Published in final edited form as:

Nat Methods. 2008 November ; 5(11): 913–914. doi:10.1038/nmeth1108-913.

A database of validated assays for the targeted mass spectrometric analysis of the *S. cerevisiae* proteome

Paola Picotti^{1,#}, Henry Lam^{3,#}, David Campbell^{3,#}, Eric Deutsch³, Hamid Mirzaei³, Jeff Ranish³, Bruno Domon¹, and Ruedi Aebersold^{1,2,3,*}

¹ Institute of Molecular Systems Biology, ETH Zurich, CH-8093 Zurich, Switzerland ² Faculty of Science, University of Zurich, CH-8006 Zurich, Switzerland ³ Institute for Systems Biology, Seattle, WA 98103, USA

To the editor

The currently most widely used mass spectrometry-based proteomic methods sample the available proteome in a quasi-random manner¹. In each analysis of a sample only a subset of the proteins it contains are identified and quantified, and repeated analyses of the same sample measure only partly overlapping segments of the proteome. This precludes the generation of consistent and reproducible datasets when the effects of different perturbations on a proteome are studied, as is the case of dosage series or time courses. Complete, quantitatively accurate data sets are however of critical importance for many studies, especially those aimed at generating data to support the mathematical modeling of a biological process, a hallmark of the emerging field of systems biology. An additional challenge in comprehensive proteomic analyses is the detection of low abundant proteins^{2, 3}. These constraints strongly limit the feasibility of quantitatively and consistently measuring defined sets of target proteins -such as functionally related proteins, proteins constituting signaling networks, or proteins involved in a specific metabolic cycle - across different samples. To overcome these limitations we recently proposed a targeted proteomic strategy^{2, 4}, which exploits the power of a mass spectrometry (MS) technique called selected reaction monitoring (SRM). The essence of this approach is the generation of specific, quantitative mass spectrometric assays for each member of a set of proteins and their subsequent application to multiple biological samples. The approach requires the generation of a list of proteins of interest for which peptides are selected that unambiguously represent these proteins and are preferentially detectable by MS. We have termed such peptides proteotypic peptides (PTP's)⁵. Next, for each PTP, precursor ion/fragment ion relationships are established, that specifically identify the respective PTP. These consists of pairs of mass-to-charge (*m/z*) values that are selected with the first and last analyzer of a triple quadrupole (QQQ)-like mass spectrometer to isolate the targeted precursor ion and corresponding, diagnostic fragment ion(s), respectively. The detector acts as a counting device for analytes matching the defined relationship(s) and returns a signal intensity over the chromatographic elution time. These relationships, commonly termed (SRM or MRM³) transitions, therefore effectively constitute mass spectrometric assays that identify a specific peptide and, by inference, the corresponding protein in a complex protein digest. These assays are accurately

*Corresponding author: aebersold@imsb.biol.ethz.ch; Phone: +41 44 633 31 70; Fax: +41 44 633 10 51.

#Authors equally contributing

³Multiple 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).

quantitative, in particular if stable isotope labeled reference peptides are used^{6, 7}. These assays are then applied to confirm the presence and determine the (absolute) quantity of the targeted proteins in biological samples. The favorable duty cycle and the two-level mass filtering of the SRM method result in a higher sensitivity - low attomole level⁸ - and specificity compared to other mass spectrometry-based proteomic techniques. Recently, we have shown that proteins spanning a range of abundance between 1.3 million down to less than 50 copies per cell could be detected and quantified by SRM in *S. cerevisiae* whole proteome digests⁹. However, despite these favourable characteristics SRM has not yet been broadly applied in proteomics. One main reason for this is the effort required to establish a validated SRM assay for every protein. This process consists of a number of steps: 1. PTP's for each target protein need to be selected, a task that is complicated by MS signal responses that vary greatly for different peptides generated from the same target protein, and by the observation that many peptides are common to multiple proteins and thus lack the specificity to identify the targeted protein. 2. Predominant fragment ions specific to the peptide of interest are selected, and define the SRM transitions. Since most fragment ion spectra are generated on instruments different than QQQs, and instrument operating conditions are usually poorly documented, the commonly accessible fragment ion spectra can at best be considered a starting point for establishing the most suitable transitions. 3. These transitions need to be validated either using co-elution with a reference peptide or by acquiring a full MS2 spectrum of the peptide in the QQQ-like instrument used for SRM. 4. Final "coordinates" of the SRM assay, such as precursor and fragment ion masses with associated charge states resulting in the strongest SRM signals and chromatographic elution time, are then compiled to be used in the actual SRM measurement. Overall, this is a lengthy and iterative process. In the absence of prior knowledge about the behaviour of a peptide with respect to precursor ion response and fragmentation patterns in LC-MS/MS experiments, there is a high potential for substandard or misleading results. However, once an SRM assay for a protein is established, it becomes universally useful, i.e. such a tedious assay development process needs to be performed only once, for a specific type of mass spectrometer and fragmentation (e.g. Q2-collision cell fragmentation).

