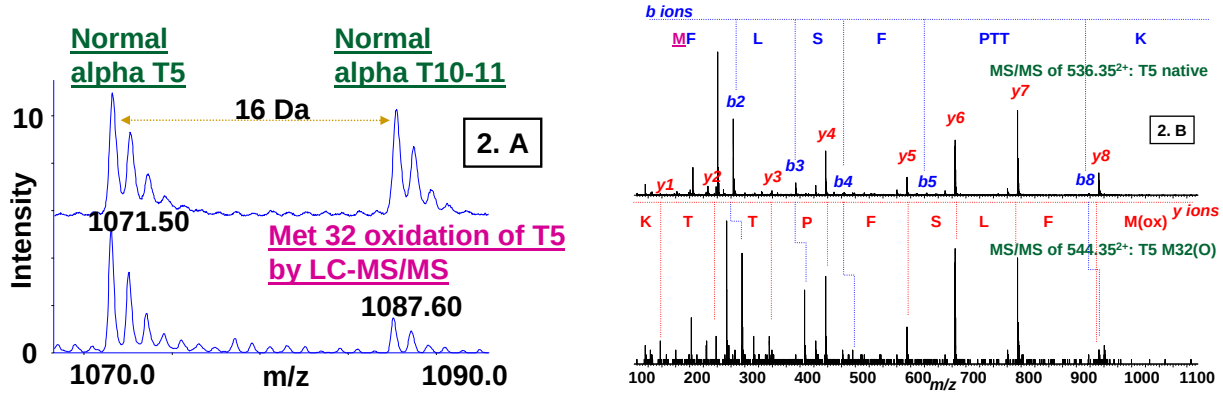

Fig.2 Identification of Met oxidation by LC-MS/MS. A: inclusive ID of normal alpha T10-11 and Met 32 oxidized alpha peptide T5; B: LC-MC/MS confirmed the oxidized form of T5.