Spectral Library Search Based Method for the Confident Identification of Genetically Variant Peptides in Human Hair

Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899 USA

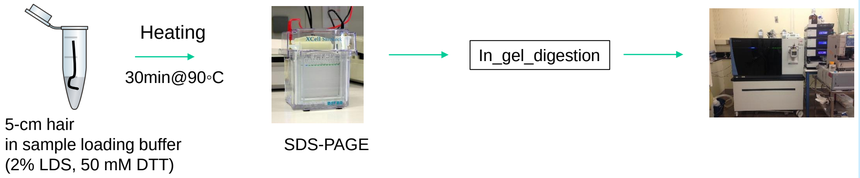
peptide.nist.gov

Introduction

Genetically variant peptides (GVPs) derived from human hair proteins have been reported to have the potential to serve as alternative evidence for human identification. We reported a sensitive direct extraction method involving one-step protein extraction from hair shaft that was found to provide reproducible results. Together with the construction of a hair specific peptide mass-spectral library, including previously reported GVPs, it provides quick and accurate means of identifying the peptides in human hair protein digests. In this study, we also describe a mass-spectral library-based procedure for GVP panel analysis and finding experimentally introduced artifactual modifications.

Methods

- ❑ Human hair samples were obtained commercially.
- ❑ Protein extraction protocols were established and coupled with digestions and desalting for peptide extraction and clean-up.
  - ▪ Direct Extraction Method
    (J Forensic Sci. 2020 Mar 65(2):406-420. PMID: 31670846 DOI: 10.1111/1556-4029.14229)



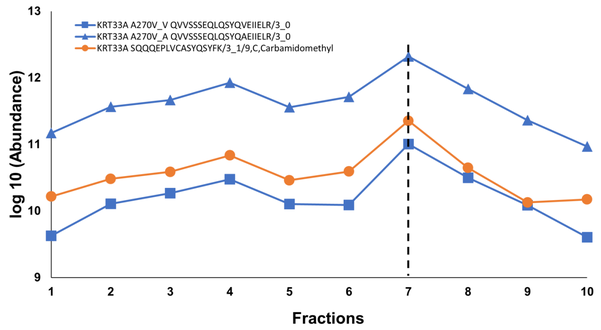
5-cm hair in sample loading buffer (2% LDS, 50 mM DTT)

SDS-PAGE

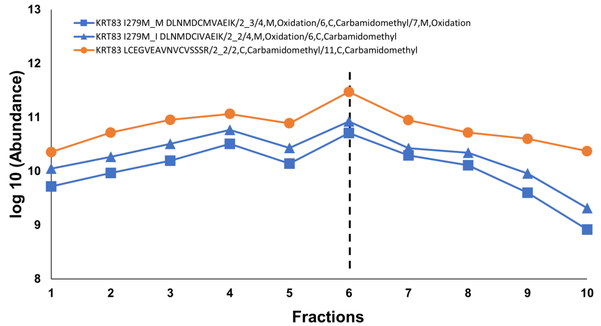
  - ▪ NaOH-based SDS repeated extraction method (NaOH+SDS method)
    (PMID:27741315 DOI:10.1371/journal.pone.0164993)
  - ▪ ProteaseMax-based Cleavable Surfactant method (Cleavable Surfactant method)
    (PMID:27603779 DOI:10.1371/journal.pone.0160653)
- ❑ Using the raw data files generated from the LC-MS/MS analyses by a mass spectrometer, a hair specific peptide mass-spectral library was constructed. This 'hair' library contains 6280 spectra (6280 peptide ions of 4343 distinct peptides, HCD=30eV), and among these – a total of 3754 spectra (3754 peptide ions of 2240 distinct peptides, HCD=30eV) arose from hair keratins or keratin associated proteins. It has a sequence coverage of hair cuticular keratins of about 70% and includes 14 previously reported GVPs.
- ❑ To achieve the best performance, we concatenated the 'hair' library with NIST 'main' library for library searching.

