NuXL in Proteome Discoverer Software - A specialized database search engine for the analysis of DNA-/RNA-protein crosslink data

Bernard Delanghe[1], Timo Sachsenberg[2,3], Luisa Welp[4], Aditi Sharma[1,9], Arslan Siraj[3], Alexander Wulf[4], Fanni Bazso[4], Aleksandar Chernev[4], Yi He[8], Kai Fritzemeier[1], Rosa Viner[8], Oliver Kohlbacher[2,3,5,6], Henning Urlaub[4,7], [1]Thermo Fisher Scientific (Bremen) GmbH, 28199 Bremen, Germany, [2]Institute of Bioinformatics and Medical Informatics, University of Tübingen, 72076 Tübingen, Germany, [3]Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, 72076 Tübingen, Germany, [4]Bioanalytical Mass Spectrometry Group, Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany, [5]Institute of Translational Bioinformatics, University Hospital Tübingen, 72076 Tübingen, Germany, [6]Biomolecular Interactions, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany, [7]Bioanalytics Group, Institute of Clinical Chemistry, University Medical Center Göttingen, 37075 Göttingen, Germany, [8]Thermo Fisher Scientific, San Jose, USA, [9]TUM School of Life Sciences, Technical University of Munich, Germany

Abstract

Purpose: Create a complete data processing workflow for DNA and RNA – protein crosslink data.

Methods: Incorporation of the NuXL algorithm into the Proteome Discoverer framework.

Results: UV- and chemically-induced crosslinks can be identified and quantified.

Introduction

Technical and methodological advances enable to apply crosslinking mass spectrometry (XL-MS) for the identification of DNA and RNA binding sites within proteins *in vitro*. The use of chemical crosslinking agents and incorporation of photoactivatable nucleotides increases crosslinking efficiencies broadening the applicability of XL-MS to *in vivo* studies of DNA-/RNA-protein interactions. Hence, the demand for specialized software tools is growing to identify crosslinked peptides, including site-localizations. We introduce the new database search engine NuXL, implemented into the Thermo Scientific™ Proteome Discoverer™ 3.0 software, which reliably identifies DNA- and RNA-crosslinked peptides supporting a variety of crosslinking techniques. Implementing the novel NuXL tool into the Proteome Discoverer software graphical user interface (GUI) enables state-of-the-art computational methods for XL discovery easy to configure, execute and visualize. The plugin is written in C# and wraps the configuration and execution of NuXL as well as parsing and processing of the produced crosslinking results.

- RNA/DNA protein complexes are essential for many cellular processes like DNA replication, DNA repair, transcription, splicing, RNA maturation, translation control etc.
- Defects in nucleic acid processing proteins are linked to severe diseases.
- Determination of the sites of interactions between proteins and RNA and/or DNA is required.
- Standard methods like X-ray crystallography, NMR, or cryo-EM is the golden standard to study large and dynamic RNA and/or DNA protein complexes but can be challenging and cannot be applied on entire cells.
- XL-MS offers a straightforward method to identify proteins, which interact with RNA and/or DNA on the molecular and atomic level not only in isolated complexes but also in entire cellular systems.

Materials and methods

The NuXL node is designed to identify UV- and chemically-induced crosslinks by taking individual MS fragmentation behavior into account. Respective presets have been assembled based on our own *in vitro* and *in vivo* datasets including human nucleosomes, HeLa cells and Escherichia coli cells as well as on published data. We calculate match-odds and subscores and employ semi-supervised score calibration using Percolator. Moreover, entrapment experiments have been employed using manually curated data ensuring proper false discovery rate (FDR) control.

The Proteome Discoverer software is a versatile suite for mass spectrometry data processing. It features a user-friendly GUI and allows processing raw spectra and perform standard proteomic analysis tasks using a flexible workflow system. By providing a public API, Proteome Discoverer software functionality can be extended with third-party plugins.

We integrate NuXL into Proteome Discover software by wrapping the novel tool as community extension. The NuXL Proteome Discoverer software node integrates seamlessly and can be configured and executed like any other node, providing in addition quantification of the crosslinked peptides. Furthermore, data acquired with FAIMS can be processed without any transformation of the data.

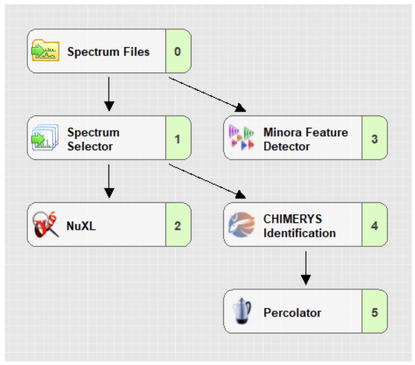
NuXL supports a wide range of crosslinking protocols, including standard UV-crosslinking, nucleotide analogs, and novel chemical crosslinkers. Users can add support for novel protocols by adding the precursor adducts, and fragment adducts according to the sample processing. The spectra will be annotated according to fragment adducts; manual validation is possible.

Table 1. Supported protocols

	UV XL	UV XL (4SU)	UV XL (6SG)	DEB	NM	FA
RNA	√	√	√	√	√	√
DNA	√			√	√	?

