Comparison of Sample Preparation Methods for Genetically Variant Peptide Identification in Human Hair

Zheng Zhang, William E. Wallace, Guanghui Wang, Meghan C. Burke Harris, Yi Liu, Sergey L. Sheetlin, Stephen E. Stein
zheng.zhang@nist.gov

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

peptide.nist.gov

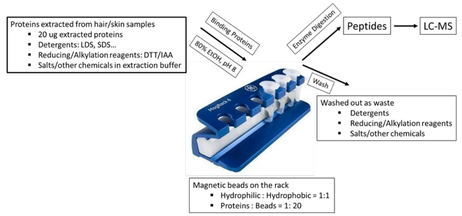
Introduction

Genetically variant peptides (GVPs) derived from human hair proteins have been proposed as alternative evidence for human identification in forensics.[1,2] The NIST-developed Direct method[3] efficiently extracts hair proteins from a 5 cm-long hair strand in 30 minutes. A new SP3 method[4] is described using fast nonselective binding of proteins on Sera-Mag Carboxylate-Modified magnetic beads. Proteins are immobilized on beads, easily separated from chemical waste in extraction buffer, making SP3 a convenient in-solution-digestion method. We optimized it for processing hair samples. We tested the efficiency and reproducibility of Direct method combined with SP3 and compared it to other in-solution-digestion methods. We also developed a pipeline that contains novel tools like protein spectrum and XIC analysis program. We applied these tools for examining experimental reproducibility, accurate peptide/GVP validation, and quantification.

Methods

- ❑ Human hair samples were obtained commercially
- ❑ Individual 5 cm-long hair strand from the same donor was used as the starting material for each experiment
- ❑ The SP3 method (SP3_A and SP3_B were two experimental repeats):

Proteins were extracted by the Direct method[3], reduced with DTT, and alkylated with IAA, followed by purification and digestion using the optimized SP3 magnetic-bead method[4]



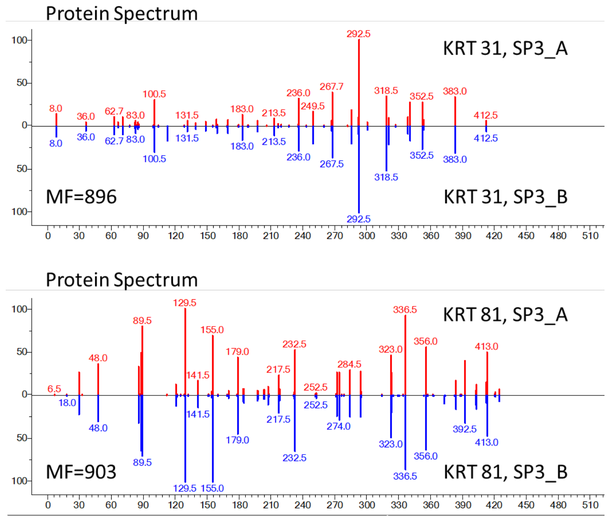
- ❑ Other tested methods include:
  1) RapiGest (RG) based method:
     i. RG_A: 1% of RapiGest with DTT at 60°C
     ii. RG_B: 1% of RapiGest at 60°C, followed by TCEP at RT
     iii. RG_C: 1% of RapiGest at 100°C, followed by TCEP at RT
  2) Guanidine (G) based method:
     i. G_A: 6M of Guanidine with DTT at 40°C
     ii. G_B: 6M of Guanidine with DTT at 95°C
- ❑ Proteins were digested by trypsin and lys-C; Peptides were cleaned up with MonoSpin C18 columns for LC-MS/MS on an orbitrap mass spectrometer (Fusion Lumos)
- ❑ We applied database searching in MSFragger (v3.7) and Sequest (Proteome Discoverer 2.4) with a newly-expanded FASTA file by adding mutations reported by BioMuta:
  1) 54 keratins (KRTs) with 7,554 proposed mutations
  2) 92 keratin-associated proteins (KRTAPs) with 4,534 proposed mutations
- ❑ We used NIST-developed pipeline that contains novel protein spectrum and XIC analysis program for checking the reproducibility, accurate peptide/GVP validation, and quantification.

