**Supporting Information**

Tables of dNSAF values and NSAF\_adj(dNSAF) values from the 20S proteasome experiments generated from MSpC (Supplemental Table 1) and ABACUS (Supplemental Table 2) are available free of charge via the Internet. Thermo .raw data, FASTA databases, and configuration files from the experiments involving the 20S proteasome were deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with ID PXD003002. Reviewers can access the data by using [reviewer33719@ebi.ac.uk](mailto:reviewer33719@ebi.ac.uk) as username and 6i2FKdy5 as a password.

**Supplemental Methods**

Sample Preparation

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. 20S proteasomes were obtained as described previously \cite{book10} from *Arabidopsis thaliana* Col-0 ecotype seedlings in which the 20S proteasome subunit PAG1 (α7) was genetically replaced with a FLAG-tagged variant, with the minor modification of switching to a more stable HEPES buffer during purification. The FLAG peptide used for elution was removed by filtering through an Amicon Ultra 4 10K filter with the elution buffer also exchanged into 8 M urea. Total protein was quantified by the bicinchoninic acid protein assay (Thermo Scientific) using bovine serum albumin as the standard. Approximately 70 µg of proteasomes were digested overnight at 37 °C using a 1:30 trypsin/sample ratio. Peptides were acidified to a final concentration of 1% TFA, desalted on a Waters C18 Sep-Pak containing 50 mg sorbent material, and lyophilized. Total *Escherichia coli* lysates were obtained from Bio-Rad (Cat. 163-2110) with 200 µg digested as above. Both proteasome and *E. coli* peptides were dissolved in 5% acetonitrile, 95% water, and 0.1% formic acid. Each MS analysis, performed in triplicate, used 5 µL volumes prepared with 3, 2, 1.5, 1, 0.5, 0.25, 0.1, or 0.05 µg of digested proteasomes mixed with 0.5 µg of digested *E. coli* proteins. This mixtures reflected proteasome/*E. coli* ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, and 0.857, respectively.

Liquid Chromatography and High-Resolution Mass Spectrometry

Samples were analyzed by ultra-high performance liquid chromatography (UPLC) (nanoAcquity, Waters Corporation) connected online to an electrospray ionization LTQ Velos Orbitrap mass spectrometer (Thermo Scientific). Separation employed a 100 x 365 µm fused silica capillary micro-column packed with 20 cm of 1.7 µm-diameter, 130-Å pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 µm using a laser puller (Sutter Instruments). Peptides were loaded at a flow-rate of 400 nL/min for 30 min and then eluted over 120 min at a flow-rate of 300 nL/min with a 2% to 30% acetonitrile gradient in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap with a mass range of 300-1500 *m/z* at a resolution of 60,000, followed by ten MS/MS high energy C-trap dissociation scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution, with a mass range starting at 100 *m/z*. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 sec and an exclusion window of 120 sec.

Data Processing

Protein identifications were determined using the Morpheus search engine \cite{wenger13}. Raw data was searched with the Thermo module of Morpheus revision 151 downloaded and compiled from source code available at: <http://sourceforge.net/projects/morpheus-ms/> using Microsoft Visual Studio 2013 professional edition. The following parameters were used to search all databases: unknown precursor charge states - +2, +3, +4; maximum number of MS/MS peaks = 400; assign charge states - enabled; de-isotope - disabled; generate target decoy database on the fly; protease trypsin (no proline rule); maximum missed cleavages = 2; initiator methionine behavior - variable; fixed modification of carbamidomethylation of cysteine; variable modification of oxidation of methionine; maximum variable modification isoforms per peptide = 1024; precursor mass tolerance = +/- 2.1 Da monoisotopic (recommended parameters to account for neutral loss); precursor monoisotopic peak correction - disabled; product mass tolerance = +/- 0.025 Da monoisotopic; consider modified forms as unique peptides - false; maximum threads = 8; minimize memory usage - false.

