



## Data Article

# Dataset for suppressors of amyloid- $\beta$ toxicity and their functions in recombinant protein production in yeast



Xin Chen<sup>a,b,\*</sup>, Xiaowei Li<sup>a</sup>, Boyang Ji<sup>a,c</sup>, Yanyan Wang<sup>a</sup>, Olena P. Ishchuk<sup>a</sup>, Egor Vorontsov<sup>d</sup>, Dina Petranovic<sup>a,b</sup>, Verena Siewers<sup>a,b,\*</sup>, Martin K.M. Engqvist<sup>a</sup>

<sup>a</sup> Division of Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

<sup>b</sup> Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

<sup>c</sup> BioInnovation Institute, Ole Møløes Vej 3, Copenhagen DK-2200, Denmark

<sup>d</sup> Proteomics Core Facility, Sahlgrenska Academy, University of Gothenburg, SE-413 90, Gothenburg, Sweden.

## ARTICLE INFO

## Article history:

Received 16 May 2022

Accepted 23 May 2022

Available online 27 May 2022

Dataset link: [Dataset for suppressors of amyloid- \$\beta\$  toxicity and their functions in recombinant protein production in yeast \(Original data\)](#)

## Keywords:

Amyloid- $\beta$

Gene engineering

Transcriptome

Proteome

Yeast *Saccharomyces cerevisiae*

## ABSTRACT

The production of recombinant proteins at high levels of often induces stress-related phenotypes by protein misfolding or aggregation. These are similar to those of the yeast Alzheimer's disease (AD) model in which amyloid- $\beta$  peptides (A $\beta$ 42) were accumulated [1,2]. We have previously identified suppressors of A $\beta$ 42 cytotoxicity via the genome-wide synthetic genetic array (SGA) [3] and here we use them as metabolic engineering targets to evaluate their potentiality on recombinant protein production in yeast *Saccharomyces cerevisiae*. In order to investigate the mechanisms linking the genetic modifications to the improved recombinant protein production, we perform systems biology approaches (transcriptomics and proteomics) on the resulting strain and intermediate strains. The RNAseq data are preprocessed by the nf-core/RNAseq pipeline and analyzed using the Platform for Integrative Analysis of Omics (PIANO) package [4]. The quantitative proteome is analyzed on an Orbitrap Fusion Lumos

DOI of original article: [10.1016/j.ymben.2022.04.005](https://doi.org/10.1016/j.ymben.2022.04.005)

\* Corresponding authors at: Division of Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden.

E-mail addresses: [cxin@chalmers.se](mailto:cxin@chalmers.se) (X. Chen), [siewers@chalmers.se](mailto:siewers@chalmers.se) (V. Siewers).

<https://doi.org/10.1016/j.dib.2022.108322>

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

mass spectrometer interfaced with an Easy-nLC1200 liquid chromatography (LC) system. LC-MS data files are processed by Proteome Discoverer version 2.4 with Mascot 2.5.1 as a database search engine. The original data presented in this work can be found in the research paper titled “Suppressors of Amyloid- $\beta$  Toxicity Improve Recombinant Protein Production in yeast by Reducing Oxidative Stress and Tuning Cellular Metabolism”, by Chen et al. [5].

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	Biotechnology, Biological science
Specific subject area	Recombinant protein production, cell factories
Type of data	Images and tables
How the data were acquired	RNA-seq was performed at the National Genomics Infrastructure (NGI) of SciLifeLab and preprocessed by the nf-core/RNAseq pipeline. Quantitative proteome was performed at the Proteomics Core Facility at the University of Gothenburg. The omics data were analyzed using the Platform for Integrative Analysis of Omics (PIANO) package and the Database for Annotation, Visualization and Integrated Discovery (DAVID).
Data format	Raw and analyzed data
Description of data collection	For RNAseq data, RNA was extracted using the RNeasy Mini Kit from 10 OD <sub>600</sub> of exponential growing cells. The preparation of mRNA samples was applied with the Illumina TruSeq Stranded mRNA Library Pre kit. The fragments were clustered on cBot and sequenced on a HiSeq 4000 with paired ends (2 × 150 bp). The number of read pairs for each sample ranged from 12.0 to 16.0 million. After quality control, the raw reads from each sample were mapped to the <i>S. cerevisiae</i> CEN.PK 113-7D reference genome using TopHat (v 2.0.12) with 84.0 – 86.7 % of the reads successfully mapped. The data were further preprocessed by the nf-core/RNAseq pipeline. For quantitative proteome data, 35 µg of total protein from each of the samples and from the reference pool were used for the modified filter-aided sample preparation. Liquid chromatography-MS (LCMS) experiments were performed on an Orbitrap Fusion Lumos mass spectrometer interfaced with an Easy-nLC1200 nanoflow LC system. Peptide and protein identification and quantification was performed using Proteome Discoverer version 2.4 with Mascot 2.5.1 as a database search engine.
Data source location	The RNA-seq raw data were collected at the National Genomics Infrastructure (NGI) of SciLifeLab, Stockholm, Sweden. The quantitative proteomics data were collected at the Proteomics Core Facility, University of Gothenburg, Gothenburg, Sweden.
Data accessibility	The RNA-seq raw data can be downloaded from the Genome Expression Omnibus website with the dataset identifier GSE185570 (GEO Accession viewer (nih.gov)). The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE [6] partner repository with the dataset identifier PXD030111 (ProteomeXchange Dataset PXD030111).
Related research article	Xin Chen, Xiaowei Li, Boyang Ji, Yanyan Wang, Olena P. Ishchuk, Egor Vorontsov, Dina Petranovic, Verena Siewers, and Martin K. M. Engqvist, Suppressors of Amyloid- $\beta$ Toxicity Improve Recombinant Protein Production in yeast by Reducing Oxidative Stress and Tuning Cellular Metabolism [5].

