Quantitative proteome dataset profiling of UBC4 and UBC5 deletion strains in Saccharomyces cerevisiae

Valentina Rossio*, Joao A Paulo*

Department of Cell Biology, Blavatnik Institute at Harvard Medical School, Boston, MA 02115, USA

A R T I C L E   I N F O

Article history:
Received 3 September 2022
Revised 25 October 2022
Accepted 4 November 2022
Available online 9 November 2022

Dataset link: Quantitative proteome dataset profiling of UBC4 and UBC5 deletion strains in Saccharomyces cerevisiae (Original data)

Keywords:
Ubiquitin proteasome system
Protein stability
Yeast
TMT
Isobaric tagging

A B S T R A C T

The Ubiquitin-Proteasome System (UPS) regulates many cellular processes in eukaryotic cells. Ubiquitylation by the UPS mainly directs proteins to proteasomal degradation, but it can also have non-degradative functions, such as regulating protein activity or localization. The small protein ubiquitin is conjugated to its substrates via a cascade of E1-E2-E3 enzymes. Dysregulation of the UPS has been implicated in the genesis and progression of many diseases, such as neurodegenerative diseases and cancer; thus, the UPS components are attractive targets for developing pharmaceutical drugs. E2s, or ubiquitin conjugating enzymes, are central players of the UPS. E2s function in tandem with specific ubiquitin ligases (E3s) to transfer ubiquitin to substrates. Here, we present the first proteome stability analysis of two closely related ubiquitin conjugating enzymes, Ubc4 and Ubc5, in S. cerevisiae. These two E2s are nearly identical, having 92% sequence identity and differing by only 11 amino acid residues. This dataset is of broad interest because higher eukaryotes express ubiquitin conjugating enzymes that are analogous to the yeast Ubc4/5. The data have been deposited in ProteomeXchange with the dataset identifier PXD037315.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

* Corresponding authors.
E-mail addresses: valentina_rossio@hms.harvard.edu (V. Rossio), joao_paulo@hms.harvard.edu (J.A. Paulo).

https://doi.org/10.1016/j.dib.2022.108737
2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Specifications Table

| Subject                       | Omics: Proteomics                                      |
|-------------------------------|--------------------------------------------------------|
| Specific subject area         | Yeast, proteomics, Ubiquitin Proteasome System         |
| Type of data                  | Figure                                                 |
|                               | Excel Table                                            |
| How the data were acquired    | Mass spectrometric data were collected on an Orbitrap Eclipse mass spectrometer coupled to a Proxeon NanoLC-1200 UHPLC (ThermoFisher Scientific). Data were acquired using Xcalibur 4.1 and Tune 3.5. RAW files were converted to mzXML using MConvert 3.0. Database searches were performed with the Comet (2022.01 rev. 0) search engine. |
| Data format                   | Raw                                                   |
| Description of data collection| Exponentially growing *S. cerevisiae* (wild type, *ubc4Δ* and *ubc5Δ*) cells were collected. Proteins were chloroform-methanol precipitated and digested with LysC followed by trypsin, prior to LC-MS/MS analysis. Digested samples were labeled with tandem mass tag (TMTpro) reagents. The pooled sample was fractionated into 96 fractions which were then concatenated into 24 superfractions prior to LC-MS/MS analysis. |
| Data source location          | Harvard Medical School                                 |
|                               | Boston, MA                                             |
|                               | USA                                                   |
| Data accessibility            | Mass spectrometry data have been shared in the ProteomeXchange consortium via the PRIDE repository.  |
|                               | Pride                                                  |
|                               | Data identification number: PXD037315                    |
|                               | Direct URL to data:                                    |
|                               | http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD037315 |
|                               | Instructions for accessing these data:                 |
|                               | None. Data are publicly accessible.                    |

Value of the Data

- This is the first dataset exploring the effect on proteome stability of Ubc4 and Ubc5.
- This dataset is of broad interest because it offers novel targets of the UPS that are involved in many biological processes. It can be useful for cell biologists, for enzymologists, and for researchers studying UPS alterations in diseases.
- The UPS is exploited as a pharmacological target (*i.e.*, for neurodegenerative diseases). These data can provide insight into the development of drugs targeting the UPS.