MSpC quantification of Universal Proteome Standard 2 (UPS2) protein sequences exploited the MS/MS analysis of individual USP2 proteins (Sigma-Aldrich) mixed at various concentrations with egg or embryo extracts from *Xenopus* *laevis* available in thePRoteomics IDEntifications (PRIDE) repository \cite{vizcaino13} using identifier - PXD000902 and the available proteomics database - pita\_v1.71.protein.name.fa. For our analysis, both the raw MS/MS data and resulting FASTA files for the USP2 and egg and embryo proteomes were obtained from PRIDE. The database used for searching the MS/MS data of *Arabidopsis* proteasomes spiked into total *E. coli* peptides was generated by combining Uniprot K12 *E. coli* reference proteome UP000000625 with a common contaminant database, and then mixing the merged dataset with FASTA sequences for all proteoforms of all known proteasome subunits and associated proteins \cite{book10} obtained from the TAIR10\_pep\_20101214 FASTA database available within The *Arabidopsis* Information Resource (TAIR) version 10. All FASTA files are available for download in the Supporting Information. The datasets were analyzed by MSpC with a 1% PSM false discovery rate (FDR) and a 1% protein group FDR to determine NSAF, dNSAF, and uNSAF values for each protein group. Two separate analyses were performed in which one unique, or alternatively two unique peptides were required to quantify a protein group. NSAF values were calculated according to formulas 1a, 2a, and 3a from Figure 1 of Zhang *et al.* \cite{zhang10}.

For the individual subunit analysis of the 20S proteasome the isoform incorporation rates were treated as follows. Given that each proteasome subunit should be incorporated at equal stoichiometry within the 20S particle, we then tested whether the Morpheus/MSpC pipeline could calculate the relative abundance of each subunit and the distribution of isoforms. Here, we divided the dNSAF values for each subunit/isoform by the total number of dNSAF values for the entire complex across all eight total proteasome/*E. coli* lysate ratios tested. This averaged value provided a concentration-independent ratio for the incorporation of each subunit/isoform. We then normalized these values based on a 1/14 stoichiometry of each subunit within the complex to calculate the estimated occupancy versus the expected occupancy of each subunit.

Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABACUS Pipeline

The speed and accuracy of MSpC combined with Morpheus was compared to the next most comparable open source software suite for calculating NSAF values; *i.e.,* TPP \cite{deutsch10} combined with ABACUS \cite{fermin11}, using the proteasome spike-in experiment files as input. The .raw files were converted to .mzmL files by TPP Build 201411201551-6764 and then searched using the multi-threaded X!Tandem MS/MS search engine \cite{craig04} with the search parameters adjusted as close as possible to that used for the Morpheus searches. A decoy database was generated using the TPP tool DecoyFASTA for use with X!Tandem. Relevant X!Tandem parameters are listed here: parent monoisotopic mass error = +/- 2.1 Da, fragment mass error = +/- 0.025 Da, fixed modifications of carbamidomethylation (57.021464) on cysteine, and variable modification of oxidation (15.994915) on methionine, fully tryptic cleavages, missed cleavage sites = 2 maximum, no refinement and 8 threads. The configuration file used and the test datasets can be found in the Supporting Information. The data were analyzed in the TPP using Peptide Prophet \cite{keller02}. Relevant settings are listed: minimum probability = 0.05; minimum peptide length = 7; accurate mass, and nonparametric decoy database to pin down false discovery rate; ignore +1 charged spectra; and run Protein Prophet \cite{nesvizhskii03} after Peptide Prophet.

Once completed, all pepxml data from Peptide Prophet contained in a single folder was combined using the command line version of Protein Prophet from the TPP binaries with the following command: ProteinProphet.exe \*.pep.xml interact-COMBINED.prot.xml. This post-analysis aggregation was required for running the spectral counting program ABACUS. Here, we note that there are no graphic user interfaces to perform this post analysis aggregation, which makes this portion of the data analysis more difficult for those unfamiliar with setting up and running programs from the command line. The combined data was analyzed by ABACUS with the following parameters: best peptide probability = 0.99; minimum peptide probability = 0.99; experimental peptide probability = 0; and combined file probability = 0.99 to most accurately match a 1% FDR stringency settings in MSpC. dNSAF values were compared in Microsoft Excel using the CORREL function and squaring the result. Additional timing tests were performed with a subset of the calibration curve data (ratios 0.091, 0.167, 0.333, and 0.500 in triplicate corresponding to 12 .raw files) by increasing file input to 24 (2x), 48 (4x), and 96 (8x) .raw files to determine the time dependence on input size between both pipelines tested (Morpheus/MSpC versus TPP/ABACUS). The timing tests and all data analyses were performed on a computer running Windows 7 Ultimate, with 16 GB of random access memory, and an Intel Core i7-2700k with hyper-threading turned on for eight logical cores.

