# **NuXL**

# Protein-RNA and Protein-DNA cross-linking Nodes

# for Proteome Discoverer 3.1

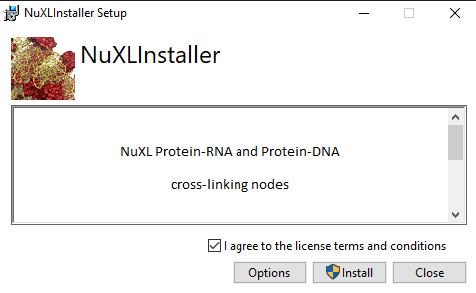
# Official beta program – Do not distribute without permission.

# User Manual

## Installation

Installing the NuXL plugin for Proteome Discoverer is easy: just run the binary installer   
*NuXL-PD3.1.exe* and it should automatically detect your PD folder and install the nodes in the correct location.

|  |  |
| --- | --- |
| ! | Proteome Discoverer should be closed before running the installer. |



After successful installation, you should find the NuXL processing and consensus nodes in your Proteome Discoverer workflow editor.

|  |  |
| --- | --- |
| ? | In rare cases, the installer does not work, because the Proteome Discoverer registry key cannot be found. Should this be the case, please contact us for instruction how to perform a manual installation. |

### Usage

### Setting up the study

1. Start Proteome Discoverer 3.1 and create a new study (e.g., name it *NuXL*).

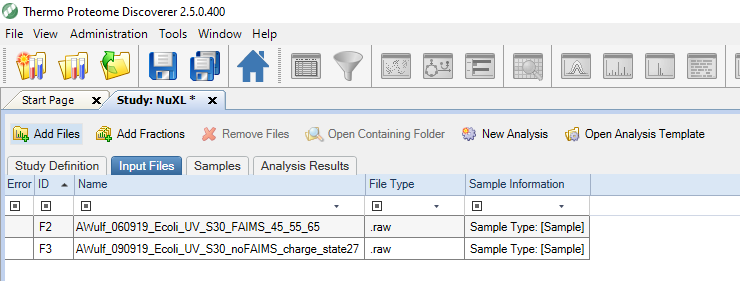
### Test Datasets and Databases

For this tutorial, we provide the following datasets consisting of spectra data, PDANALYSIS files with preconfigured nodes and protein databases (see Supp. for details):

|  |  |  |
| --- | --- | --- |
| **Spectra files (.raw)** | **PDANALYSIS file** | **Database (.fasta)** |
| LWelp\_100221\_160221\_BPSHeLanucleosomes\_DNA\_UV  F\_Bazso\_120919\_130209\_Monoonucleosomes\_DNA\_DEBLWelp\_100221\_120221\_BPSHeLanucleosomes\_DNA\_NM  AChernev\_080219\_HeLa\_RNA\_UV  MR\_AWulf\_010819\_150121\_Ecoli\_RNA\_DEB  LW\_101117\_101117\_Hsh49\_polyG\_RNA\_NM | DNA-UV  DNA-DEB  DNA-NM  RNA-UV  RNA-DEB  RNA-NM | 200315\_BPSHeLanucleosomes\_linearpeptideDB  CoreHistones\_H14  200315\_BPSHeLanucleosomes\_linearpeptideDB  210302\_hsapiens\_SwissProt\_20381prot  210302\_ecoli\_K12\_SwissProt\_4389prot  Hsh49 |

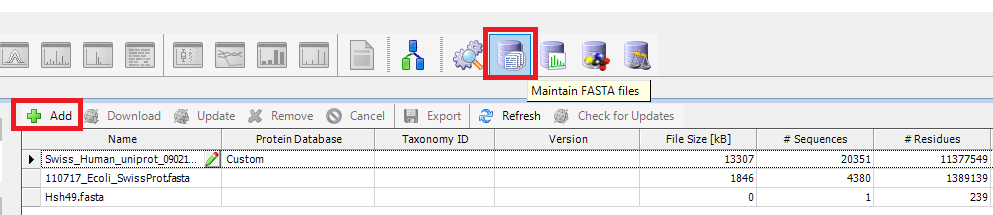
Table 1: Tutorial data

Add the Thermo raw files you want to analyze.



### Adding a protein database

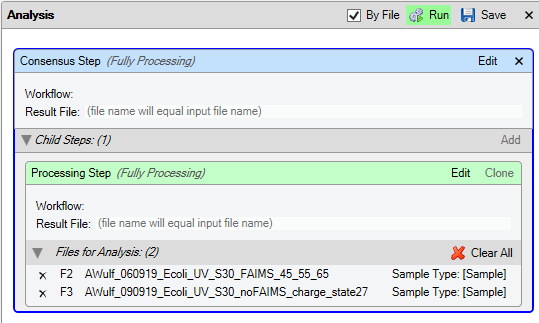
Click on Maintain FASTA file to add a new protein database to proteome discoverer.



### Setting up the workflows from a preconfigured analysis template

The easiest way of getting started is to use the preconfigured processing and consensus workflows from the analysis templates:

1. Make sure to open the correct analysis template for the input files.
2. Add the input files by dragging them on the Processing Step in the **Analysis** tab.
3. Check the **By File** checkbox.



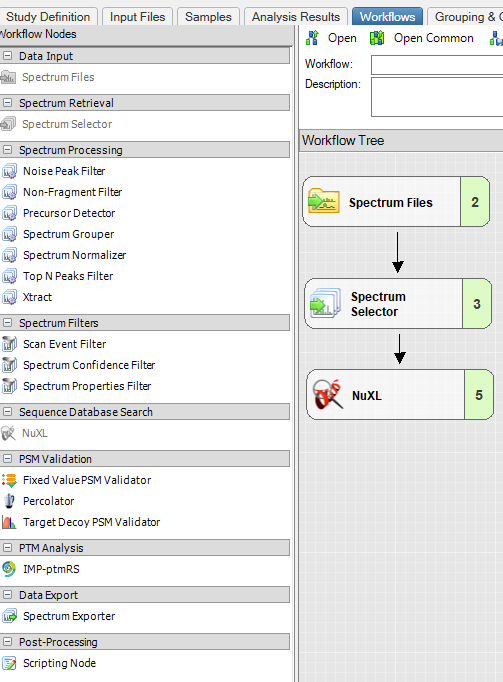
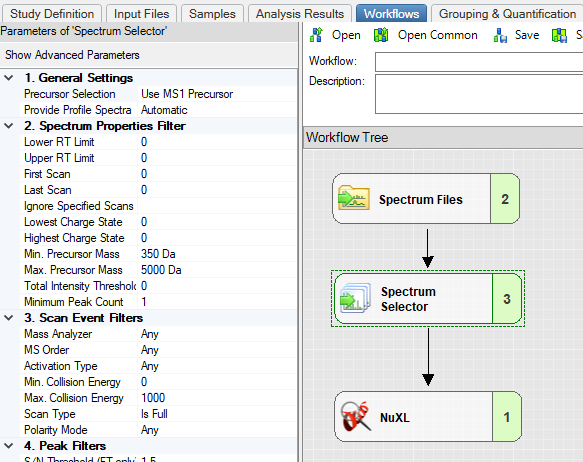
|  |  |
| --- | --- |
| ! | Checking the By File checkbox is crucial for proper processing in this beta version. |