Result A. Abundance (Peak Area) and Numbers of PSMs of Identified GVPs in Different In-Solution Digestion Methods

Method	DNVELENLIR 2+ KRT31_A82V		QVVSSSEQLQSYQVEIIELRR 3+ KRT31_A270V		VSAMYSSSPCK 2+_1(10,C,CAM) KRT35_S36P		VSAMYSSSPCKLPSLSPVAR 3+_2(4,M,Oxi)(10,C,CAM) KRT35_S36P		DLNMDCMVAEIK 2+_1(6,C,CAM) KRT83_I279M		TYVIAASTMSVCSSDVGR 2+_1(12,C,CAM) KRTAP10-8_H26R	
	Abundance	PSM	Abundance	PSM	Abundance	PSM	Abundance	PSM	Abundance	PSM	Abundance	PSM
SP3_A	4.87E+09	8	3.42E+09	30	1.10E+08	4	-	-	7.39E+08	7	-	-
SP3_B	6.49E+09	11	3.50E+09	33	1.13E+08	2	-	-	6.91E+08	5	-	-
RG_A	8.77E+07	7	-	-	-	-	3.86E+08	7	-	-	1.21E+07	1
RG_B	5.28E+07	3	-	-	-	-	-	-	-	-	-	-
RG_C	9.42E+07	1	-	-	-	-	-	-	-	-	-	-
G_A	3.34E+07	3	-	-	-	-	2.01E+07	2	-	-	-	-
G_B	3.85E+07	1	-	-	-	-	4.28E+07	2	-	-	-	-

Abundance (peak area) and numbers of peptide spectral matches (PSMs) of identified GVPs in different in-solution digestion methods on hair samples from the same donor. The mutated sites were highlighted in red.

Result B. Protein Spectra Showing Reproducibility of the SP3 Method



Protein spectra were applied to show the reproducibility of the SP3 method. Each vertical line represents an identified peptide. X-axis is the position in a protein; Y-axis is the abundance. SP3_A and SP3_B were two experimental repeats. KRT 31 was used as an example protein for the type I cuticular KRTs; KRT 81 was used for the type II cuticular KRTs. MF about 900 in both cases indicates high reproducibility.

Result C. Method Comparison of Hair Proteins, KRTs, and KRTAPs

Method	Hair Proteins			Hair KRTs & KRTAPs			Cuticular KRTs			KRTAPs		
	Proteins	Peptides	PSMs	Proteins	Peptides	PSMs	Proteins	Peptides	PSMs	Proteins	Peptides	PSMs
SP3_A	185	1,847	12,838	52	1,660	12,566	14	1,068	10,751	28	520	1,633
SP3_B	153	1,686	12,510	50	1,553	12,301	13	1,008	10,607	26	477	1,552
RG_A	150	768	5,390	45	541	4,592	13	426	4,196	24	90	321
RG_B	119	516	3,017	26	366	2,679	8	78	361	1	1	1
RG_C	225	876	4,503	30	385	2,901	12	187	1,606	2	2	6
G_A	49	403	1,371	34	387	1,344	9	242	903	22	140	424
G_B	58	689	2,237	37	663	2,191	9	413	1,510	25	243	665

Method comparison of hair proteins, hair KRTs (cuticular and cytoskeleton keratins), KRTAPs (keratin-associated proteins) in different in-solution digestion studies on hair samples from the same donor. Individual 5 cm-long hair strand was used as the starting material in each experiment.

References

(1) Parker G.J.; Leppert T.; Anex D.S.; Hilmer J.K.; Matsunami N.; Baird L.; Stevens J.; Parsawar K.; Durbin-Johnson B.P.; Rocke D.M.; Nelson C.; Fairbanks D.J.; Wilson A.S.; Rice R.H.; Woodward S.R.; Bothner B.; Hart B.R.; Leppert M. Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome. PLoS One. 2016, 11(9), e0160653.
(2) Wu J.; Liu J.; Ji A.; Ding D.; Wang G.; Liu Y.; Zhang L.; Feng L.; Ye J. Deep coverage proteome analysis of hair shaft for forensic individual identification. Forensic Sci Int Genet. 2022, 60, 102742.
(3) Zhang Z.; Burke M.C.; Wallace W.E.; Sheetlin S.L.; Mirokhin Y.A.; Tchekhovskoi D.V.; Stein S.E. Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin. J Forensic Sci 2020, 65(2), 406-420.
(4) Hughes C.S.; Moggridge S.; Müller T.; Sorensen P.H.; Morin G.B.; Krijgsveld J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc. 2019, 14(1), 68-85.

Summary & Conclusions

- ❑ For GVPs identified in the tested different methods, abundances (peak area) and numbers of PSMs were significantly higher in the SP3 method
- ❑ Reproducibility of the SP3 method was in general substantially better possibly due to the fact that the majority of peptides, as well as GVPs, were identified with significantly higher abundance using this method
- ❑ The SP3 method found many more peptides/PSMs derived from hair proteome (main components are hair cuticular/cytoskeleton keratins and keratin-associated proteins)
- ❑ Putting these findings together, the SP3 method provides a straightforward, efficient, and reliable in-solution sample preparation method for processing hair proteins after successful extraction from trace-amount hair by the Direct method[3]
- ❑ When combined with NIST-developed pipeline which contains novel protein spectrum and XIC analysis program, it can serve as a robust tool for demonstrating experimental reproducibility, accurate peptide/GVP validation, and quantification, which is critically important for potential forensic applications