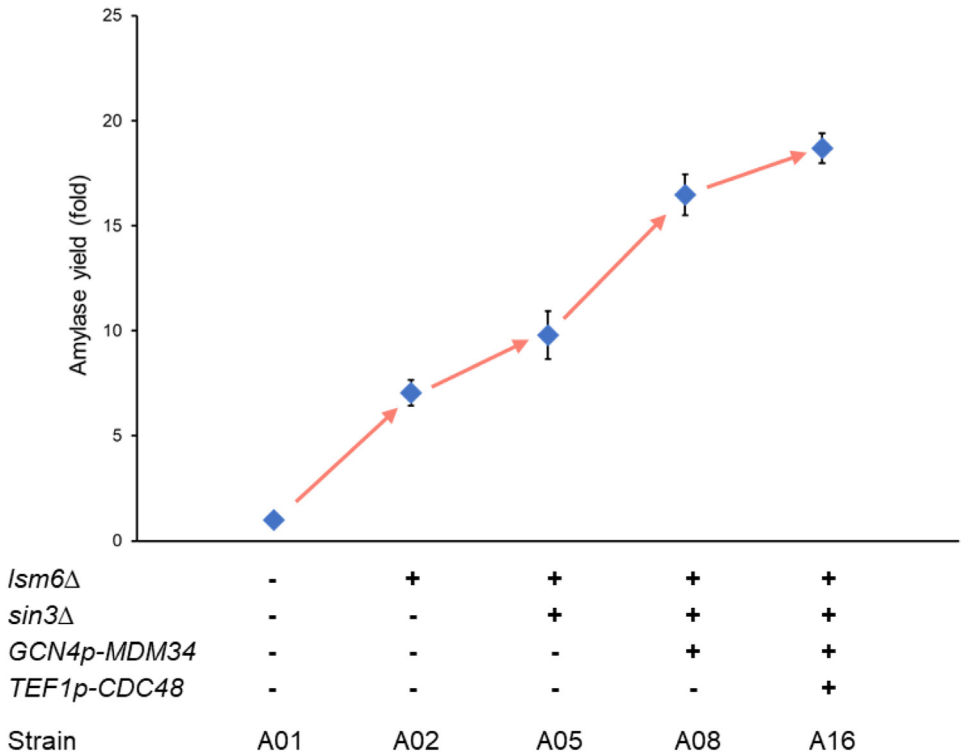
## Values of the Data

- The data provide detailed information on the 2615 differentially expressed genes (DEGs) which are commonly expressed in all engineered strains compared with control strain.

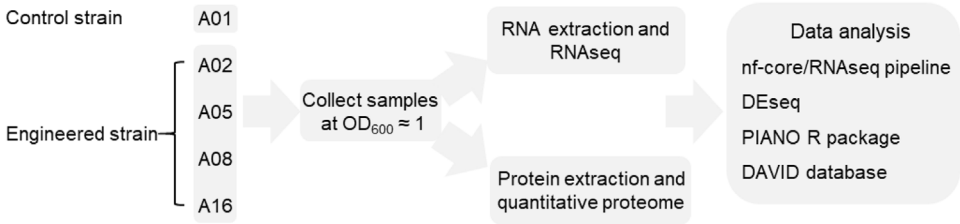
- The data provide detailed information on the 1171 differentially expressed proteins (DEGs) which are commonly expressed in all engineered strains compared with control strain.
- Information on this data article will be helpful for the researchers in the eukaryotic protein production and metabolic engineering fields.
- This data set will be applied to look for potential targets to develop synthetic cell factories for biosynthesis of valuable proteins.