**Supplemental Discussion (Requiring Two Unique Peptides to Quantify a Protein)**

Occasionally, some researchers may want to use a more stringent criterion for quantification such as requiring a protein to have more than one unique identifying peptide. To see how this might affect our data analysis, we re-analyzed our results shown in Figure 1, this time requiring two unique peptides to quantify an individual protein. The results point to a very small increase in linearity observed in the average plots; however, there is a slight decrease in linearity for the egg sample (0.886 to 0.865) and a larger decrease in linearity for the individual UPS2 protein plot for the embryo sample (0.827 to 0.723). The decrease in linearity in the embryo sample is due to the analysis removing a low abundance UPS2 protein (O00762ups) identified with only one unique peptide. While some have suggested that requiring more than one unique peptide to identify a protein is an ideal approach, requiring two peptides for identifications in database searches reduces the number of protein identifications in the target database more than those in the decoy database and results in increased false discovery rates . While we recognize that researchers may want to implement more stringent requirements than what is typically used in database searching to quantify a set of proteins, there are two cases where requiring two unique peptides may not be ideal in a quantitative analysis. Firstly, low abundance proteins that have few PSMs might be identified by only a single peptide and thus be erroneously thrown out of the analysis. Secondly, there may be only one unique peptide that can differentiate between families of homologous proteins. In this second case, requiring two unique peptides would remove these homologous proteins from the MSpC analysis, even if they had a large number of PSMs. Because of these reasons and because of the decreased linearity observed when requiring proteins to have two unique peptides (Supplemental Figure 2A) as compared to one unique peptide (Figure 1A), we suggest caution in requiring more than one unique peptide per protein.

**SUPPORTING INFORMATION REFERENCES**

@article{book10,

author = {Book, A. J. and Gladman, N. P. and Lee, S. S. and Scalf, M. and Smith, L. M. and Vierstra, R. D.},

title = {Affinity purification of the Arabidopsis 26S proteasome reveals a diverse array of plant proteolytic complexes},

journal = {J. Biol. Chem.},

volume = {285},

number = {33},

pages = {25554-25569},

keywords = {Arabidopsis/\*metabolism

Arabidopsis Proteins/chemistry/\*metabolism

Chromatography, High Pressure Liquid

Electrophoresis, Polyacrylamide Gel

Mass Spectrometry

Proteasome Endopeptidase Complex/chemistry/\*metabolism

Spectrometry, Mass, Electrospray Ionization

Tandem Mass Spectrometry},

ISSN = {1083-351X (Electronic)

0021-9258 (Linking)},

DOI = {10.1074/jbc.M110.136622},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/20516081}](http://www.ncbi.nlm.nih.gov/pubmed/20516081%7d),

year = {2010},

type = {Journal Article}

}

@article{craig04,

author = {Craig, R. and Beavis, R. C.},

title = {TANDEM: matching proteins with tandem mass spectra},

journal = {Bioinformatics},

volume = {20},

number = {9},

pages = {1466-1467},

keywords = {\*Algorithms

Databases, Protein

Information Storage and Retrieval/\*methods

Mass Spectrometry/\*methods

Proteins/\*analysis/\*chemistry

Sequence Alignment/methods

Sequence Analysis, Protein/\*methods

Sequence Homology, Amino Acid

\*Software},

ISSN = {1367-4803 (Print)

1367-4803 (Linking)},

DOI = {10.1093/bioinformatics/bth092},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/14976030}](http://www.ncbi.nlm.nih.gov/pubmed/14976030%7d),

year = {2004},

type = {Journal Article}

}

@article{deutsch10,

author = {Deutsch, E. W. and Mendoza, L. and Shteynberg, D. and Farrah, T. and Lam, H. and Tasman, N. and Sun, Z. and Nilsson, E. and Pratt, B. and Prazen, B. and Eng, J. K. and Martin, D. B. and Nesvizhskii, A. I. and Aebersold, R.},

title = {A guided tour of the Trans-Proteomic Pipeline},

journal = {Proteomics},

volume = {10},

number = {6},

pages = {1150-1159},

keywords = {Computational Biology

\*Databases, Protein

Information Storage and Retrieval

Isotope Labeling

Proteomics/\*methods

Sequence Analysis, Protein

\*Software

Tandem Mass Spectrometry},

ISSN = {1615-9861 (Electronic)

1615-9853 (Linking)},

DOI = {10.1002/pmic.200900375},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/20101611}](http://www.ncbi.nlm.nih.gov/pubmed/20101611%7d),

year = {2010},

type = {Journal Article}

}

@article{fermin11,

author = {Fermin, D. and Basrur, V. and Yocum, A. K. and Nesvizhskii, A. I.},

title = {ABACUS: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis},

journal = {Proteomics},

volume = {11},

number = {7},

pages = {1340-1345},

keywords = {Automation, Laboratory

Computational Biology/\*methods

Databases, Protein

Humans

Proteins/\*analysis

Proteomics/\*methods

Sequence Analysis, Protein

Software

Tandem Mass Spectrometry},

ISSN = {1615-9861 (Electronic)

