Supplementary Material

Top-Down Garbage Collector: a tool for selecting high-quality top-down proteomics mass spectra

Diogo B Lima1&, André R F Silva2&,Mathieu Dupré1, Marlon D M Santos2, Milan A Clasen2, Louise U Kurt2, Priscila F Aquino3, Valmir C Barbosa4, Paulo C Carvalho2+, Julia Chamot-Rooke1+

1Mass Spectrometry for Biology Unit, CNRS USR 2000, Institut Pasteur, Paris, France 2Computational Mass Spectrometry & Proteomics Group, Carlos Chagas Institute, Fiocruz, Paraná, Brazil, 3Leônidas and Maria Deane Institute, Fiocruz, Amazonas, Brazil, 4Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Materials & Methods

Sample preparation

*Escherichia coli* K12 fresh cultures were harvested at late exponential growth phase. After cycles of wash and centrifugation, cell pellets were recovered in (80% H2O, 10% ACN, 10% FA). Cell lysis was performed by mechanical disruption using a high-speed homogenizer. After cell debris removal, bacterial lysate samples were transferred into protein LoBind tubes (Eppendorf) and stored at -80°C.

LC-MS/MS

A Dionex UltiMate 3000 RSLC Nano System coupled to an Orbitrap Fusion Lumos mass spectrometer fitted with a nano-electrospray ionization source (Thermo-Scientific) was used for all experiments. Five µL of protein sample in solvent A were loaded at a flow rate of 10 µL.min-1 onto an in-house packed C4 (5µm, Reprosil) trap column (0.150 mm i.d. x 30 mm) and separated at a flow rate of 0.5 µL.min-1 using a C4 (5 µm, Reprosil) column (0.075 mm i.d. x 600 mm). The following gradient was used: 2.5% B from 0–5 min; 15% B at 6.6 min.; 60% B at 124 min.; 99% B from 126–131 min.; and 2.5% B from 132–150 min. Solvent A consisted of (98% H2O, 2% ACN, 0.1% FA) and solvent B of (20% H2O, 80% ACN, 0.1% FA).

MS scans were acquired at 60,000 resolving power (at m/z 400) with a scan range set to 550–1,750 m/z, two microscans (µscans) per MS scan, an automatic gain control (AGC) target value of 5x105 and maximum injection time of 50 ms. MS/MS scans were acquired using the Data-Dependent Acquisition mode (Top 4) at 60,000 resolving power (at m/z 400) with an isolation width of 1.5 m/z, two µscans, an AGC target value of 5x105 and maximum injection time of 250 ms. For fragmentation, electron transfer dissociation with 10 ms of reaction injection time and a supplemental higher-energy collisional dissociation with normalized collision energy (NCE) of 10% (EThcD) was used.

Data analysis

All data were processed with ProSightPC v3.0 (Thermo-Scientific) and Proteome Discoverer v2.1 (Thermo-Scientific) using the ProSightPD 1.1 node. Spectral data were first deconvoluted and deisotoped using the cRAWler algorithm. Spectra were then searched using a Three Tier Search tree with searches against an *E. coli* (strain K12, taxon identifier 83333) protein database (Uniprot release 2017-11, XML version, 6,174 protein entries). The Uniprot database was created using the database manager application within ProSightPC v3.0. Potential initial methionine cleavage and N-terminal acetylation, as well as known modifications, were included, resulting in a database in a ProSight Warehouse File (.pwf) format containing 69,838 proteoforms. The first search consisted of a ProSight Absolute Mass search with MS1 tolerance of 2 Da and MS2 tolerance of 5 ppm. The second search was a ProSight Biomarker search with MS1 tolerance of 2 Da and MS2 tolerance of 5 ppm. The third search was a ProSight Absolute Mass search performed with MS1 tolerance of 10,000 Da and MS2 tolerance of 5 ppm. Identifications with E-values better than 1e-10 (–log (E-value) =10) were considered.