Result A. Type I & II Hair Cuticular Keratin-GVPs

(A) Type I Cuticular Keratin KRT33A



(B) Type II Cuticular Keratin KRT83



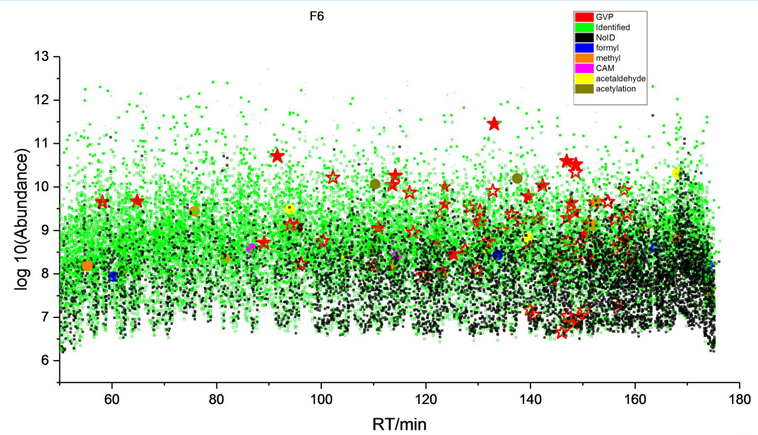
The range of the intensities of example peptide ions across all ten fractions from the Direct method in type I (A) and type II (B) hair cuticular keratins. For each type of hair cuticular keratin, a high-abundance GVP and its regular non-variant form, as well as another unique peptide ion were shown as examples of high-confidence GVP identification from the gel-based Direct method. x-axis: Fractions, y-axis: Abundance in log 10 scale.

Result B. GVP Panel Analyses in Three Methods

ONE 5-CM HAIR, ASIAN DONOR	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP 10-8	TGM3
	R1738Q_Q	V128L_L	A82V_V	S222Y_Y	A270V_V	V279L_L	P443A_A	S36P_P	S13R_R	T458M_M	G362S_S	I279M_M	H26R_R	T13K_K
D_LG_F1_TO_F10_R1#	X		X		X			X	X		X	X	X	X
D_LG_F1_TO_F10_R2	X		X		X	X		X	X		X	X	X	X
D_LG_F1_TO_F10_R3	X		X	X	X	X		X	X		X	X	X	X
D_LG_COMBINED_R1			X		X	X		X	X			X		X
D_LG_COMBINED_R2			X		X	X		X	X			X		X
D_LG_COMBINED_R3			X		X			X	X			X		X
D_SG_COMBINED_R1			X		X			X	X			X		X
D_SG_COMBINED_R2			X		X			X	X			X		X
D_SG_COMBINED_R3			X		X		X	X	X			X		X
NS_LG_F1_TO_F10_R1	X	X	X		X		X	X	X		X	X	X	X
NS_LG_F1_TO_F10_R2	X	X	X	X	X			X	X		X	X		X
NS_LG_F1_TO_F10_R3	X		X	X	X		X	X	X		X	X	X	X
CS_R1			X				X			X		X		X
CS_R2			X				X	X		X		X		X
CS_R3			X				X			X				X

GVP panel analyses by the Direct method with all 10 fractions from a long-gel (30 min run at 200 V) which have been individually processed by LC-MS/MS and then summarized the results in one row are labeled as 'D_LG_F1_TO_F10'; GVP panel analyses with combined fractions processed as a mixture from a long-gel run by the Direct method are labeled as 'D_LG_COMBINED'; with combined fractions from a short-gel run (10 min run at 200 V) are labeled as 'D_SG_COMBINED'; GVP panel analyses by NaOH+SDS method with all 10 fractions from a long-gel run individually processed and then summarized are labeled as 'NS_LG_F1_TO_F10'; GVP panel analyses by the Cleavable Surfactant method are labeled as 'CS'.R1, R2, and R3 are three experiment repeats.

Result C. Classification of GVP and Other Ions by the Hybrid Search



IonPlot shows the classification of GVP, identified, and not identified (NoID) ions, as well as several artifactual modifications: formylation (formyl), methylation (methyl), alkylation (CAM), acetaldehyde, and acetylation that present in fraction 6 (F6), a representative gel fraction from a protein gel separating proteins derived from a 5 cm-long hair shaft of this Asian donor by the Direct method. Solid: identified by regular library search; Hollowed: identified by hybrid library search. x-axis: Retention Time (RT) in minute (min), y-axis: Abundance in log 10 scale.

Summary & Conclusions

- ❑ A gel-based method of analysis reveals a wide distribution of molecular weights of proteins yielding keratin-based GVPs.
- ❑ The finding that different sample preparation methods can identify different GVPs suggests the inadequacy of any current method of finding all potentially identifiable GVPs in a hair sample.
- ❑ A future version of a comprehensive hair peptide mass-spectral library shall include all identifiable peptides with major and minor experimentally introduced artifactual modifications.
- ❑ Using hair-derived peptide library to make peptide spectral assignments provides a sensitive and reliable means of peptide identification. Ultimately, the hair-derived peptide spectral library can contain spectra of all known GVPs reported in the literature and online open resources (UniProtKB, dbSNP).