1. Data Description

This article contains both the RAW data files (*.raw) and the search engines results (*.mzIdentML) from a multiplexed mass spectrometry experiment in yeast cells, as well as the analyzed data. These data have been deposited in the PRIDE repository (ProteomeXchange Consortium) with the identifier PXD037315. Fig. 1 illustrates the sample preparation workflow for LC-MS/MS analysis (Fig. 1A) and also illustrates distributions of the coefficient of variation for both the proteins and peptides analyzed in the proteomics experiment (Fig. 1B). Supplemental Table 1 lists the RAW files, the fraction associated with that file, along with the compensation voltage (CV) sets used for the associated analysis. Supplemental Table 2 lists the protein identifications. Columns include: protein ID, gene symbol, description, protein group ID, the number of peptides assigned to a given protein, 9 columns of TMT signal-to-noise values and 9 columns of TMT signal-to-noise values that have been scaled to 100 across all channels.
2. Experimental Design, Materials and Methods

2.1. Yeast Strains and Growth Conditions

All yeast strains used in this study are isogenic to W303 (ade2-1, trp1-1, leu2-3112, his3-11, 15, ura3). Standard yeast genetics was used to generate ubc4Δ and ubc5Δ strains [1]. Cells were grown as described previously [2]. Briefly, triplicate cultures of wild type, ubc4Δ and ubc5Δ cells were grown overnight at 25 °C in YEPD medium (1% yeast extract, 2% bactopeptone, 50 mg/l adenine, 2% glucose). The next day, the cultures were diluted with fresh medium to OD600=0.3 (wild type and ubc5Δ) or OD600=0.4 (ubc4Δ) and allowed to grow at 25 °C until mid-exponential phase (OD600~1). Cells were collected by centrifugation at 2000 g for 2 min, rinsed with 1 ml cold water, flash-frozen in liquid nitrogen and stored at 80 °C until LC-MS/MS analysis.

2.2. Sample Processing for TMTpro-Proteomic Analysis

Lysis buffer (8 M urea in 200 mM EPPS, pH 8.5 plus protease and phosphatase inhibitors) was added to the yeast cell pellets. Samples were lysed by bead beating (5 cycles of 30 s). Protein concentration of each sample was determined with a BCA assay (according to manufacturer's instruction). Disulfide bonds were reduced with 5 mM TCEP (15 min), alkylated with 10 mM NEM (15 min) and the NEM was then quenched with 5 mM DTT (15 min). Alkylation and quenching were performed in the dark. 100 μg of protein from each sample was precipitated with chloroform-methanol [3], and then resuspended in 200 mM EPPS pH 8.5. Samples
were digested by Lys-C (overnight at 24 °C) and trypsin (6 h at 37 °C). 1 µg of each enzyme was used per 100 µg of protein.

As preparation for TMT labeling, acetonitrile was added to a final volume of 30% to each digest. For each sample, 50 µg of peptide were labeled with 100 µg of TMTpro reagents at room temperature for one hour. The labeling scheme was as follows: wt triplicates: 126,127n,127c; ubc4Δ triplicates: 128n,128c,129n; ubc5Δ triplicates: 129c,130n,130c. Upon verifying that labeling efficiency was >97% [4], the reactions were quenched by adding hydroxylamine to give a final concentration of ~0.3% and then incubated 15 min at room temperature. The label-check was a quality control step prior to the final pooling of TMT-labeled samples. Here, we combined a small amount (1–3 µL or ~2 µg) of each sample and analyzed it by mass spectrometry to confirm that the protein digestion was successful, if the degree of labeling was sufficient, and if the labeled samples contained approximately equal amount of peptides. During database searching, the TMTpro label was considered a variable modification at the N-terminus and at lysine residues. We then determined the labeling efficiency for the N-terminus and the lysine residues by dividing labeled N-terminal peptides by total peptides, and also the labeled lysine-terminating peptides by the total lysine-terminating peptides.