|  |  |
| --- | --- |
| ! | Choosing the correct analysis template is crucial for valid search results. E.g., searching for chemical cross-links in UV data will not yield meaningful results. |

Your project is now set up and you can click on **Run** to execute the analysis

### Advanced: setting up a custom analysis workflow

If you prefer to set up the workflows yourself, we recommend building the processing workflow so that it looks like the figure to the right. You will find the NuXL node in the “Sequence Database Search” category.



|  |  |
| --- | --- |
| ! | Make sure to set “MS Order” to “Any” in the Spectrum Selector node (otherwise, MS1 spectra are discarded and the workflow makes use of them). Adapt the parameters of the NuXL processing node as needed. |

The consensus workflow is even simpler:

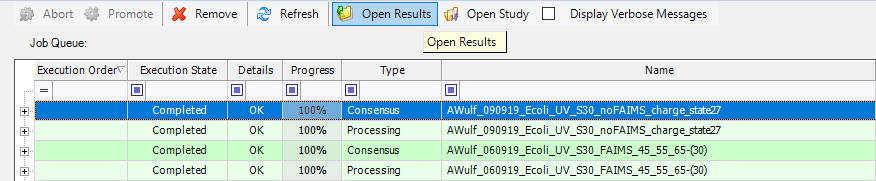
1. Connect an MSF Files node to an NuXL Consensus node.

|  |  |
| --- | --- |
| ! | Make sure to set “Spectra to Store” to “All” in the MSF Files node. |

1. Drag&drop the input files by dragging them on the Processing Step in the **Analysis** tab.
2. Check the **By File** checkbox.
3. Run the workflow

# Results

Once your jobs have completed you can open the results from the **Administration** tab by double-clicking (or selecting “**Open Results**”) on a row that shows Type “Consensus”.



The “Proteins” and “PSMs” tables are empty in this beta version. The results will be contained in the “XLs” table (see Figure below). The most important columns are **Proteins** and **Peptide** as well as the **Score** (= Cross-link PSM-level q-value).

|  |  |
| --- | --- |
| ! | Note that the q-value is calculated in the group of cross-link spectrum matches (CSM). It is usually higher than a PSM-level q-value (which considers both cross-linked and non-cross-linked peptides). |

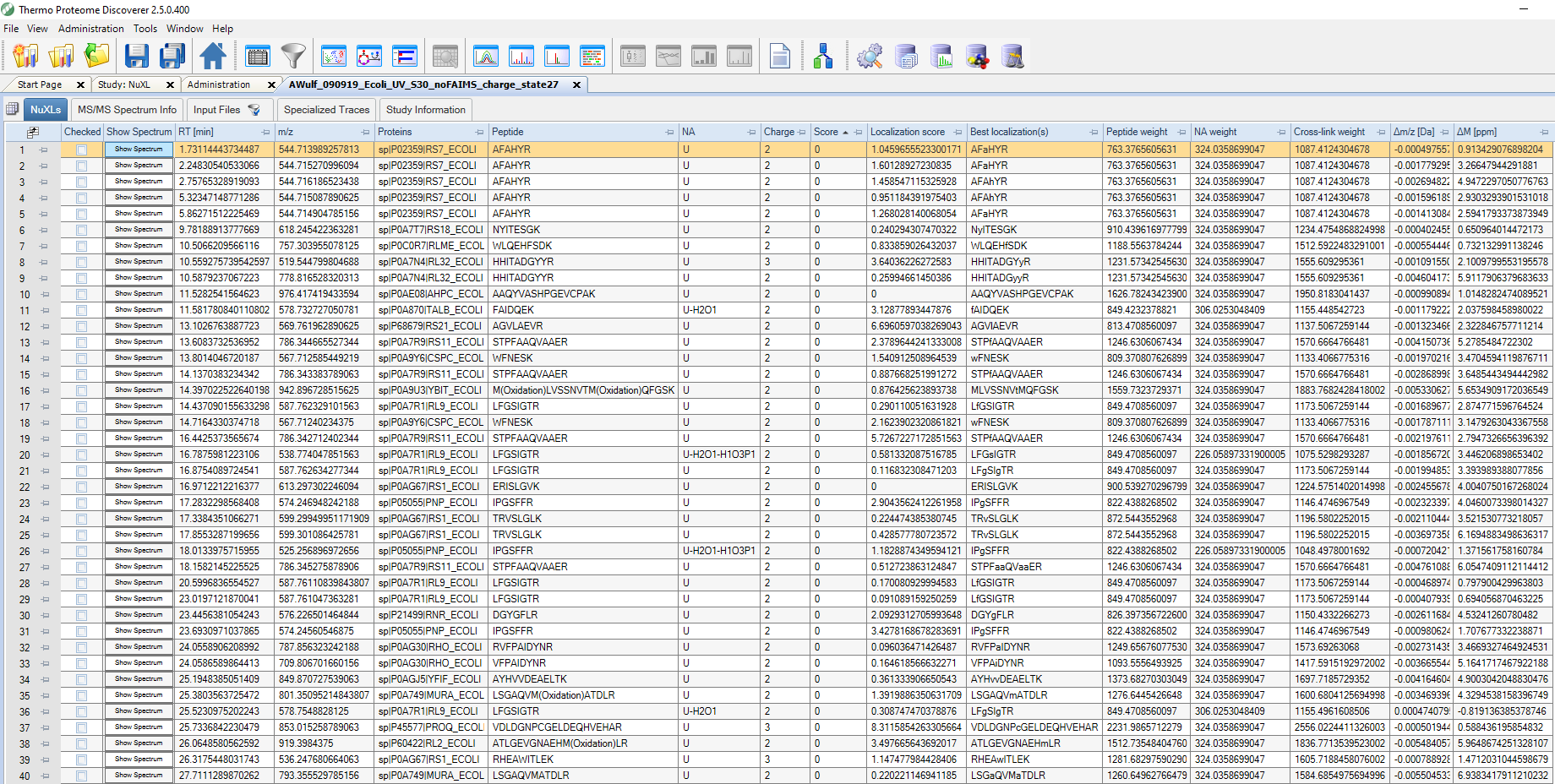
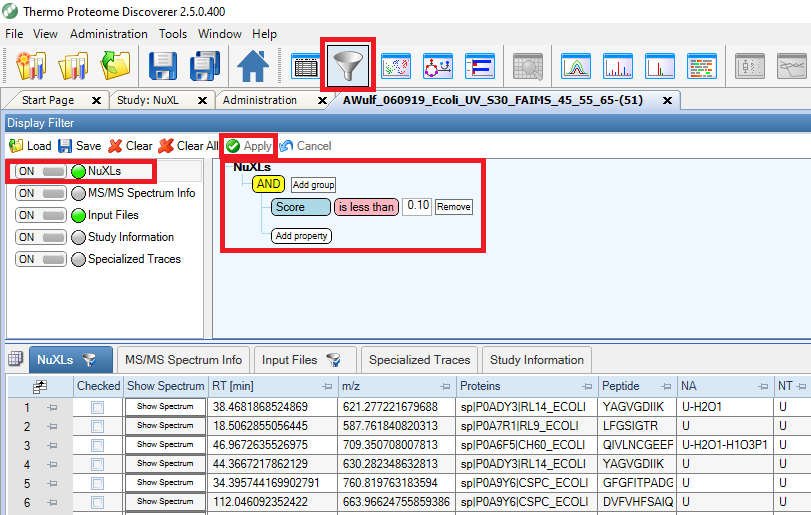