## 1. Data Description

In our previous study, we applied a yeast genome-wide synthetic genetic array (SGA) technology with our  $A\beta 42$  expression model to screen for genetic mutants in which the  $A\beta 42$  cytotoxic phenotype was altered [3]. Two yeast deletion mutant libraries were included, a nonessential genes collection (~ 4300 deletion strains) and an essential genes collection (~ 1200 temperature sensitive mutant strains) [3]. According to the SGA screen, 46 resistant mutants with reduced  $A\beta 42$  cytotoxicity and 20 sensitive mutants with enhanced  $A\beta 42$  cytotoxicity were selected to evaluate their effects on recombinant protein production. The gene deletions and gene overexpression strategies were applied to resistant mutants and sensitive mutants, respectively. The recombinant  $\alpha$ -amylase was used as a model protein [7], with the  $\alpha$ -factor leader in front of open reading frame (ORF), which has been proved to increase protein production [8]. The genes led to an improved  $\alpha$ -amylase production yield, were chosen for combinatorial engineering to further enhance protein production. After four rounds of genetic screening,  $\alpha$ -amylase production was serially increased to 18.7-fold in the final engineered strain A16 compared with control strain A01 (Fig. 1).



**Fig. 1.** Summary of combinatorial gene engineering. The best performance strain from each round of genetic screening is presented. The results are shown as the average values  $\pm$  SD from four independent biological replicates.



**Fig. 2.** Schematic workflow for transcriptomic and proteomic processes of engineered strains.

These identified genes (*LSM6*, *SIN3*, *MDM34*, and *CDC48*) were reported to be involved in different cellular networks. To get systems-level insight on how these genes were involved in improved recombinant protein production, we performed the transcriptomics and proteomics analysis on the control strain A01 and engineered strains including A02, A05, A08, and A16 (Fig. 2).

Using the transcriptomic data (raw data available at GEO Accession viewer (nih.gov)), we first analyzed the global expression pattern via the principal component analysis (PCA). The results showed that engineered strains had a different gene expression profile compared to control strain A01. Strains were clustered into three groups from the first and second PCA components: group 1 (strain A01), group 2 (strain A02), and group 3 (strains A05, A08, and A16). To examine the extent of changes in the engineered strains, we further performed the pair-wise comparisons. Using the strain A01 as control, the differentially expressed genes (DEGs) were identified from all engineered strains. 2615 genes were found commonly differentially expressed in all engineered strains ( $p\text{-adj} < 0.01$ , dataset 1). The global changes of proteome were also quantified (raw data available at ProteomeXchange Dataset PXD030111). The PCA showed similar separation between control strain A01 and engineered strains. The differentially expressed proteins were analyzed and 1171 proteins were differentially expressed in all engineered strains compared with control strain A01 ( $p\text{-adj} < 0.05$ , dataset 2).

## 2. Experimental Design, Materials and Methods

### 2.1. Materials and Methods

Yeast strains were cultured at 30 °C in the SD-2 × SCAA media, which is an optimized culture media for protein production [9]. To collect cells for transcriptome and proteome analysis, a single colony was inoculated into 1 mL of SD-2 × SCAA medium and cultivated at 30 °C with 200 rpm agitation overnight. Biological triplicates were taken from each strain. The next day, the precultures were diluted to an initial  $OD_{600}$  of 0.05 in 20 mL of SD-2 × SCAA medium and cultivated to early exponential phase ( $OD_{600} \approx 1$ ) for sample collection.

### 2.2. Transcriptome Profiling

10  $OD_{600}$  of cells were collected in 50 ml falcon tubes with ice and centrifuged at 5000 rpm for 5 min at 4 °C. The cell pellets were snap-frozen in liquid nitrogen and stored at -80 °C before RNA extraction. RNA was extracted using the RNeasy Mini Kit (Qiagen), and RNA quality was evaluated using the 2100 Bioanalyzer (Agilent Technologies). RNA-seq was performed by the National Genomics Infrastructure (NGI) of SciLifeLab in Stockholm, Sweden (<https://ngisweden.scilifelab.se/>). The mRNA samples were prepared for sequencing via the Illumina TruSeq Stranded mRNA Library Pre kit (Illumina). The fragments were clustered on cBot and sequenced on a HiSeq 4000 with paired ends (MID Output 2 × 150 bp). The number of read pairs for each sample ranged from 12.0 to 16.0 million. After quality control, the raw reads from