1615-9853 (Linking)},

DOI = {10.1002/pmic.201000650},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/21360675}](http://www.ncbi.nlm.nih.gov/pubmed/21360675%7d),

year = {2011},

type = {Journal Article}

}

@article{

author = {Gupta, N. and Pevzner, P. A.},

title = {False discovery rates of protein identifications: a strike against the two-peptide rule},

journal = {J Proteome Res},

volume = {8},

number = {9},

pages = {4173-81},

keywords = {Algorithms

Cluster Analysis

Databases, Protein

Mass Spectrometry

Peptides/\*chemistry

Proteins/\*chemistry

Proteomics/\*methods

Terminology as Topic},

ISSN = {1535-3893 (Print)

1535-3893 (Linking)},

DOI = {10.1021/pr9004794},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/19627159}](http://www.ncbi.nlm.nih.gov/pubmed/19627159%7d),

year = {2009},

type = {Journal Article}

}

@article{keller02,

author = {Keller, A. and Nesvizhskii, A. I. and Kolker, E. and Aebersold, R.},

title = {Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search},

journal = {Anal. Chem.},

volume = {74},

number = {20},

pages = {5383-5392},

keywords = {Algorithms

Databases, Factual

Mass Spectrometry

Models, Statistical

Peptides/\*analysis},

ISSN = {0003-2700 (Print)

0003-2700 (Linking)},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/12403597}](http://www.ncbi.nlm.nih.gov/pubmed/12403597%7d),

year = {2002},

type = {Journal Article}

}

@article{nesvizhskii03,

author = {Nesvizhskii, A. I. and Keller, A. and Kolker, E. and Aebersold, R.},

title = {A statistical model for identifying proteins by tandem mass spectrometry},

journal = {Anal. Chem.},

volume = {75},

number = {17},

pages = {4646-4658},

keywords = {Amino Acid Sequence

Humans

Mass Spectrometry/\*methods

\*Models, Statistical

Molecular Sequence Data

Peptides/analysis/chemistry

Proteins/\*analysis/\*chemistry},

ISSN = {0003-2700 (Print)

0003-2700 (Linking)},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/14632076}](http://www.ncbi.nlm.nih.gov/pubmed/14632076%7d),

year = {2003},

type = {Journal Article}

}

@article{vizcaino13,

author = {Vizcaino, J. A. and Cote, R. G. and Csordas, A. and Dianes, J. A. and Fabregat, A. and Foster, J. M. and Griss, J. and Alpi, E. and Birim, M. and Contell, J. and O'Kelly, G. and Schoenegger, A. and Ovelleiro, D. and Perez-Riverol, Y. and Reisinger, F. and Rios, D. and Wang, R. and Hermjakob, H.},

title = {The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013},

journal = {Nucleic Acids Res.},

volume = {41},

number = {Database issue},

pages = {D1063-9},

note = {Vizcaino, Juan Antonio

Cote, Richard G

Csordas, Attila

Dianes, Jose A

Fabregat, Antonio

Foster, Joseph M

Griss, Johannes

Alpi, Emanuele

Birim, Melih

Contell, Javier

O'Kelly, Gavin

Schoenegger, Andreas

Ovelleiro, David

Perez-Riverol, Yasset

Reisinger, Florian

Rios, Daniel

Wang, Rui

Hermjakob, Henning

eng

BB/I024204/1/Biotechnology and Biological Sciences Research Council/United Kingdom

WT085949MA/Wellcome Trust/United Kingdom

Research Support, Non-U.S. Gov't

England

2012/12/04 06:00

Nucleic Acids Res. 2013 Jan;41(Database issue):D1063-9. doi: 10.1093/nar/gks1262. Epub 2012 Nov 29.},

abstract = {The PRoteomics IDEntifications (PRIDE, <http://www.ebi.ac.uk/pride>) database at the European Bioinformatics Institute is one of the most prominent data repositories of mass spectrometry (MS)-based proteomics data. Here, we summarize recent developments in the PRIDE database and related tools. First, we provide up-to-date statistics in data content, splitting the figures by groups of organisms and species, including peptide and protein identifications, and post-translational modifications. We then describe the tools that are part of the PRIDE submission pipeline, especially the recently developed PRIDE Converter 2 (new submission tool) and PRIDE Inspector (visualization and analysis tool). We also give an update about the integration of PRIDE with other MS proteomics resources in the context of the ProteomeXchange consortium. Finally, we briefly review the quality control efforts that are ongoing at present and outline our future plans.},

keywords = {\*Databases, Protein

Internet

Mass Spectrometry

Peptides/chemistry/metabolism

Proteins/chemistry/metabolism

\*Proteomics

Software},

ISSN = {1362-4962 (Electronic)