Figure 1 Example output of cross-link search.

Note: You can remove bad scoring CSMs by adding a filter to the score column that retains all cross-links at a CSM-FDR of 10%. Just add a “**Score** *is less than* **0.1**” filter on the **NuXLs** Table for q-value < 0.1 as shown in the picture below.

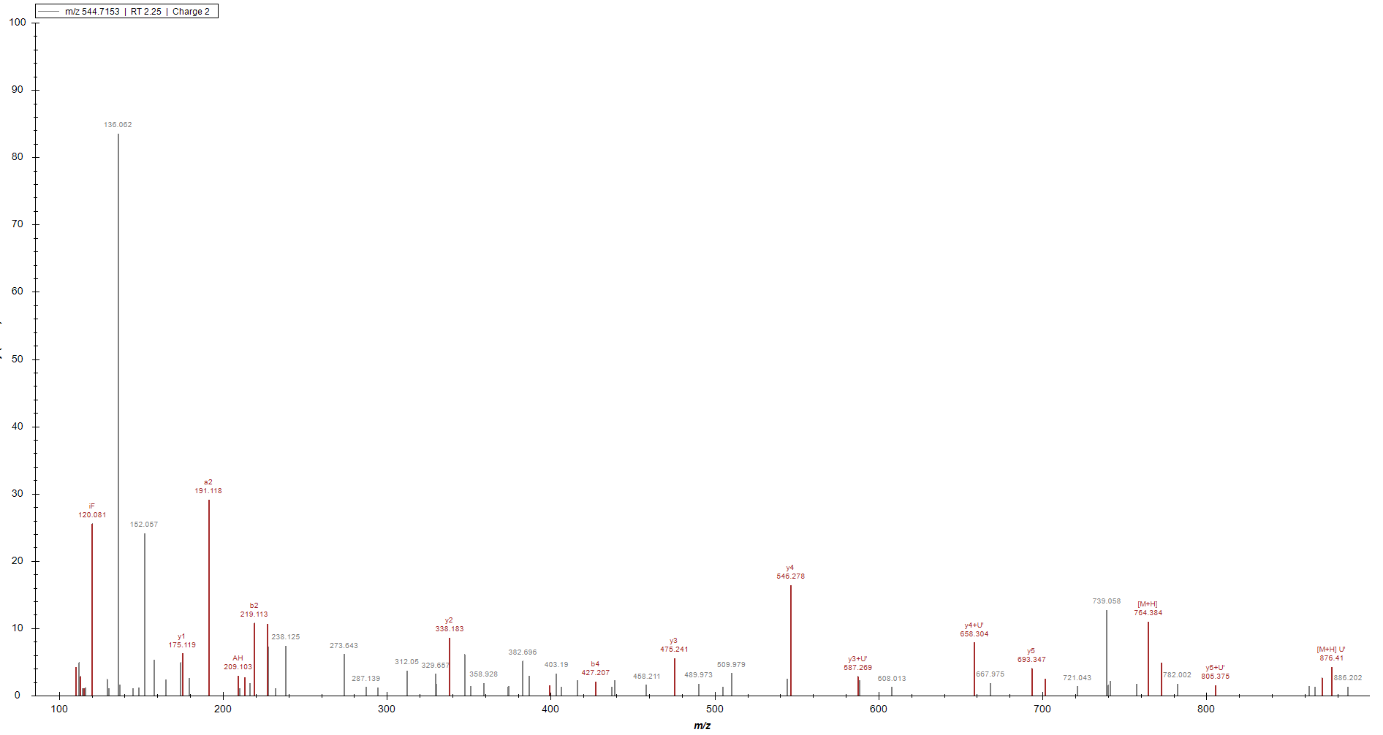


See the table below for a description of table content:

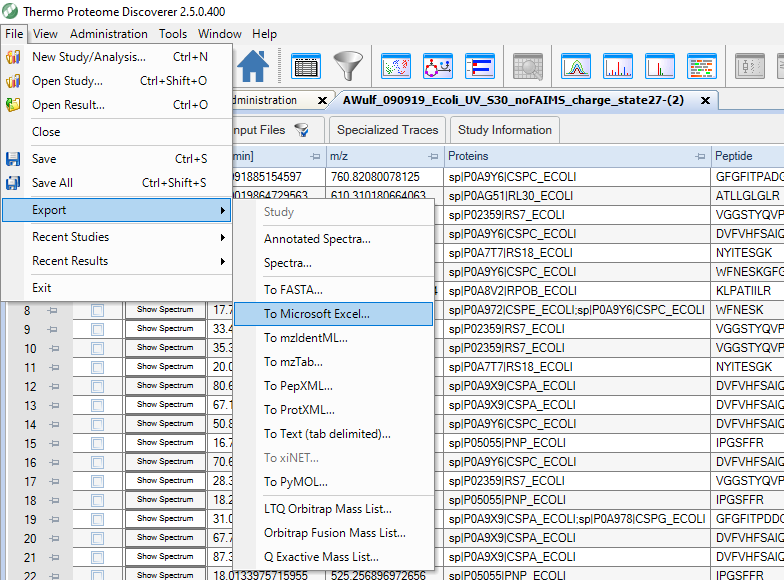
|  |  |
| --- | --- |
| Column | Description |
| Show Spectrum | Click button to visualize annotated spectrum |
| RT [min] | Retention time of the MS/MS spectrum |
| m/z | Precursor m/z |
| **Proteins** | **Protein** **accessions** |
| **Peptide** | **Peptide sequence** |
| NA | Cross-linked precursor nucleotide adduct |
| NT | Cross-linked nucleotide |
| Charge | Precursor charge |
| **Score** | **Cross-link PSM-level q-value (lower is better)** |
| Localization score | Score of the best localization site (higher is better) |
| Best localizations | The cross-linked position(s) (marked in lower-case) |
| Peptide weight | for z=0 |
| NA weight | for z=0 |
| XL weight | for z=0 |
| ∆m/d [Da] | Absolute precursor mass error |
| ∆M [ppm] | Relative precursor mass error |

Table 2 Result columns

You can also visualize cross-link PSMs by clicking on the “Show Spectrum” button.

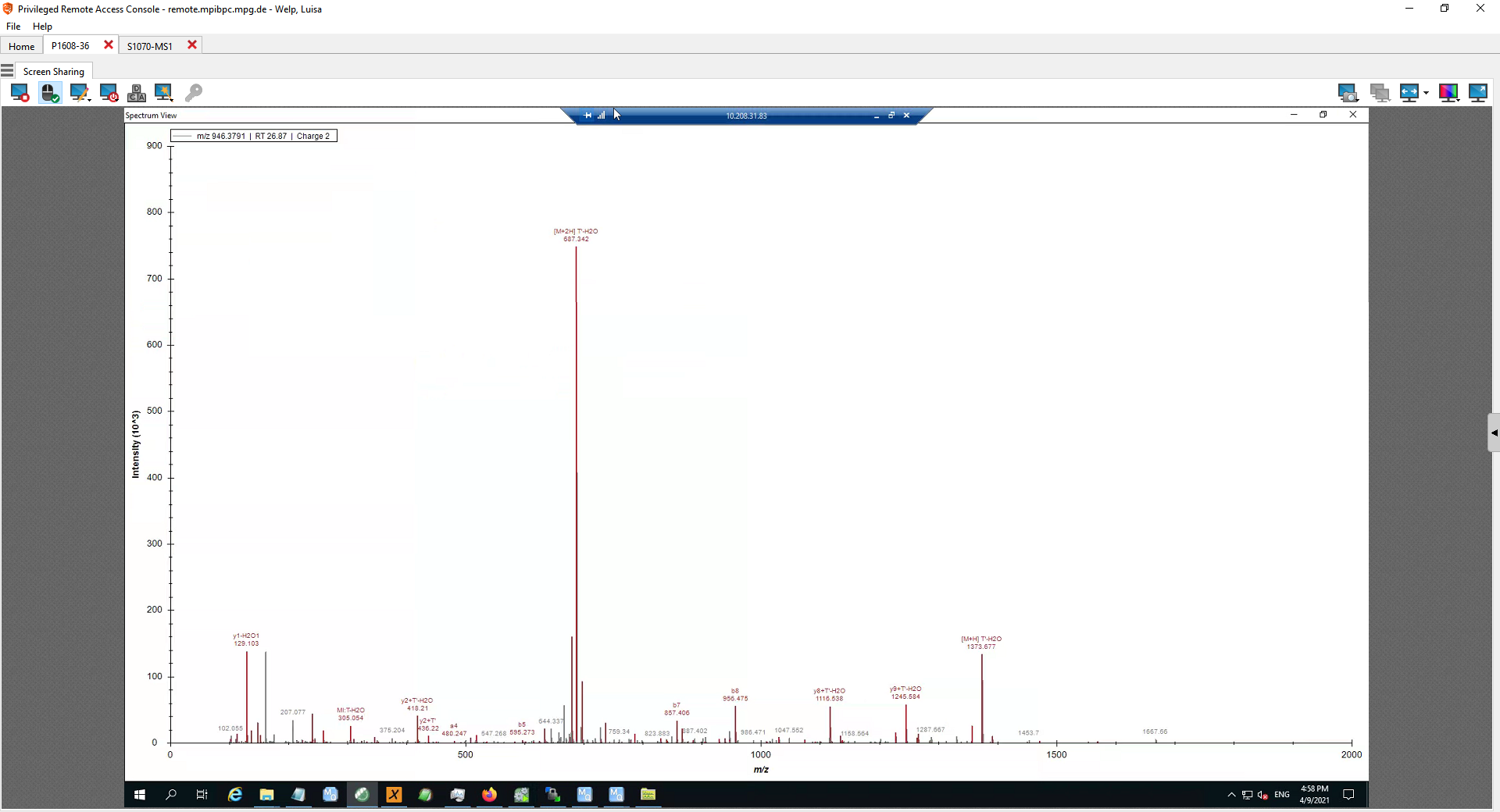
Figure 2 Spectrum visualization

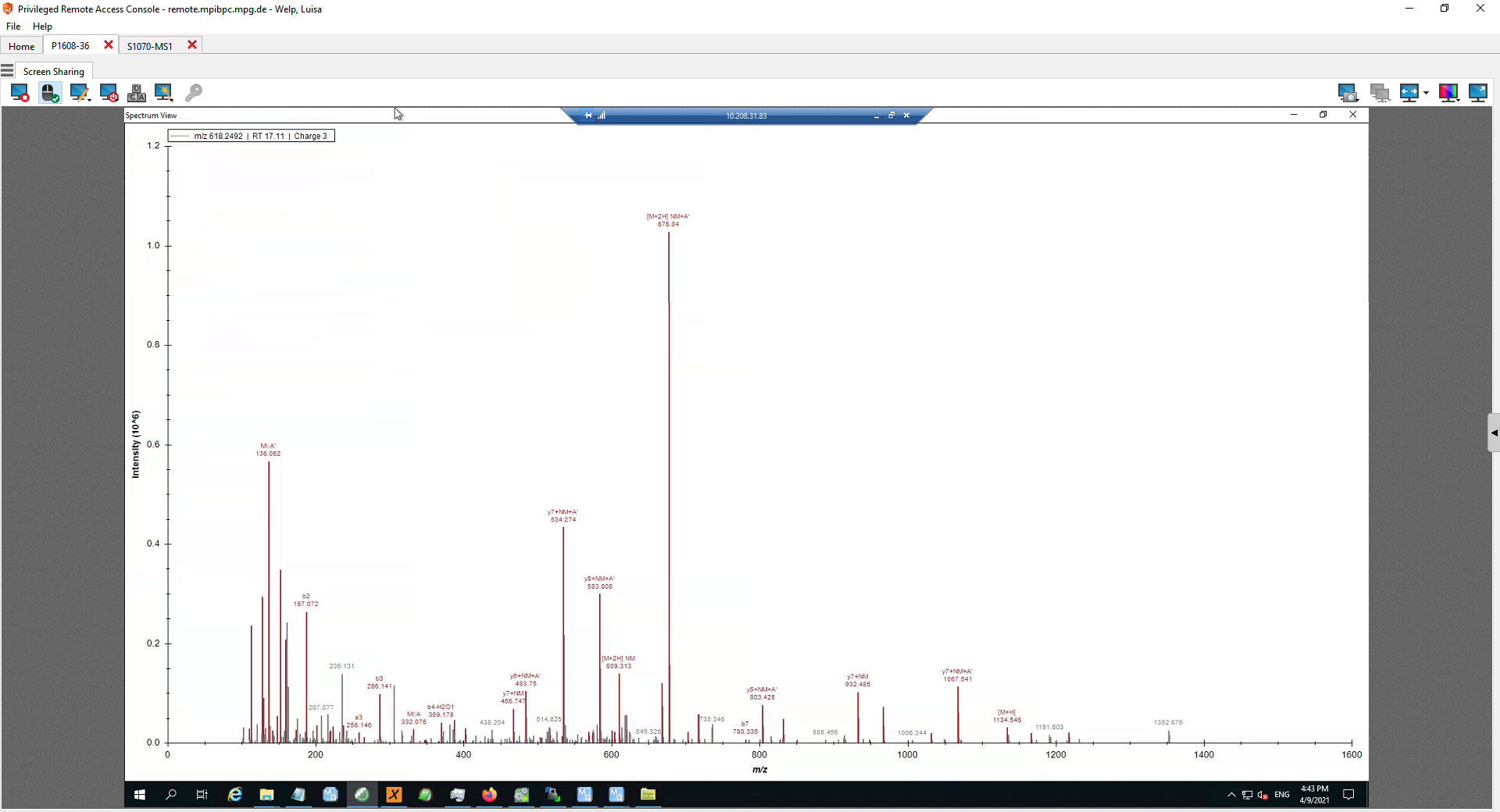
To export the result table to a spreadsheet software (e.g., Microsoft Excel) click on File->Export->To Microsoft Excel

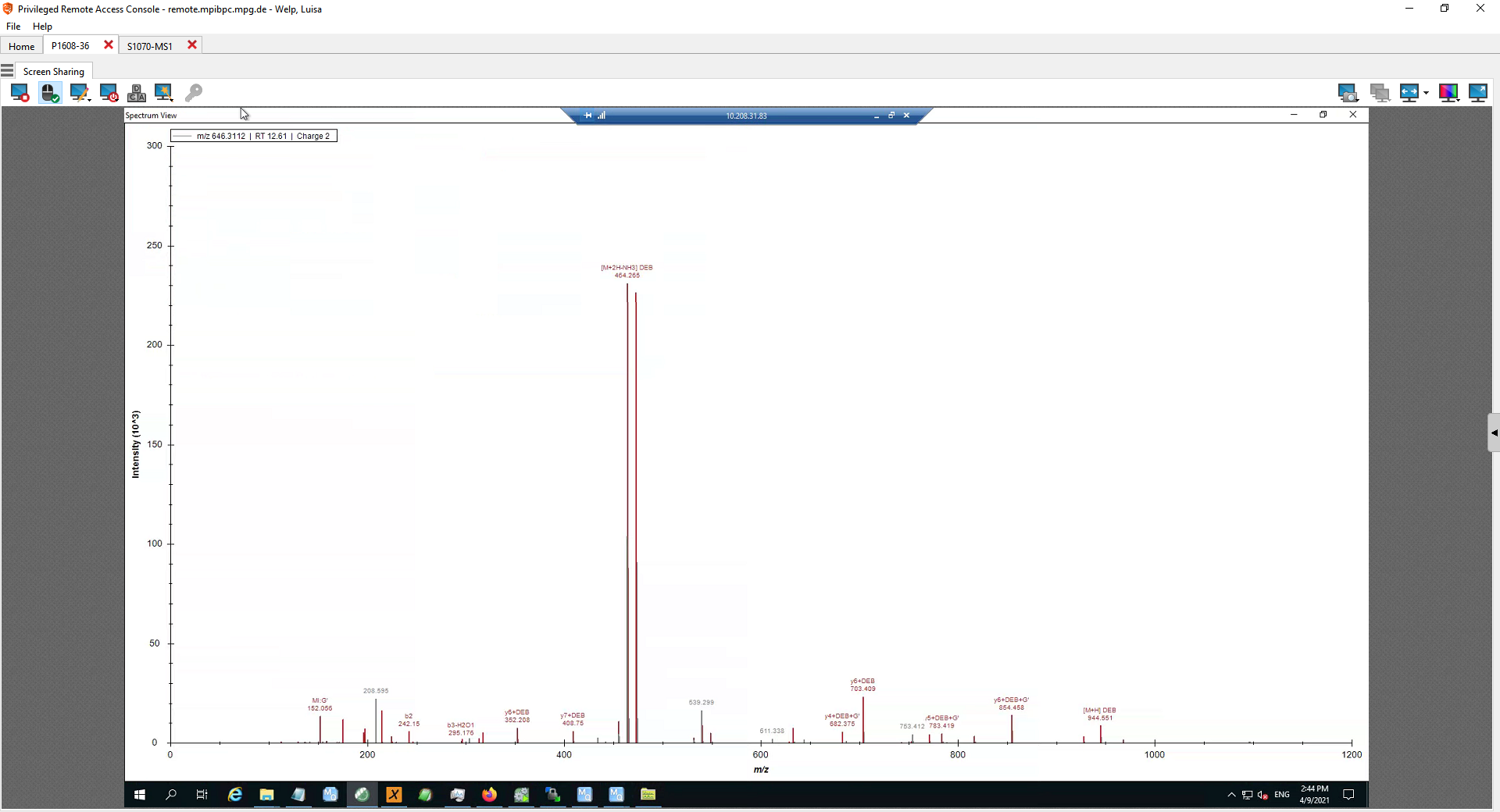


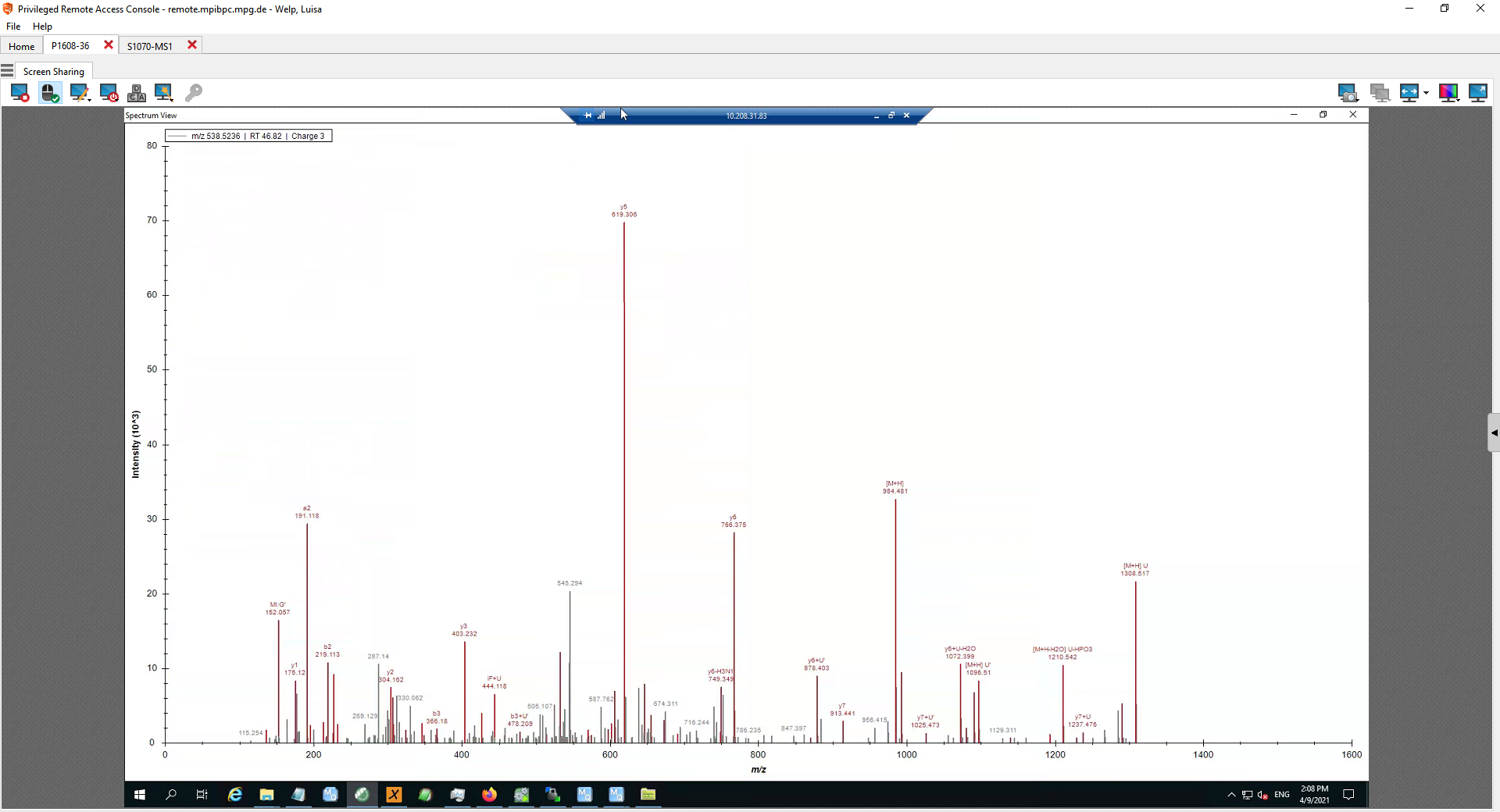
## Spectrum Quality Assessment

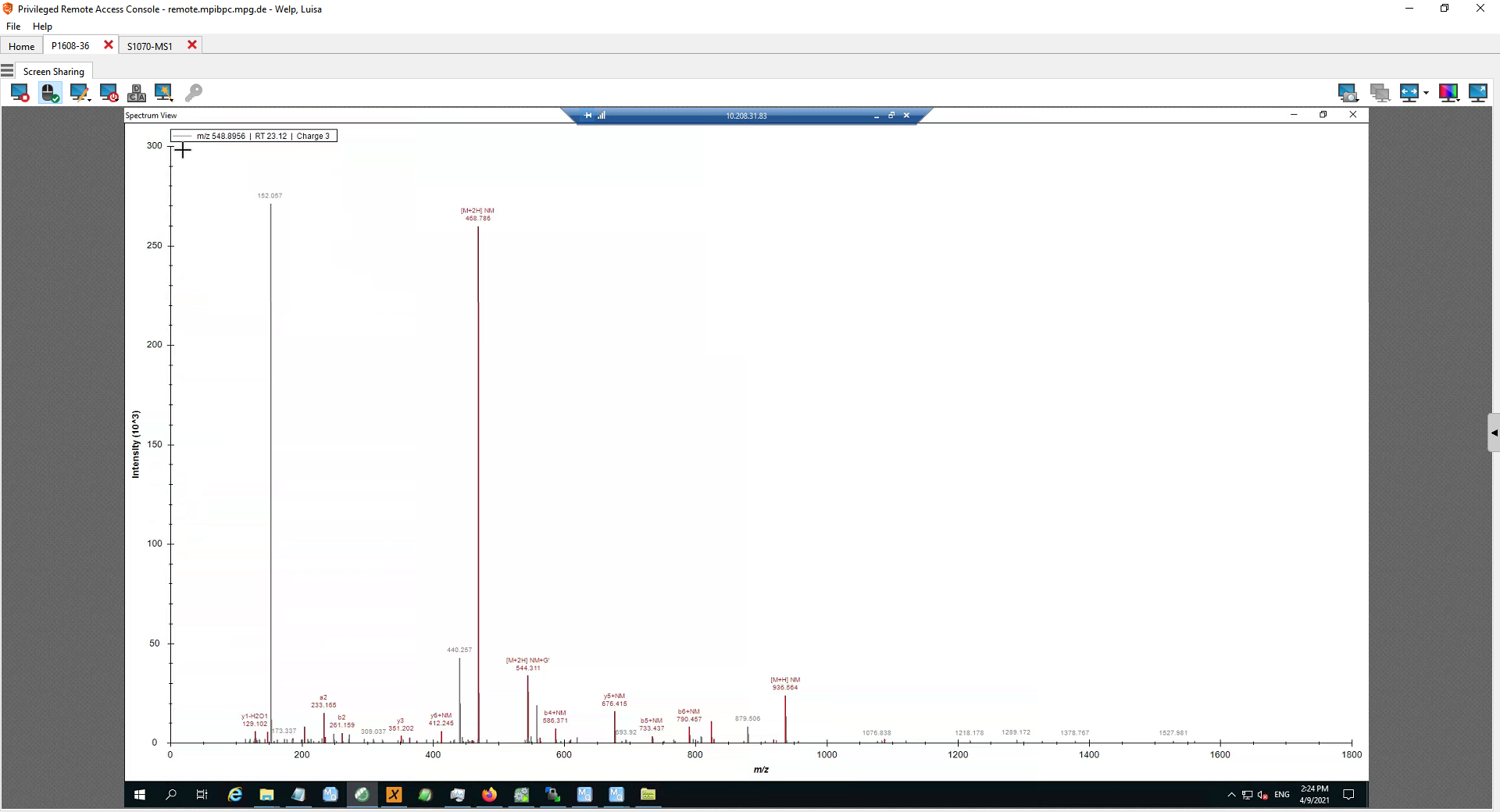
After filtering the cross-link data, the spectra can be manually validated by opening the spectrum view as described. Figure 3 to 8 show examples of good quality spectra in the given test datasets. Features that need to be