A database of mass spectrometric assays for the yeast proteome

Paola Picotti, Henry Lam, David Campbell, Eric Deutsch, Hamid Mirzaei, Jeff Ranish, Bruno Domon & Ruedi Aebersold

Supplementary text:

<table>
<thead>
<tr>
<th>Supplementary Methods</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Discussion</td>
<td>Construction of the MRMAAtlas and exportability of the optimal transitions</td>
</tr>
<tr>
<td>Supplementary Note</td>
<td>Tour of the MRMAAtlas user interface</td>
</tr>
</tbody>
</table>
Supplementary Methods

Sample preparation – The database was constructed from a collection of mass spectrometry analyses of different yeast samples. Wild-type *S. cerevisiae* cells were grown to different OD₆₀₀₅ (0.5-2.0) in glucose-based yeast extract peptone dextrose (YPED) medium, or in synthetic defined (SD) medium based on glucose or galactose or in complex yeast medium. Cells were disrupted by vortexing in the presence of acid-washed glass beads in a lysis buffer including 50 mM Hepes, pH 7.5, 5% glycerol, 15 mM dithiothreitol (DTT), 100 mM KCl, 5 mM EDTA, and a complete protease inhibitors cocktail (Roche, Mannheim, Germany). Alternatively a nuclear extract preparation was obtained as described at http://www.fhcrc.org/science/labs/hahn/methods/biochem_meth/polii_nuc_ext.html.

Yeast lysates were centrifuged to remove cellular debris, the supernatants were transferred to a fresh tube and the protein concentration in each extract was determined by Bradford assay. Proteins were precipitated by adding six volumes of cold (-20°C) acetone and resolubilized in a digestion buffer containing 8M urea and 0.1 M NH₄HCO₃. Proteins were reduced with 12 mM dithiothreitol for 30 min at 35°C and alkylated with 40 mM iodoacetamide for 45 min at 25°C, in the dark. Samples were diluted with 0.1 M NH₄HCO₃ to a final concentration of 1.5 M urea and sequencing grade porcine trypsin (Promega) was added to a final enzyme:substrate ratio of 1:100. Peptide mixtures were either directly destined to mass spectrometry analysis or firstly separated according to the isoelectric point of the peptides by off-gel electrofocusing (OGE) using a pH 3-10 IPG strip and 24 wells (Agilent Technologies). Peptides collected in each well as well as peptides from the unfractionated sample were cleaned by Sep-Pak tC18 cartridges (Waters, Milford, MA, USA) eluted with 60% acetonitrile. Peptides were evaporated on a vacuum centrifuge to dryness, resolubilized in 0.1% formic acid and immediately analyzed.

Development and validation of MRM assays. Samples were analyzed on a hybrid triple quadrupole/ion trap mass spectrometer (4000QTrap, ABI/MDS-Sciex, Toronto) equipped with a nanoelectrospray ion source. Chromatographic separations of peptides were performed on a Tempo nano LC system (Applied Biosystems) coupled to a 15 cm fused silica emitter, 75 μm diameter, packed with a Magic C18 AQ 5 μm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4°C) Tempo autosampler and separated with a linear gradient of acetonitrile/water, containing 0.1% formic acid, at a flow rate of 300 nL/min. A gradient from 5 to 30% acetonitrile in 30 or 60 minutes was used. The mass spectrometer was operated in multiple reaction monitoring mode, triggering acquisition of a full MS² spectrum upon detection of an MRM trace (threshold 300 ion counts). MRM acquisition was performed with Q1 and Q3 operated at unit resolution (0.7 m/z half maximum peak width) with an average of 60 transitions (dwelltime 20 ms/transition) per run. MS² spectra were acquired in enhanced product ion (EPI) mode for the two highest MRM transitions, using dynamic fill time, Q1 resolution low, scan speed 4000 amu/s, m/z range 300-1400, 2 scans summed. Collision energies used for both MRM and MS² analyses were calculated according to the formulas: CE = 0.044 * m/z + 5.5 and CE = 0.051 * m/z + 0.5 (CE, collision energy, m/z, mass-to-charge ratio of the precursor ion) for doubly and triply charge precursor ions.
respectively. Where available (for overall < 100 proteins), synthetic peptides or heavy-labelled synthetic peptides were used to extract optimal SRM transitions. All possible fragment ion masses of the targeted peptide were calculated and experimentally tested in SRM mode, analyzing the peptides by direct infusion in the QQQ mass spectrometer. Transitions resulting in the most intense signals were manually selected as optimal.

**Identification of MS2 spectra by sequence searching.** Raw MS2 .mzWiff data were converted to .mzXML format with the program mzWiff and searched against the yeast SGD database (version dated 01/26/2006) using Sequest (version 27). Human keratins and porcine trypsin were added to the protein database. A decoy database was generated by randomly reshuffling amino acids in between tryptic cleavage sites, and appended to the target database. Precursor mass tolerance was set at 2.1 Da. Data were searched allowing one non-tryptic terminus, with methionine oxidation as a variable modification and carboxamidomethylation of cysteine residues as a static modification. The search results were validated and assigned probabilities using the PeptideProphet program implemented in the Trans-Proteomic-Pipeline, with decoy-assisted semiparametric model and retention-time model enabled (Nesvizhskii et al., in preparation). Identifications were filtered to retain only those identifications with a PeptideProphet probability above 0.9, yielding a total of 38,804 identifications (7,529 distinct peptide ions and 6,698 distinct peptide sequences). Of these, 190 were decoy identifications. The spectrum-level false discovery rate of the retained set of non-decoy identifications predicted by PeptideProphet is 0.84%, and that estimated by decoy counting was 0.49% (190/38,614).

**Consensus spectral library creation -** A consensus spectral library was created from the confident identifications using the software SpectraST. Namely, whenever available, multiple observations of the same peptide ion were combined to form a consensus spectrum, averaging out experimental variations to yield a representative fragmentation pattern of the peptide ion. Spectral quality filters were applied to remove low-quality and potentially incorrectly identified spectra (Lam et al., in preparation). A total of 5,679 non-decoy and 59 decoy spectra remained after filtering. Based on decoy counting, the peptide-level false discovery rate of 1.0% (59/5,679) is expected for the set of non-decoy spectra. They cover a total of 1,234 proteins with distinct SGD identifiers, with an estimated protein-level false discovery rate of 4.8% (59/1,234). Alternatively, using the MRMAtlas web interface, the user is empowered to increase the initial PeptideProphet cutoff to control the false discovery rate more stringently as desired (e.g. PeptideProphet probability cutoffs of 0.95 or 0.99 correspond to a protein-level false discovery rate of 2.6% or 0.5%, respectively). For each spectrum, 8 peaks were selected as recommended SRM transitions from the most intense 20 peaks (fragments due to neutral loss from precursor excluded) based on the following preference: y ions > b ions > any neutral loss from y/b fragments. More intense peaks were favored in case of a tie in preference. Fragments with m/z values close to the precursor ion m/z (| m/z2 - m/z1 | ≤ 5 Th) were discarded, as such transitions result in high noise levels.
SRM transitions associated with the low abundant proteins contained in the atlas (< 150 copies/cell, according to Ghaemmaghammi et al.) and identified using yeast extracts, were confirmed by analyzing the synthetic analogues of the corresponding peptides.

**Construction of database** – The MRMAtlas is a derivative of the UniPep and PeptideAtlas databases\(^4,5\), and built upon the Systems Biology Experiment Analysis Management System (SBEAMS; [http://www.sbeams.org](http://www.sbeams.org)), where original data in .mzXML format\(^6\) are stored. This framework utilizes open-source programming tools such as perl, php, and mysql, as well as bioinformatics resources such as KEGG\(^7\), SGD ([http://www.yeastgenome.org/](http://www.yeastgenome.org/)), and Ensembl ([www.ensembl.org/](http://www.ensembl.org/)). The dataset was processed as described in Deutsch et al\(^8\) with some changes to ensure consistency between the build and the SRM assays extracted from the consensus library. Specifically, the final list of identified peptides was filtered against the consensus library, such that no peptide ions (sequence, charge, modifications) which do not have a representative consensus spectrum were retained in the Atlas build. In addition, the PeptideProphet probabilities for each peptide ion were not adjusted based on the ProteinProphet results, since this would have excluded a small number of peptides from the build which are present in the consensus library. The build was loaded into the backend database, and summary statistics generated. The extracted SRM assays were also uploaded and associated with the MRMAtlas, which is the publicly accessible ([www.mrmatlas.org](http://www.mrmatlas.org) or [www.srmatlas.org](http://www.srmatlas.org)) interface to the data and tools described herein. The data can be queried via the web-interface for peptides, individual proteins, protein sets, or cellular pathways.

