Top Down Mass Spectrometry With Multiple MS/MS Strategies to Identify Age-related Proteoform Changes in Tear Fluid

Daniel Lopez-Ferrer[1], Romain Huguet[1], David Horn[1], Peter M Raus[2], Greg Foster[1], Vlad Zabrouskov[1], Andreas FR Huhmer[1], Peter DEM Verhaert[3]
[1]Thermo Fisher Scientific, 355 River Oaks Pkwy, San Jose, CA, USA; [2]Miró Center for Oculoplastic Surgery, Geel, Belgium; [3]ProteoFormiX, Beerse, Belgium

ABSTRACT

Mass spectrometry based proteomics is the preferred method for in-depth characterization of the protein components of biological systems. Here we want to demonstrate the potential of top-down proteomics for mapping proteoform diversity. To this end, we have developed an standardized method to map proteoform diversity in tear fluid using UVPD fragmentation and high resolution accurate mass detection using Orbitrap technology. This method enables not only the identification and quantification of thousands of proteoforms with minimal sample preparation. But it has also the potential to stratify individuals according to their proteoform fingerprint. In addition, this workflow requires minimal sample preparation and allows for the interrogation of proteins in their intact state. This workflow was applied to map the tear fluid proteoform landscape in the context of age-related changes in the protein composition.

INTRODUCTION

Tear fluid is of key importance to maintain the health of the surface of the eye and to provide clear vision. Tears represent a promising body fluid that may help in the diagnosis and prognosis of various (eye) diseases and that is easily acquired by non-invasive means. Here we explore the use of top-down proteomics to map proteoform diversity in the tear fluid. The proposed workflow uses UVPD as the preferential top-down fragmentation technique on a chromatographic time scale that allows for identification of over 1000 proteoforms simultaneously in less than 1 hour. This workflow was successfully applied to a cohort of thirteen donors with ages ranging from 17 up to 59 years old to identify molecular changes due to age.

MATERIALS AND METHODS

Tear samples were purchased from Lee Biosolutions. Tears were directly transferred to an injection vial and analyzed by LC-MS using a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer modified with UVPD. MS/MS acquisition was performed using ETD, EThcD, HCD and UVPD fragmentations at a resolution of 120,000 at *m/z* 200. However, in this poster, only UVPD results are shown. Top-down protein identification was performed using ProSightPC™ 4.0 software, and the ProSightPD™ nodes in Thermo Scientific™ Proteome Discoverer™ 2.3 software. Label free analysis was done using Thermo Scientific™ BioPharma Finder™ 3.0 software. Statistical analysis to identify differentially expressed or processed proteoforms was done using InfernoRDN software (https://omics.pnl.gov/software/infernordn).

Figure 1. (A) Orbitrap Fusion laser source allows UVPD capabilities, (B) Schematic of different hardware parts including UVPD source (far right).

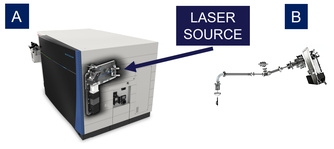
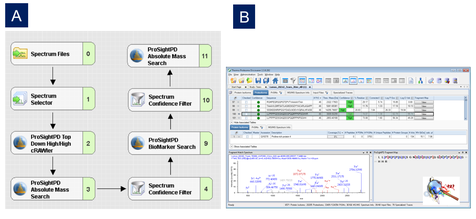


Figure 2. (A) Processing workflow for top-down proteomics analysis in Proteome Discoverer 2.3 SW using the ProSightPD nodes, (B) Screenshot of analysis output on Proteome Discoverer 2.3 SW, highlighting an identified proteoform of proline-rich protein 4, and showing the corresponding fragmentation spectrum and sequence coverage.



Study Design

Figure 3. Schematic representation of our simplified top down proteomics sample workflow (left panel) and the acquisition queue that includes system suitability runs, blanks and three randomized sample batches for each of the replicate measurements (right panel).

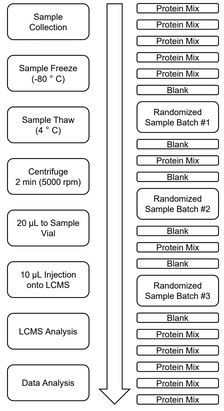


Table 1. Donor characteristics.

Cohort	Donor	Year	Age	Race	Gender
Young	5632	2001	17	Caucasian	Male
Young	6207	2000	18	Hispanic	Male
Young	5631	1999	19	Caucasian	Male
Young	4665	1993	25	Caucasian	Male
Young	5717	1991	27	Caucasian	Male
Young	6376	1989	29	Caucasian	Male
Young	6368	1988	30	Caucasian	Male
Old	5129	1986	32	Caucasian	Male
Old	6350	1976	42	Caucasian	Male
Old	5640	1970	48	Caucasian	Male
Old	6382	1967	51	Caucasian	Male
Old	6208	1966	52	Asian	Male
Old	6367	1959	59	Caucasian	Male

QC and Performance Monitoring

Figure 4. Representative chromatogram from a 60 min gradient of tear fluid without any processing.

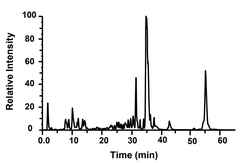


Figure 5. Representative chromatogram from a 30 min QC gradient for the Pierce Intact Protein Standard Mix. Inserts show different full MS scans across the chromatogram.

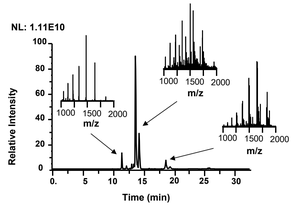
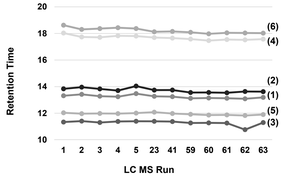


Figure 6. Retention time trends for the Thermo Scientific™ Pierce™ Intact Protein Standard Mix containing a mixture of six highly pure recombinant proteins, including IGF-1 LR3 (1), thioredoxin (2), Protein G (3), Carbonic Anhydrase II (4), Protein AG (5), and Exo Klenow (6). Retention time CVs are 0.9%, 1.2%, 1.5%, 0.9%, 0.6% and 1.1%, respectively.



Proteoform Quantitation

Figure 7. Histograms showing the number of quantified proteins (left panel) and proteoforms across the different LCMS runs of the study.

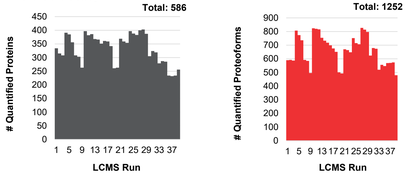


Figure 8. Tree map providing a hierarchical view of the quantified proteoforms per protein across all different donors. Most of the proteins are identified by 2 or 3 proteoforms. The most abundant tear proteoforms belong to lysozyme, lactoferrin, secretory immunoglobulin A, lipocalin, and dermicidin. In addition, Proline-rich protein 4 is represented by the largest number of proteoforms with almost 200 different proteoforms. This contradicts previous studies that have reported only two different proteoforms or cleavage products identified from this protein[1,2].

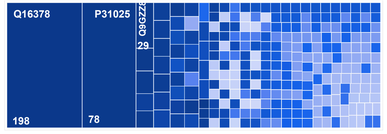


Figure 9. Pearson correlation plot from raw proteoform abundances. The high correlation among replicates is determined by the high reproducibility of the analytical setup. Interestingly, the graph shows large differences among donors.

