**Untargeted Feature Extraction**

The raw data (.raw files) may be found in the “raw data” directory. Note that the analysis in the paper is based on: C120171218\_MR002\_YckK\_AF\_inj1.raw, C120171218\_MR004\_DppA\_AF\_inj1.raw and C120171218\_MR006\_AF\_Buffer\_inj3.raw. The other files correspond to samples which were not considered further but are included here for the sake of completeness as they were all used in the MZmine analysis.

|  |  |
| --- | --- |
| File name | Description |
| C120171218\_MR001\_YckK\_AF\_NaCl\_inj1.raw | Analysis of the extract of the YckK protein provided in NaCl buffer. |
| **C120171218\_MR002\_YckK\_AF\_inj1.raw** | Analysis of the extract of the YckK protein provided in ammonium formate buffer. This was used as a control in the analysis in the paper. |
| C120171218\_MR003\_DppA\_AF\_NaCl\_inj1.raw | Analysis of the extract of the DppA protein provided in NaCl buffer. |
| **C120171218\_MR004\_DppA\_AF\_inj1.raw** | Analysis of the extract of the DppA protein provided in ammonium formate buffer. This was the sample of interest discussed in the paper. |
| C120171218\_MR005\_AF\_NaCl\_Buffer\_inj1.raw | A blank extraction of NaCl buffer |
| C120171218\_MR005\_AF\_NaCl\_Buffer\_inj2.raw | A blank extraction of NaCl buffer |
| C120171218\_MR006\_AF\_Buffer\_inj1.raw | A blank extraction of ammonium formate buffer |
| C120171218\_MR006\_AF\_Buffer\_inj2.raw | A blank extraction of ammonium formate buffer |
| **C120171218\_MR006\_AF\_Buffer\_inj3.raw** | A blank extraction of ammonium formate buffer |
| C120171218\_MR006\_AF\_Buffer\_inj4.raw | A blank extraction of ammonium formate buffer |
| C120171218\_MR006\_AF\_Buffer\_inj5.raw | A blank extraction of ammonium formate buffer |

Each data file (datafile.raw) was centroided and converted to mzXML format using MSConvert1 using the command:

msconvert datafile.raw --filter "peakPicking true 1-" –mzXML

The resulting .mzXML files were then imported into MZmine 2.262. Each file is then processed individually by sequentially applying the following modules/conditions as follows:

Mass detection module with the scans parameter set to MS level: 1, mass detector set to centroid with a noise level 1.0E5. Chromatogram builder module with scans parameter set to MS level: 1, Min time span (min) = 0.5, Min height = 1.0E7, m/z tolerance = 0.002 m/z or 2.0 ppm. Duplicate peak filter module with m/z tolerance = 0.002 m/z or 2.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = false. Isotopic peak greater module with m/z tolerance = 0.001 m/z or 1.0 ppm, retention time tolerance 0.2 absolute (min), monotonic shape = false, Maximum charge = 2, representative isotope = lowest m/z. Chromatogram deconvolution module with local minimum search algorithm with parameters; chromatographic threshold = 10%, Search minimum in RT range (min) 0.20, Minimum relative height = 10%, Minimum absolute height = 1.0E3, Min ratio of peak top/edge = 2, Peak duration range (min) = 0.50 – 50.00. Peak filter module with FWHM = 0.00 – 5.00 (all other parameters set to false). Peak list rows filter module with minimum peaks in a row = 1, all other parameters set to false.

The resulting peak lists from all samples in the experiment were then aligned using the RANSAC aligner module using the following conditions: m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), RT tolerance after correction = 0.5 absolute (min), RANSAC iterations = 1000, Minimum number of points = 50.0%, Threshold value = 0.2, Linear model = false, require same charge state = false. The combined peak list was then further processed using the same RT and m/z range gap filler module with m/z tolerance = 0.002 m/z or 5.0 ppm and then the duplicate peak filter module with m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), require same identification = false. These commands are encoded in the MZmine batch file: “20171218\_Exp1\_Pos\_UT\_5.0.xml” along with the MZmine project file “20171218\_Exp1\_Pos\_UT\_5.0.mzmine“ in the “Untargeted\_Analysis\_with\_MZmine” directory.