**Protein property analysis** - We used yeast protein abundance datasets previously published by Ghaemmaghami et al\(^9\), and yeast protein properties were obtained from the SGD ([http://www.yeastgenome.org/](http://www.yeastgenome.org/)) and GO ([http://www.geneontology.org](http://www.geneontology.org)) databases, using the tool goSlimMapper ([http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl](http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl)).
Supplementary Discussion

Construction of the MRMAAtlas and exportability of the optimal transitions

When developing an SRM/MRM assay for a peptide of interest, the fragment ion masses of the target peptide can be calculated and experimentally tested in LC-SRM mode on a triple quadrupole instrument. This results in an overlay of SRM traces with different intensities, and with perfectly aligned intensity peak apexes in an elution time scale (Fig. 1S). The SRM transitions resulting in the most intense signal peaks can then be selected to be used as final assay for the peptide of interest, thus maximizing the sensitivity of the assay. However, if different ion series are taken into account, and more than one precursor charge state, this can result in more than 30 transitions per peptide to be tested. This high number of transitions, coupled to the typical SRM cycle time constraints, result in a limited number of peptides for which transitions are tested in a single LC-SRM analysis.

![Overlay of SRM traces corresponding to different transitions of a given peptide.](image)

**Figure 1S.** Overlay of SRM traces corresponding to different transitions of a given peptide.

Alternatively, MS/MS spectra for the target peptide can be acquired on a triple quadrupole instrument to derive optimal SRM transitions. This is ideally performed using only a couple of calculated transitions for the peptide of interest to trigger acquisition of the corresponding full tandem mass spectrum. This protocol is referred to as “SRM-triggered MS/MS scanning”. In such experiments, the MS is programmed to acquire a full fragment ion spectrum only when the signal of a target transition(s) is/are detected. The acquired MS2 spectra are then compared to the predicted peptide fragments to assure that the major MS2 peaks are matched, manually or by automatic database search. Although this protocol resembles a classical data dependent acquisition mode, the triggering by SRM traces instead of MS1 spectra increases specificity and thereby allows the identification of peptides of low abundance in complex mixtures which would not be detected in data dependent acquisition mode. In this workflow, the optimal SRM transitions, resulting in the most intense signals can be directly derived from the tandem mass spectrum acquired on the triple quadrupole-like mass spectrometer and under the same fragmentation conditions as the ones used in
SRM mode. This approach allows, in a single analysis, extraction of optimal SRM transitions and validation that the detected SRM signals actually derive from the targeted peptide.

In this work, we mostly took advantage of the latter approach to construct the MRMAAtlas. For each proteotypic peptide (either previously observed, or predicted) we derived one or a few calculated transitions. These transitions were used to trigger acquisition of full MS2 spectra of the peptide of interest, using tryptic digests of yeast samples grown under different biological conditions. SRM-triggered MS2 experiments were performed on the triple quadrupole mass spectrometer QTrap. Peptide fragmentation was performed in the Q2 collision cell, both when acquiring SRM traces and MS2 spectra, and under the same nominal fragmentation conditions (e.g. collision gas pressure and collision energy). Under this experimental setup, relative intensities of different SRM traces acquired in SRM mode perfectly match relative intensities of the associated fragment ions in the full MS2 spectrum of the peptide (Fig. 2S). This allowed us to confidently select optimal fragment ions to be used in the final SRM assays for the peptide of interest directly from the corresponding MS2 spectrum, derived from Q2-collision cell fragmentation in the QQQ-like MS. When available, multiple MS2 spectra of the same peptide were combined to form a consensus spectrum, to average out experimental variations and yield an even more representative fragmentation pattern of the peptide.

The need to acquire MS2 spectra for the validation of transitions can be obviated if synthetic peptides are used. In this work, where available (for overall < 100 proteins), synthetic peptides or heavy-labeled synthetic peptides were used to extract optimal SRM transitions. All possible fragment ion masses of the targeted peptide were calculated and experimentally tested in SRM mode directly on the QQQ instrument, analyzing the peptides by direct infusion in the QQQ mass spectrometer. Transitions resulting in the most intense signals were selected as optimal. Retention times were extracted from an LC-SRM analysis of the synthetic peptides.