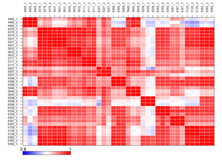
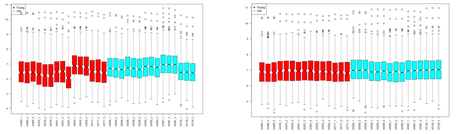
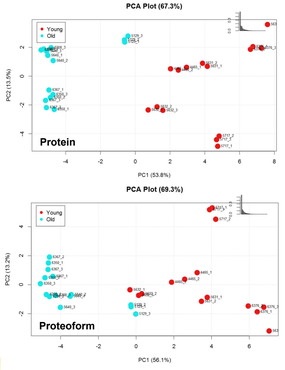


Figure 10. Box plots for all datasets before and after normalization. Normalization consisted of using local regression estimates to fit simple models to localized subsets of the data. Then, linear regression was used for each dataset within the age factor against the dataset with the least amount of data. This method detected proteins with abundances that span almost 5 orders of magnitude.



Proteoform Diversity

Figure 11. PCA plots from the differentially expressed (A) proteoforms and at the protein level (B), where proteoform abundances were rolled up at the protein level using a similar approaches as with bottom up proteomics. In both cases, old and young individuals can be discriminated. However, additional discrimination is observed at the proteoform level where donors cluster together in sub-classes. This effect is not seen when rolling up the proteoform abundances to the protein level.



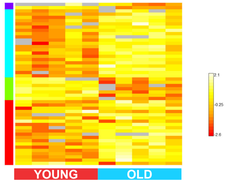


Figure 12. Cluster heatmap using k-means clustering from the differentially expressed proteoforms. We chose k=4 because we wanted to see if there would be other properties that cluster proteins among the two groups due to other factors rather than age. The data suggest that there are only two groups: those proteoforms that are more abundant in the young cohort and those more abundant in the old one.

Same Protein Different Proteoform Trends

Figure 13. (A) Box plots from differentially expressed proteoforms from Proline Rich Protein 4. Of the almost 200 detected proteoforms, 19 show statistically significant differences (p-value<0.01) between sample groups. 13 proteoforms of these proteoforms are more abundant in young donors while the remaining 6 proteoforms are more abundant in the old donors. (B) Box plots for differentially expressed proteoforms from Polymeric immunoglobulin receptor. In this case, one proteoform is more abundant in young donors and the other one is slightly more abundant in the older sample cohort. Another 6 more proteoforms were quantified, but no significant changes were detected between sample groups.

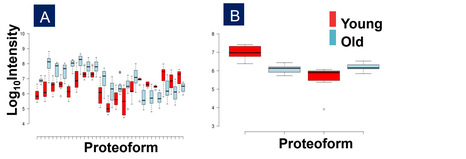
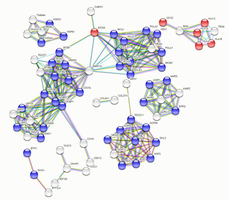


Figure 14. Network analysis using String 10.5 for the differentially expressed proteoforms. Blue circles denote proteins that interact selectively and non-covalently with organic cyclic compounds including DNA and RNA. Red circles denote proteins with antigen binding functions.



CONCLUSIONS

UVPD coupled to high resolution mass spectrometry and sophisticated bioinformatics tools represents a simple and versatile technique for proteoform profiling of human body fluids such as tears.

- The combination of UVPD dissociation, high resolution and accurate mass Orbitrap mass spectrometry with ProSightPD for Proteome Discoverer provides sensitive analysis using small sample amounts.
- This work illustrates that robust profiling and monitoring of hundreds of proteoforms in tears in less than 60 min is possible. This approach shows promise for single donor tear analysis in ophthalmological clinical research applications.

REFERENCES

1. Perumal, N.; Funke, S.; Wolters, D.; Pfeiffer, N.; Grus, F.H. Characterization of human reflex tear proteome reveals high expression of lacrimal proline-rich protein 4 (PRR4). *Proteomics* **2015,***15* (19), 3370–81.
2. Aluru, S. V.; Agarwal, S.; Srinivasan, B.; et al. Lacrimal proline rich 4 (LPRR4) protein in the tear fluid is a potential biomarker of dry eye syndrome. *PLoS One* **2012,** *7* (12).

TRADEMARKS/LICENSING

PO65252-EN 0518S

ThermoFisher SCIENTIFIC