each sample were mapped to the *S. cerevisiae* CEN.PK 113-7D reference genome using TopHat (v 2.0.12) with 84.0–86.7% of the reads successfully mapped. The data were preprocessed by the nf-core/RNAseq pipeline (<https://github.com/nf-core/rnaseq>, SciLifeLab). The differential gene expression was analyzed using DESeq. The reporter GO terms and reporter TFs were analyzed using the Platform for Integrative Analysis of Omics (PIANO) package [4] in R. The Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) was used to analyze functional enrichment of KEGG pathways and biological processes. Heatmaps were generated using the pheatmap package in R.

### 2.3. Sample Preparation for Proteomic Analysis

At early exponential phase ( $OD_{600} \approx 1$ ), 10 ml of cell culture was collected into pre-chilled Falcon tubes and centrifuged at 4 °C for 5 min ( $2000 \times g$ ). Biological triplicates were taken from each strain. Cells were washed once with ice-cold PBS buffer, snap-frozen in liquid nitrogen and stored at -80 °C. Samples were suspended in 200  $\mu$ l of the lysis buffer containing 2% sodium dodecyl sulfate (SDS) and 50 mM triethylammonium bicarbonate (TEAB), and homogenized using lysis matrix D (1.4 mm ceramic spheres) on a FastPrep-24 instrument (MP Biomedicals, OH, USA) for 4 repeated 40 s cycles at 6.5 m/s. Lysed samples were centrifuged at 16,200 $\times g$  for 10 min and the supernatants were transferred to clean tubes. The pellets were washed with 200  $\mu$ l of the lysis buffer, centrifuged at 16,200 $\times g$  for 10 min, and the supernatants were combined with the corresponding lysates from the previous step. Protein concentrations in the lysates were determined using Pierce BCA Protein Assay Kit (Thermo Fischer Scientific) and the Benchmark Plus microplate reader (Bio-Rad Laboratories, Hercules, CA, USA) with bovine serum albumin (BSA) solutions as standards. The proteomic reference pool was prepared by taking an equal aliquot from each individual sample.

### 2.4. Quantitative Proteome Analysis

Proteomic analysis was performed as described previously with modification [10]. In brief, 35  $\mu$ g of total protein from each of the samples and from the reference pool were used for the modified filter-aided sample preparation (FASP) [11], which included the two-stage digestion of each sample with trypsin (Pierce Trypsin Protease, Thermo Fisher Scientific) in digestion buffer containing 0.5% sodium deoxycholate (SDC) and 50 mM TEAB, and subsequent labeling with TMTpro16plex reagents (Thermo Fisher Scientific) according to the manufacturer's instructions. The labeled samples were combined into one TMTpro set and fractionated into 404 primary fractions using a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific). Peptide pre-fractionation was performed on an XBridge BEH C18 column (3.5  $\mu$ m, 3.0  $\times$  150 mm, Waters Corporation) at pH 10 and primary fractions were concatenated into final 20 fractions for liquid chromatography-MS (LC-MS) analysis. Each fraction was analyzed using an Orbitrap Fusion Lumos mass spectrometer interfaced with an Easy-nLC1200 liquid chromatography system (both Thermo Fisher Scientific).

LC-MS data files were processed in Proteome Discoverer version 2.4 (Thermo Fisher Scientific). Reference database for *Saccharomyces cerevisiae* ATCC 204508 / S288c was downloaded from Uniprot (November 2019) and supplemented with common proteomic contaminant sequences. Mascot version 2.5.1 (Matrix Science, London, UK) was used as a database search engine with 5 ppm precursor tolerance and 0.6 Da fragment tolerance; trypsin with 1 allowed missed cleavage as an enzyme rule; oxidation on methionine as a variable modification; methylthiolation on cysteine, TMTpro on lysine and peptide N-termini as a fixed modification; Percolator for PSM validation with the strict FDR threshold of 1%. Reporter ions were identified within 3 mmu in MS3 spectra, quantification was based on unique peptides and reporter S/N, with the correction for isotopic impurities, the average reporter S/N threshold of 10 and SPS match threshold of 40%, reporter normalization on total peptide amount.

Differential protein expression (Log<sub>2</sub>-FoldChange) and corresponding significance (*p*-adj) were calculated by the Benjamini–Hochberg method and used as input. The reporter GO terms were analyzed using the PIANO R package [4]. The DAVID Database (<https://david.ncicfcrf.gov/>) was used to analyze functional enrichment of KEGG pathway and biological process.