Using the methods described above, optimal transitions for overall ~1500 yeast proteins were determined on the triple quadrupole mass spectrometer QTrap (Applied Biosystems, AB) and stored in the MRMAAtlas. To test the exportability of the suggested transitions, we selected a subset of 50 peptides stored in the MRMAAtlas and we analyzed the corresponding 6–10 SRM transitions recommended in the atlas on other brands of QQQ mass spectrometers, i.e. an Agilent QQQ (Agilent Technologies) and a TSQ Quantum (Thermo Fischer Scientific). The transitions appeared to be extremely exportable to other QQQ instruments (see example in Fig. 3S). In other words, the relative intensity of SRM transitions corresponding to different fragments for the same peptide was strikingly well conserved across different QQQ MS. This indicates that the optimal transitions presented in the MRMAAtlas can be exported to another QQQ and used directly to quantify the proteins of interest in SRM mode, without additional method development, other than a realignment of the suggested collision energies to the local ranges (e.g. in our hands, CE \text{Agilent-QQQ} = \text{CE } \text{AB-QTrap} – 10 \text{ V}). This reinforces the concept of storing validated, optimal SRM transitions in centralized databases such as the MRMAAtlas. The use of the transitions suggested in the MRMAAtlas on mass spectrometers other than QQQs (e.g. on ion traps operated in pseudo-SRM mode) remains instead still to be explored.
Figure 25. Similarity between SRM traces relative intensities and fragment ion relative intensities as derived from MS2 spectra, measured on a QQQ-type instrument operated in SRM-triggered MS2 mode.
Figure 3S. Example of exportability of SRM transitions. Relative intensities of SRM transitions for two peptides as measured on different QQQ mass spectrometers (top, AppliedBiosystems QTrap; middle, Agilent Technologies QQQ; bottom, Thermo Fischer TSQ Quantum). The plots show the normalized intensities of SRM transitions for different fragment ions versus the m/z value of the fragment used as Q3 value. The m/z ratios of the doubly charged peptides are used as Q1 values.
Supplementary Note

Tour of the MRMAAtlas user interface

The MRMAAtlas can be accessed via the websites www.mrmatlas.org or www.srmatlas.org. With the following screenshots, a quick tour of the MRMAAtlas user interface is presented.

The entry page (Fig. 4S) introduces the MRMAAtlas project and provides links to the main tools and to the background information (Fig. 5S). The MRMAAtlas interface features several query options. Users can browse for peptides, single proteins, or protein lists, via the “MRM Transitions” query interface (Fig. 6S) and retrieve the corresponding SRM/MRM assays (Fig. 7S and 8S). Each assay is presented as a set of eight selected, optimal SRM/MRM coordinates for the peptides that represent a protein. The annotated information consists of the following values: m/z of the precursor peptide ion, charge state, m/z of the fragment ion, fragment relative intensities, calculated hydrophobicity and observed elution times, collision energy and MS type on which the transitions were measured. This information can be downloaded in a spreadsheet format. For each peptide, links are provided to the consensus spectrum and to each individual tandem mass spectrum of the peptide stored in the database and from which the consensus spectrum was constructed. Dropdown help menus provide explanation of the abbreviations used in the MRMAAtlas web pages. The protein-centric view (Fig. 9S) provides an overview of the set of peptides from the protein of interest for which an SRM/MRM assay is available.

The MRMAAtlas can also be queried for cellular pathways. The link “Pathway search” displays all cellular pathways as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database4 for the species of interest. Upon selection of a pathway of interest, a schematic structure of the pathway is displayed (Fig. 10S) which highlights proteins for which SRM/MRM assays have been developed.

![Figure 4S. Entry page of the MRMAAtlas.](image-url)
Figure 5S. Background information: web page of the MRMAAtlas explaining the development of targeted proteomic assays based on MRM/SRM.
**Figure 6S.** “MRM Transitions” query interface. In the example shown, the user browses the MRMAtlas for the *S. cerevisiae* proteins YOR230W, YOR096W, YKL060C, to retrieve the corresponding SRM assays.
**Figure 7S.** SRM assays are retrieved for the peptide ILELVFPTEIVGK. The table with the final coordinates of the SRM assay can be downloaded in a spreadsheet format. Links to the consensus and individual MS2 spectra of the peptide are provided.
Figure 8S. SRM assays are retrieved for the peptide AEEAATTDLTYR.
Figure 9S. Example of protein-centric view for the *S. cerevisiae* protein YKL060C.
Figure 10S. Example of “Pathway Search” for the biological pathway “pyruvate metabolism”. Proteins for which SRM assays are available are displayed in yellow/green. Proteins for which no peptides were observed are highlighted in light blue.
References


Note

a. Multiple SRM transitions can be measured within the same experiment by rapidly toggling between the different transitions. The term multiple reaction monitoring (MRM) is frequently used to describe such parallel acquisition of SRM transitions, but might be in the future deprecated by the IUPAC nomenclature (Murray et al., IUPAC Current Provisional Recommendations, August 2006, prepared for publication).