Samples were combined 1:1 so that each channel contained the same amount of starting protein. The pooled protein sample was desalted by solid phase extraction (with a 50 mg Sep-Pak column). Once dried and reconstituted in 10 mM ammonium bicarbonate/5% acetonitrile, 300 µg of peptide was fractionated by basic pH reversed-phase (BPRP) using an Agilent 1200 pump with an Agilent 300Extend C18 column (250 mm long, 2.1 mm ID, with 3.5 µm particles). Peptides were separated over a 50-min linear gradient spanning 5% to 35% acetonitrile in 10 mM ammonium bicarbonate pH 8 while using a flow rate of 0.25 ml/min. In total, 96 fractions were collected which were then concatenated to 24 superfractions which were divided into two sets of 12 non-adjacent superfractions [3]. The acidified fractions (1% formic acid, final volume) were vacuum centrifuged to near dryness. Finally, each was desalted via StageTip [4], dried again, and reconstituted in 5% acetonitrile, 5% formic acid for analysis by LC-MS/MS.

2.3. Mass Spectrometry Data Collection and Analysis

An Orbitrap Eclipse mass spectrometer was used to collect the data. The instrument was coupled to a Proxeon NanoLC-1200 UHPLC and to a FAIMSpro interface [5]. The sample was analyzed across a 100 µm capillary column packed with 35 cm of Accucore 150 resin (2.6 µm, 150 Å; ThermoFisher Scientific).

In total 24 RAW files were acquired. Data for the first 12 superfractions were collected using a Orbitrap Eclipse with a CV set of –40/–60/–80 V while the remaining 12 superfractions were analyzed with a CV set of –30/–50/–70 V also over a 90 min gradient. A one second TopSpeed cycle was used for each CV. The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 60,000, 400–1600 Th, automatic gain control (AGC) target “standard”, maximum injection time “auto”). The hrMS2 stage consisted of fragmentation by higher energy collisional dissociation (HCD, normalized collision energy 36%) and analysis using the Orbitrap (AGC 200%, maximum injection time 86 ms, isolation window 0.7 Th, resolution 50,000).

The spectra, which were converted to mzXML with MSconvert [6], were searched with the Saccharomyces Genome Database (SGD; August 2021). Our forward database consisted of 6077 reviewed yeast entries and 115 common contaminants. This database was concatenated with a decoy database in which each protein sequence is reversed so as to enable the calculation of the target-decoy strategy for false discovery rate. These searches were performed using 50 ppm and 0.03 Da mass tolerances for precursors and fragments, respectively. We have traditionally used the 50 ppm mass tolerance for our Sequest and now Comet database searches. These wide mass tolerance windows were chosen to maximize sensitivity in conjunction with Comet searches and linear discriminant analysis [7]. In addition, oxidation of methionine residues (+15.995 Da) was set as a variable modification, whereas alkylation with N-ethylmaleimide at cysteine residues (+125.048 Da) and TMTpro tag modifications at peptide N-termini and lysine
residues (+304.207 Da) were set as static modifications. A linear discriminant analysis was performed for PSM filtering, such that a 1% false discovery rate (FDR) for peptide-spectrum matches (PSMs) was set [8,9], after which the data were assembled further to a final protein-level FDR of 1% [10]. The manufacturer-provided isotopic impurities of each TMTpro reagent were used to correct reporter ion intensities [11]. The peptide signal-to-noise (S/N) measurements for each protein were summed. These values were normalized to equate the sum of all protein signals in each channel, so that equal protein loading was taken into account.

**Ethics Statement**

This work does not include human subjects, animal experiments or data collected from social media platforms.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work presented in this paper.