considered are the peptide sequence coverage, the presence of a (mass shifted) precursor, immonium-, fragment- and internal ions as well as the presence of marker ions. The quality of a cross-link spectrum also depends on the level of noise/not annotated peaks.

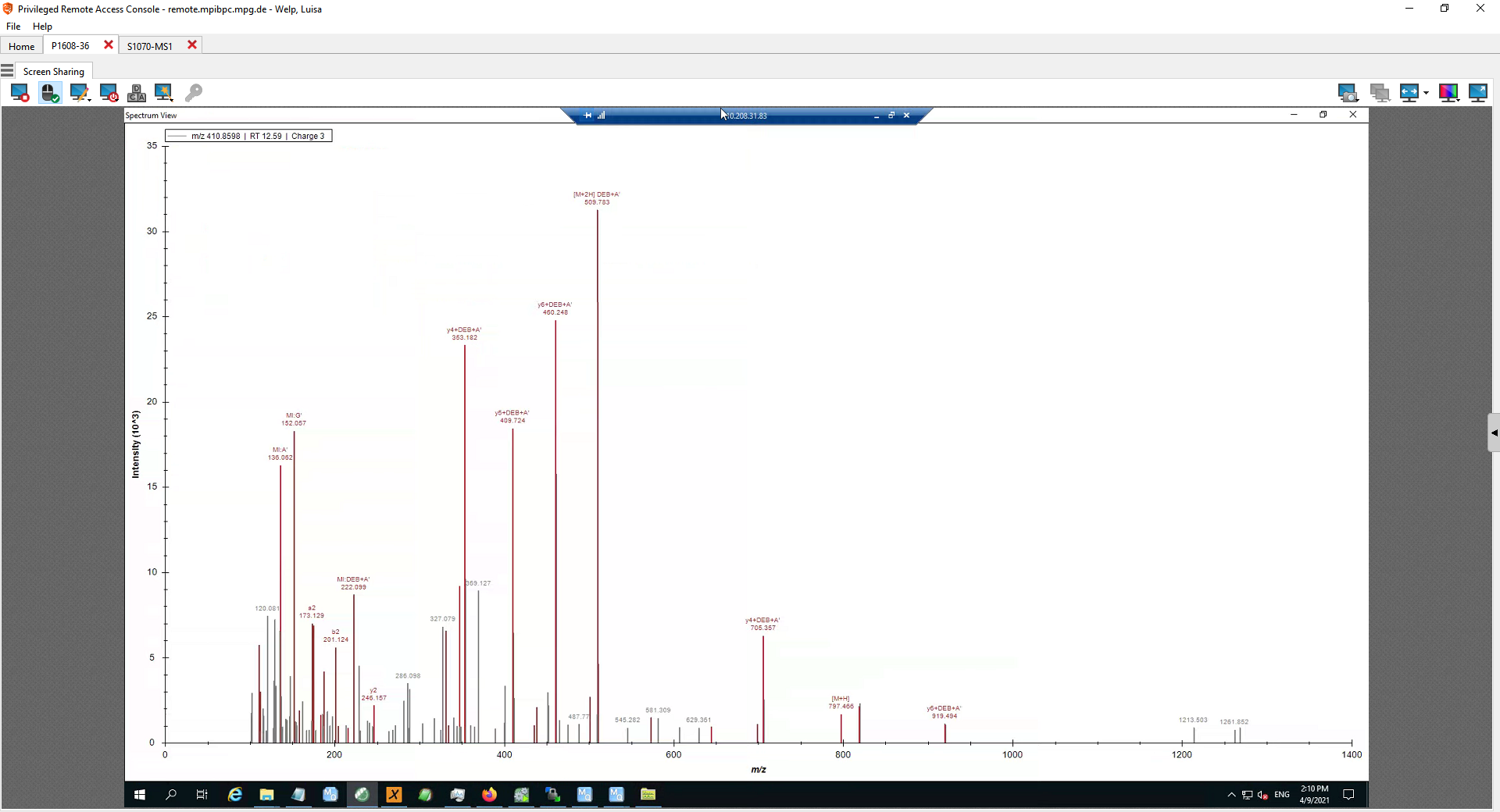
Figure 3 Cross-link spectrum of H2B peptide KESYSVYVYK in DNA-UV dataset