Here we present the first database of validated SRM assays for ~1500 yeast proteins. It was constructed by merging the results of more than 650 SRM-triggered MS2 analyses of *S. cerevisiae* protein digests, carried out on a triple quadrupole-type mass spectrometer. 1324 proteins are represented by assays for at least one of their PTP's. The resource also contains assays for a small number of peptides common to a maximum of two proteins. All peptides were selected because they show intense signal response by electrospray ionization mass spectrometry. Each peptide identification was validated by collecting a full tandem mass spectrum of the peptide in the QQQ-like mass spectrometer also used for SRM measurements. A detailed description of the methods used to construct the database is provided in Supplementary Methods and Discussion online.

Each assay is presented as a set of selected, optimal and validated SRM coordinates for the peptide(s) 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 ions with the highest signal intensities, calculated hydrophobicity and observed elution times, both as a measure of relative retention time in C₁₈ chromatography, suggested collision energy and relative intensities of fragment ion signals. The type of QQQ mass spectrometer on which the transitions were measured is also indicated. The SRM coordinates can be downloaded in a spreadsheet format which can be directly pasted into a SRM/MRM method of a triple quadrupole instrument and used to specifically detect and quantify the protein of interest in a complex protein digest. The intensity ratios of transitions for a specific peptide can optionally be used to further validate the specificity of the assay in the biological background of interest. Retention time constraints can be used for time-scheduled SRM acquisition⁸ which relieves SRM cycle-time constraints

and allows the acquisition of more than 1000 SRM transitions in a single LC-SRM analysis without compromising on sensitivity.

The database is at present the largest resource of validated SRM/MRM assays of any organism. It currently contains assays for 22% of the yeast proteome and the coverage is expected to rapidly increase since more SRM datasets are generated and will be submitted. The database contains assays for yeast proteins involved in all biological processes, as defined by gene ontology (GO) nomenclature (Fig. 1). In addition, proteins spanning all ranges of abundance in yeast are present in the dataset, down to a concentration below 50 molecules/cell¹⁰ (Fig. 2). Overall, this indicates that the database supports the quantitative analyses of cellular networks and processes in *S. cerevisiae*, even when they contain or consist of very low abundance proteins. The dataset was uploaded to MRMatlas (www.mrmatlas.org or www.srmatlas.org, see Supplementary Tutorial online) which is a publicly accessible instance of the PeptideAtlas project¹¹ (www.peptideatlas.org) and can be queried via the web-interface for peptides, individual proteins, protein sets, or cellular pathways.

The data and informatics resources we describe here offer the unique possibility of retrieving targeted proteomic assays for yeast proteins to be directly applied to the high-throughput, quantitative analysis of deliberately chosen protein sets and samples of interest, without the need of additional method development. This resource and its further developments, including the extension to other species, will open new possibilities to study and model biological processes, specifically in cases that require, complete, reproducible and accurate quantitative datasets. The resource we describe is therefore expected to find wide application in quantitative proteomics and to broadly impact systems biology.

Acknowledgments

We acknowledge Chee-Hong Wong, Lukas Reiter and Vinzenz Lange for valuable assistance and insightful discussions. We also thank Roeland Costenoble and Anne Kümmel. This project has been funded in part by ETH Zurich, the Swiss National Science Foundation, with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract No. N01-HV-28179 and by SystemsX.ch the Swiss initiative for systems biology. PP is the recipient of a Marie Curie Intra-European fellowship.

References

1. Aebersold R, Mann M. Nature 2003;422:198–207. [PubMed: 12634793]
2. Picotti P, Aebersold R, Domon B. Mol Cell Proteomics 2007;6:1589–1598. [PubMed: 17533221]
3. de Godoy LM, et al. Genome Biol 2006;7:R50. [PubMed: 16784548]
4. Domon B, Aebersold R. Science 2006;312:212–217. [PubMed: 16614208]
5. Kuster B, Schirle M, Mallick P, Aebersold R. Nat Rev Mol Cell Biol 2005;6:577–583. [PubMed: 15957003]
6. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 2003;100:6940–6945. [PubMed: 12771378]
7. Kuhn E, et al. Proteomics 2004;4:1175–1186. [PubMed: 15048997]
8. Stahl-Zeng J, et al. Mol Cell Proteomics 2007;6:1809–1817. [PubMed: 17644760]
9. Picotti, P.; King, N.; Domon, B.; Aebersold, R. Proceedings of the 54th ASMS Conference on Mass Spectrometry; June 3–7; Indianapolis, Indiana, USA. Santa Fe, NM: MPZ8, Society for Mass Spectrometry; 2007.
10. Ghaemmaghami S, et al. Nature 2003;425:737–741. [PubMed: 14562106]
11. Deutsch EW, et al. Proteomics 2005;5:3497–3500. [PubMed: 16052627]

Figure 1. Functional categorization of the proteins for which targeted proteomic assays are available in the MRMatlas

Grouping of ORFs into biological processes in *S. cerevisiae* was obtained from the SGD and GO databases. The same ORF can map to more than one GO biological process.

Figure 2. Distribution of *S. cerevisiae* protein abundances in the MRMAAtlas
 Protein abundances are derived from the dataset published by Ghaemmaghammi et al (2003)
 10.