**Data Availability**

Quantitative proteome dataset profiling of UBC4 and UBC5 deletion strains in *Saccharomyces cerevisiae* (Original data) (Pride).

**CRediT Author Statement**

Valentina Rossio: Investigation, Conceptualization, Writing – original draft; Joao A Paulo: Resources, Conceptualization, Data curation, Software, Writing – original draft.

**Acknowledgments**

We would like to acknowledge S. P. Gygi and the Taplin Mass Spectrometry Facility at Harvard Medical School for use of their mass spectrometers and software. We thank S. Piatti for the construction of the yeast deletion strains and for performing the yeast experiment. We thank R.W. King for helpful suggestions. This work was supported by the Cell Biology Education and Fellowship Fund (V.R) and NIH grant R01 GM132129 (JAP).

**Supplementary Materials**

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2022.108737.

**References**

[1] C. Janke, M.M. Magiera, N. Rathfelder, C. Taxis, S. Reber, H. Maekawa, A. Moreno-Borchart, G. Doenges, E. Schwob, E. Schiebel, M. Knop, A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes, Yeast 21 (11) (2004) 947–962, doi:10.1002/yea.1142.

[2] V. Rossio, E. Galati, M. Ferrari, A. Pellicioli, T. Sutani, K. Shirahige, G. Lucchini, S. Piatti, The RSC chromatin-remodeling complex influences mitotic exit and adaptation to the spindle assembly checkpoint by controlling the Cdc14 phosphatase, J. Cell Biol. 191 (5) (2010) 981–997, doi:10.1083/jcb.201007025.
[3] J. Navarrete-Perea, Q. Yu, S.P. Gygi, J.A. Paulo, Streamlined tandem mass tag (SL-TMT) protocol: an efficient strategy for quantitative (phospho)proteome profiling using tandem mass tag-synchronous precursor selection-MS3, J. Proteome Res. 17 (6) (2018) 2226–2236, doi:10.1021/acs.jproteome.8b00217.

[4] J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, Anal. Chem. 75 (3) (2003) 663–670, doi:10.1021/ac026117i.

[5] J. Saba, E. Bonnei, C. Poree, K. Eng, P. Thibault, Enhanced sensitivity in proteomics experiments using FAMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer, J. Proteome Res. 8 (7) (2009) 3355–3366, doi:10.1021/pr801106a.

[6] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egerton-Koff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (10) (2012) 918–920, doi:10.1038/nbt.2377.

[7] E.L. Huttlin, M.P. Jedrychowski, J.E. Elias, T. Goswami, R. Rad, S.A. Beausoleil, J. Villen, W. Haas, M.E. Sowa, S.P. Gygi, A tissue-specific atlas of mouse protein phosphorylation and expression, Cell 143 (7) (2010) 1174–1189, doi:10.1016/j.cell.2010.12.001.

[8] S.A. Beausoleil, J. Villen, S.A. Gerber, J. Rush, S.P. Gygi, A probability-based approach for high-throughput protein phosphorylation analysis and site localization, Nat. Biotechnol. 24 (10) (2006) 1285–1292, doi:10.1038/nbt1240.

[9] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat. Methods 4 (3) (2007) 207–214, doi:10.1038/nmeth1019.

[10] M.M. Savitski, M. Wilhelm, H. Hahne, B. Kuster, M. Bantscheff, A Scalable approach for protein false discovery rate estimation in large proteomic data sets, Mol. Cell. Proteom. 14 (9) (2015) 2394–2404, doi:10.1074/mcp.M114.046995.

[11] G.C. McAlister, E.L. Huttlin, W. Haas, L. Ting, M.P. Jedrychowski, J.C. Rogers, K. Kuhn, J. Pike, R.A. Grothe, J.D. Blethrow, S.P. Gygi, Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses, Anal. Chem. 84 (17) (2012) 7469–7478, doi:10.1021/ac301572t.