The resulting peak list was then exported as a .csv file “APL\_GF\_Filt.csv” which was opened in Excel for further visualization and processing. Specifically, we produced bubble plots in which m/z is given on the x-axis, retention time (min) on the y-axis and the bubble size is proportional to the chromatographic peak area. In order to identify those features which are uniquely associated with the protein of interest rather than small molecule background we excluded all features with peak areas less than ten times that of a corresponding control sample. See “20171218\_Exp1\_Pos.xlsx” in the Untargeted\_analysis\_with\_MZmine” directory.

**De novo peptide sequencing using PEAKS**

The .mzXML files generated following conversion with MSCovert (described above) were imported into Peaks Studio 64-bit v 7.5 (Bioinformatics Solutions Inc.) and processed using the default setting for Orbitrap (orbi-orbi). Initial data refinement was then performed with the following parameters: Precursor options = corrected, charge options = no correction, Filter Quality > 0.65, process = true and default = true. This was followed by de novo sequencing with the following instrument settings: Parent mass error tolerance = 10.0 ppm, fragment mass error tolerance = 0.01 Da, enzyme = none, report # peptides = 3, data refinement dependencies = 7. The resulting de novo identified peptides with ALC scores > 50 were then exported as “de novo peptides.csv files”. These can be found in the directory “PEAKS\_processed\_data” and then the subdirectory corresponding to each sample.

**Extraction of the chromatographic features corresponding to the peptides identified by PEAKS**

Extracted ion chromatograms for all putatively identified peptides from the PEAKS analysis were extracted in MZmine 2.26 for each sample. The following modules/parameters were used: After importing the .mzXML files into MZmine 2.26 the targeted peak detection module was used with the parameters, intensity tolerance = 50%, noise level = 1.0E3, m/z tolerance = 0.001 m/z or 1.0 ppm and retention time tolerance = 1.0 absolute (min). The peak list file with the series of m/z and RTs to search was generated from the PEAKS data by taking the parent ion m/z and RT from all MS/MS scans for which a sequence had been assigned. Duplicate peak filter module with m/z tolerance = 0.02 m/z or 10.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = false. Peak filter module with FWHM = 0.0 to 1.0, all other parameters = false. Peak list rows filter module with minimum peaks in a row = 1, all other parameters = false. Once this had been performed for each sample, the resulting peak lists were aligned using the RANSAC aligner module using the following conditions: m/z tolerance = 0.001 m/z or 1.0 ppm, RT tolerance = 1.0 absolute (min), RT tolerance after correction = 0.5 absolute (min), RANSAC iterations = 1000, Minimum number of points = 50.0%, Threshold value = 0.2, Linear model = false, require same charge state = false. The combined peak list was then further processed using the same RT and m/z range gap filler module with m/z tolerance = 0.002 m/z or 5.0 ppm and then the duplicate peak filter module with m/z tolerance = 0.005 m/z or 5.0 ppm, RT tolerance = 5.0 absolute (min), require same identification = true. See “PEAKS\_ALC\_50.xml” for the MZmine batch file. The resulting peak list was then exported as a .csv file, “APL\_GF\_Filt.csv” which was opened in Excel for further visualization and processing “IDs\_from\_PEAKS\_ALC\_50\_de\_novo.xlsx”. The corresponding MZmine project is saved as “PEAKS\_ALC\_50.mzmine” The files associated with this analysis are located in the PEAKS\_processed\_data\Targeted\_analysis\_of\_features\_IDed\_by\_PEAKS directory.

**References:**

1 Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012**, *30*, 918.

2 Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395.