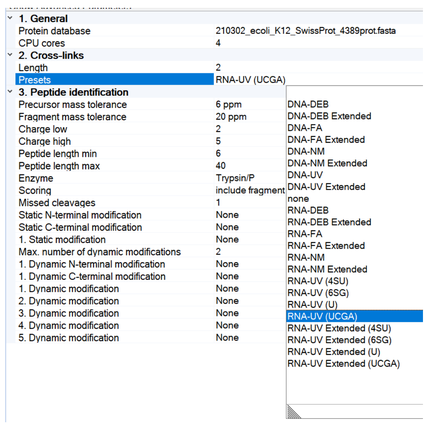
DEB: Diepoxybutane, NM: nitrogen mustard, 4SU: 4 Thiouridine, 6SG: 6-Thioguanosine, FA: Formaldehyde

Figure 1. Processing workflow including label-free quantification (LFQ)



Crosslinked peptides are identified by the NuXL search node, regular peptides by the CHIMERYS™ search engine. The Minora nodes is calculating the abundances for all the peptides

Figure 2. Presets available in the NuXL node



Results

NuXL robustly identifies crosslinked peptides and amino acids from XL-MS data offering optimized search settings for a broad spectrum of crosslinking agents (UV, 4SU, 6SG, formaldehyde, 1,2:3,4-diepoxybutane and mechlorethamine) for both RNA- and DNA-protein crosslink samples. Further, NuXL provides FDR control on the level of crosslinked peptide spectrum matches.

Figure 3. Proteome Discoverer software result file

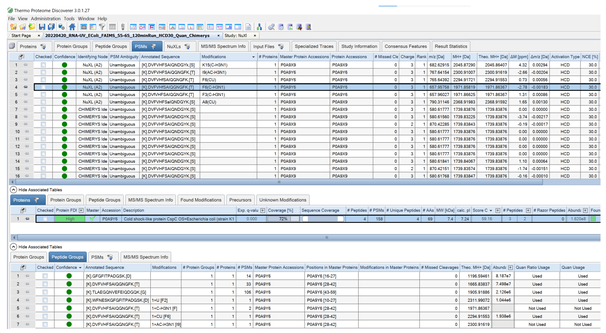
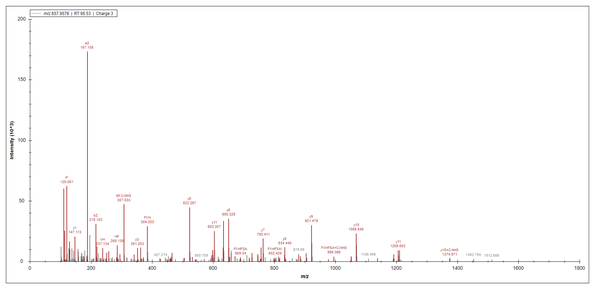


Figure 4. Annotated MS2 spectrum of a UV induced RNA-peptide crosslink



The MS2 spectrum view in Proteome Discoverer software of the peptide DVFVHfSAIQGNGFK crosslinked with C-H3N

Outlook

We are offering standardized and crosslinked RNA-peptide samples suitable for subjecting directly to LC-MS/MS in order to evaluate the data search and annotation of the NuXL tool in Proteome Discoverer software. Moreover, we are offering standardized and crosslinked RNA-protein samples for performing own crosslink purification for subsequent LC-MS/MS and NuXL data analysis. Additional features to be integrated into the NuXL node of Proteome Discoverer software:

- Visualization of crosslinking sites on 3D structures of proteins and RNA/DNA protein complexes
- Support for studying peptides with crosslinked RNA comprising more than four nucleotides.
- Improved site localization

Availability

The binary installer, documentation, and example data for the NuXL Proteome Discoverer software Nodes are available free of charge as part of a beta testing agreement**.**

Conclusions

We introduce the user-friendly database search engine NuXL, implemented in Proteome Discoverer software developed for the analysis of DNA-/RNA-protein UV- and chemically-induced crosslink data. Default settings for commonly used crosslinking techniques are provided and the workflow allows for custom defined settings that can be individually adapted to other crosslinking agents. As NuXL is embedded in the Proteome Discoverer software environment, it can be combined with downstream analysis nodes enabling LFQ of crosslink sites.

References

1. Stützer, A., Welp, L. M., Raabe, M., Sachsenberg, T., Kappert, C., Wulf, A., ... & Urlaub, H. (2020). Analysis of protein-DNA interactions in chromatin by UV induced cross-linking and mass spectrometry. Nature Communications, 11(1), 1-12
2. Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., & Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nature Methods, 11(10), 1064-1070.
3. Veit, J., Sachsenberg, T., Chernev, A., Aicheler, F., Urlaub, H., & Kohlbacher, O. (2016). LFQProfiler and RNPxl: open-source tools for label-free quantification and protein–RNA cross-linking integrated into proteome discoverer. Journal of Proteome Research, 15(9), 3441-3448.

Acknowledgements

This project has received funding from the European

Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement Nº 956148.

Trademarks/licensing

Learn more at thermofisher.com/MS-Structure