Figure 4 Cross-link spectrum of H4 peptide DAVTYTEHAK in DNA-NM dataset

Figure 5 Cross-link spectrum of H3 peptide QLATKAAR in DNA-DEB dataset

Figure 6 Cross-link spectrum of PA2G4 peptide AFFSEVER in RNA-UV dataset

Figure 7 Cross-link spectrum of HSH49 peptide IFNKFGK in RNA-NM dataset

Figure 8 Cross-link spectrum of RS3 peptide VTIHTAR in RNA-DEB dataset

## Sample and Result Description – Test Datasets

Samples were cross-linked either with UV-light for 10 min at 254 nm, or by the addition of 50 mM 1,2:3,4-diepoybutane (DEB) or 1/10 mM Mechlorethamine (NM).

DEB and NM are *bis*-electrophilic molecules. The cross-linking reaction involves two consecutive nucleophilic attacks by RNA/DNA bases and/or nucleophilic amino acids (e.g., cysteines).

### DNA-UV, DNA-NM: BPS purified HeLa nucleosomes

HeLa nucleosomes were purchased from BPS Biosciences and cross-linked with 10 mM NM or UV-light. Cross-linked peptide-deoxy(oligo)nucleotides were analyzed by LC-MS/MS and acquired data was analyzed with NuXL in Proteome Discoverer 3.1. The most abundant proteins in this sample are the four core histones H2A, H2B, H3 and H4. In addition, the sample contains other nuclear and cytosolic proteins that were not completely removed by nucleosome purification. For data analysis, we provide a FASTA file including sequences of all proteins identified in a proteome analysis of the same sample.

Among the top-scoring (1% FDR, "fragment adducts included") cross-linked proteins in the DNA-UV experiment, all four core-histones are identified as well as other DNA-binding proteins like Heterogeneous nuclear ribonucleoproteins C1/C2 (P07910), Heterogeneous nuclear ribonucleoprotein D-like (O14979), DNA replication licensing factor MCM3 (P25205), Lamin-B1 (P20700), Lysine-specific histone demethylase 1A (O60341) etc., and known RNA-binding proteins like several ribosomal proteins from the 40S and 60S subunit (P62913, P62888, P30050, P60866 etc.).

The NM cross-linking results (10% FDR, "fragment adducts included"), as identified with NuXL, are consistent with the UV cross-linking results. All four core histones are found to be cross-linked to DNA as well as other DNA binders (NF-X1-type zinc finger protein NFXL1 (Q6ZNB6), Heterogeneous nuclear ribonucleoprotein D-like (O14979) and C1/C2 (P07910) etc.) and RNA binders (60S ribosomal protein L10a (P62906), 60S ribosomal protein L4 (P36578) etc.).

### DNA-DEB: *in vitro* reconstituted mononucleosomes

Mononucleosomes were *in vitro* reconstituted from 187bp nucleosomal DNA and core histones H2A, H2B, H3 and H4. Reconstituted mononucleosomes were cross-linked with 50 mM DEB. Cross-linked peptide-deoxy(oligo)nucleotides were analyzed by LC-MS/MS and acquired data was analyzed with NuXL in Proteome Discoverer 3.1. At 1% FDR ("fragment adducts included"), 61 different core histone peptides are identified carrying adducts comprised of DEB plus single nucleotides or dinucleotides. Almost all possible combinations of nucleotide adducts are present in this dataset.

### RNA-UV: HeLa cytosolic extract

HeLa S3 cytoplasmic extract was cross-linked for 10 min at 254 nm UV-light. Cross-linked peptide- (oligo)nucleotides were analyzed by LC-MS/MS and acquired data was analyzed with NuXL in Proteome Discoverer 3.1. The provided database contains all reviewed *Homo sapiens* protein entries from Swissprot (downloaded on 02/03/2021; 20,381 entries). 51 out of the 229 cross-linked proteins (1%FDR, "fragment adducts included") are ribosomal proteins. Also, you will find other known RNA binders like several translation initiation factors, Heterogenous nuclear ribonucleoproteins and splicing factors cross-linked to RNA in this dataset.

### RNA-NM: Hsh49-poly(G)25 complex

Spliceosomal protein Hsh49 from *Saccharomyces cerevisiae* was reconstituted with a synthetic oligonucleotide comprised of 25 guanosines. Hsh49-poly(G)25 complex was cross-linked with 10 mM NM and cross-linked peptide- (oligo)nucleotides were analyzed by LC-MS/MS and acquired data was analyzed with NuXL in Proteome Discoverer 3.1. 11 different cross-linked Hsh49 peptides are identified at FDR 1 % ("fragment adducts included"). 10 of these peptides show a lysine missed cleavage which is in most cases the cross-link site.

### RNA-DEB: *Escherichia coli* S30 extract

*E. coli* BL21 (DE3) strain was cross-linked with 50 mM DEB in medium and S30 extract was prepared and cross-linked peptide- (oligo)nucleotides were analyzed by LC-MS/MS. Acquired data was analyzed with NuXL in Proteome Discoverer 3.1. The provided database contains all reviewed *E. coli* K12 protein entries from Swissprot (downloaded on 02/03/2021; 4,389 entries). 105 proteins are identified as cross-linked to RNA at an FDR of 1 % ("fragment adducts included") out of which 30 proteins are ribosomal proteins.