## 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 reported in this paper.

## Data Availability

Dataset for suppressors of amyloid- $\beta$  toxicity and their functions in recombinant protein production in yeast (Original data) (DIB).

## CRediT Author Statement

**Xin Chen:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing; **Xiaowei Li:** Investigation, Formal analysis, Writing – review & editing; **Boyang Ji:** Software, Writing – review & editing; **Yanyan Wang:** Methodology, Validation; **Olena P. Ishchuk:** Investigation, Writing – review & editing; **Egor Vorontsov:** Methodology, Writing – review & editing; **Dina Petranovic:** Funding acquisition, Project administration, Writing – review & editing; **Verena Siewers:** Project administration, Supervision, Writing – review & editing; **Martin K.M. Engqvist:** Project administration, Supervision, Writing – review & editing.

## Acknowledgements

We acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We acknowledge the Proteomics Core Facility at the University of Gothenburg, Sweden. The work was financially supported by the VINNOVA center CellNova (2017-02105) and Novo Nordisk Foundation (NNF10CC1016517).

## References

- [1] X. Chen, D. Petranovic, Amyloid- $\beta$  peptide-induced cytotoxicity and mitochondrial dysfunction in yeast, *FEMS Yeast Res.* 15 (6) (2015), doi:[10.1093/femsyr/fov061](https://doi.org/10.1093/femsyr/fov061).
- [2] X. Chen, M.M.M. Bisschops, N.R. Agarwal, B. Ji, K.P. Shanmugavel, D. Petranovic, Interplay of energetics and ER Stress exacerbates alzheimer's amyloid- $\beta$  ( $A\beta$ ) toxicity in yeast, *Front. Mol. Neurosci.* 10 (2017) 232, doi:[10.3389/fnmol.2017.00232](https://doi.org/10.3389/fnmol.2017.00232).
- [3] X. Chen, B. Ji, X. Hao, X. Li, F. Eisele, T. Nyström, D. Petranovic, FMN reduces amyloid- $\beta$  toxicity in yeast by regulating redox status and cellular metabolism, *Nat. Commun.* 11 (1) (2020) 867, doi:[10.1038/s41467-020-14525-4](https://doi.org/10.1038/s41467-020-14525-4).
- [4] L. Våremo, J. Nielsen, I. Nookaew, Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods, *Nucl. Acids Res.* 41 (8) (2013) 4378–4391, doi:[10.1093/nar/gkt111](https://doi.org/10.1093/nar/gkt111).
- [5] X. Chen, X. Li, B. Ji, Y. Wang, O.P. Ishchuk, E. Vorontsov, D. Petranovic, V. Siewers, M.K.M. Engqvist, Suppressors of amyloid- $\beta$  toxicity improve recombinant protein production in yeast by reducing oxidative stress and tuning cellular metabolism, *Metab. Eng.* 72 (2022) 311–324, doi:[10.1016/j.ymben.2022.04.005](https://doi.org/10.1016/j.ymben.2022.04.005).
- [6] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucl. Acids Res.* 47 (D1) (2019) D442–d450, doi:[10.1093/nar/gky1106](https://doi.org/10.1093/nar/gky1106).

- [7] Z. Liu, K.E. Tyo, J.L. Martinez, D. Petranovic, J. Nielsen, Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*, *Biotechnol. Bioeng.* 109 (5) (2012) 1259–1268, doi:[10.1002/bit.24409](https://doi.org/10.1002/bit.24409).
- [8] J. Hou, K. Tyo, Z. Liu, D. Petranovic, J. Nielsen, Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*, *Metab. Eng.* 14 (2) (2012) 120–127, doi:[10.1016/j.ymben.2012.01.002](https://doi.org/10.1016/j.ymben.2012.01.002).
- [9] K.D. Wittrup, V. Benig, Optimization of amino acid supplements for heterologous protein secretion in *Saccharomyces cerevisiae*, *Biotechnol. Technol.* 8 (3) (1994) 161–166, doi:[10.1007/BF00161582](https://doi.org/10.1007/BF00161582).
- [10] R. Yu, K. Campbell, R. Pereira, J. Björkeröth, Q. Qi, E. Vorontsov, C. Sihlbom, J. Nielsen, Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast, *Nat. Commun.* 11 (1) (2020) 1881, doi:[10.1038/s41467-020-15749-0](https://doi.org/10.1038/s41467-020-15749-0).
- [11] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods* 6 (5) (2009) 359–362, doi:[10.1038/nmeth.1322](https://doi.org/10.1038/nmeth.1322).