0305-1048 (Linking)},

DOI = {10.1093/nar/gks1262},

url = {[http://www.ncbi.nlm.nih.gov/pubmed/23203882}](http://www.ncbi.nlm.nih.gov/pubmed/23203882%7d),

year = {2013},

type = {Journal Article}

}

@article{wenger13,

author = {Wenger, C. D. and Coon, J. J.},

title = {A proteomics search algorithm specifically designed for high-resolution tandem mass spectra},

journal = {J. Proteome Res.},

volume = {12},

number = {3},

pages = {1377-1386},

keywords = {\*Algorithms

Chromatography, Liquid

Escherichia coli Proteins/isolation & purification/metabolism

Humans

\*Proteomics

Saccharomyces cerevisiae Proteins/isolation & purification/metabolism

Tandem Mass Spectrometry/\*methods},

ISSN = {1535-3907 (Electronic)

1535-3893 (Linking)},

DOI = {10.1021/pr301024c},

url = {<http://www.ncbi.nlm.nih.gov/pubmed/23323968>

[http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3586292/pdf/nihms439262.pdf}](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3586292/pdf/nihms439262.pdf%7d),

year = {2013},

type = {Journal Article}

}

@article{zhang10,

author = {Zhang, Y. and Wen, Z. and Washburn, M. P. and Florens, L.},

title = {Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins},

journal = {Anal. Chem.},

volume = {82},

number = {6},

pages = {2272-2281},

keywords = {Albumins/analysis

Animals

Cattle

Fungal Proteins/analysis

Humans

Mice

Peptides/\*analysis

Proteins/\*analysis

Proteome/\*analysis

Proteomics/\*methods

Rabbits

Rats

Reproducibility of Results

Saccharomyces cerevisiae/chemistry

Swine},

ISSN = {1520-6882 (Electronic)

0003-2700 (Linking)},

DOI = {10.1021/ac9023999},

url = {<http://www.ncbi.nlm.nih.gov/pubmed/20166708>

[http://pubs.acs.org/doi/abs/10.1021/ac9023999}](http://pubs.acs.org/doi/abs/10.1021/ac9023999%7d),

year = {2010},

type = {Journal Article}

}

**SUPPLEMENTAL FIGURE LEGENDS**

**Supplemental Figure 1.** MSpC Graphic User Interface (GUI) and software flow chart. (**A**) Screenshot of the GUI. The input requires the user to select a Morpheus search summary file containing experiments to be analyzed. The user can optionally select a whitelist to filter output, and select an output directory. Additional options can set peptide and protein FDR cutoffs, and method of quantification for output, including Normalized Spectral Abundance Factor (NSAF), distributed NSAF (dNSAF), and unique NSAF (uNSAF). A progress bar highlights completion of the analysis. **(B)** Data analysis flow chart. Experiments and groups of experiments to be analyzed are imported through the Morpheus summary file. PSMs and protein groups are filtered at the specified FDR cutoff with a default of 1%. Due to shared peptides being attributed to only one instance of a protein group in Morpheus’s PSM file, PSMs are re-matched to all possible protein groups. PSMs are then cataloged as shared or as unique (distinctly matching one protein group) to generate NSAF, dNSAF, and uNSAF outputs. Finally, the output can be filtered for proteins of interest by specifying a comma delimited file containing unique identifiers and descriptions.

**Supplemental Figure 2.** Re- Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2). The array of UPS2 standards were spiked into *Xenopus laevis* egg **(Top)** and embryo **(Bottom)** extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein was determined by Morpheus and MSpC with a change from Figure 1 in that two unique peptides were required to quantify a protein. **(A)** A log-log plot of dNSAF versus concentrationfor each UPS2 protein detected across each fmol range. **(B)** A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

**Supplemental Table 1.** Table of dNSAF values for each 20S proteasome subunit generated by analyzing the proteasome spike in experiments with the Morpheus and MSpC pipeline. The top half of the table lists α1-7 (PAA-PAG where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms).

**Supplemental Table 2.** Table of adj\_NSAF values for each 20S proteasome subunit (equivalent to dNSAF) generated by analyzing the proteasome spike in experiments with the TPP and ABACUS pipeline. The top half of the table lists α1-7 (PAA-PAG where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms).