Understanding the effect of cysteine on proteomic profiles in Saccharomyces cerevisiae with high consistency and accuracy using data independent acquisition

Ajay Bhat[1,2], Trayambak Basak[1,2], **Dipankar Malakar[3]**, Manoj Pillai[3], and Shantanu Sengupta[1,2]

1CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi, India, 2.Academy of Scientific and Innovative Research, New Delhi, India , 3. AB Sciex, 121 Udyog Vihar, Gurgaon, Haryana, India

ABSTRACT

In yeast *Saccaromyces cerevisiae* a good depth of coverage was obtained for the protein identification experiment, with 1334 proteins found at 1% global FDR. Using this information to interrogate the Data Independent Acquisition (DIA), termed as SWATH™ acquisition, we obtained good quantitation data on 963 proteins across the biological samples with only 100 ng on column injection of yeast protein digest. We show that suitable combinations of fragment ions extracted from these data sets are sufficiently specific to confidently identify and quantify query proteins over a dynamic range more than 4 orders of magnitude. Good reproducibility across the analytical replicates obtained as evident from very low percentage CV (less than 10%), highlighting the quality of the dataset. A comparative study between the SWATH™ acquisition and an isobaric tagging workflow using SCIEX iTRAQ® reagents for relative quantification of proteins was carried out to compare the strengths and utilities of each workflow for consistent, accurate and sensitive quantification of proteins in biological studies. A significant numbers of unique deregulated proteins were identified in both SWATH™ and iTRAQ® reagent workflow. SWATH™ analysis showed more robustness in quantitation specially for low abundant proteins. Glycolysis and TCA cycle are the two most prominent pathways that are enriched with the differentially expressed proteins found both in SWATH™and iTRAQ® reagent datasets.

This study indicate that SWATH™ could be used as effective tool for both identification and verification of deregulated proteins.

INTRODUCTION

Most proteomic studies use LC/MS/MS to identify and quantify peptides generated by the proteolysis of a biological sample. In proteomics, and many other sample types, the complexity and dynamic range of compounds are very large. This poses challenges for the traditional data dependent workflows, requiring very high speed MS/MS acquisition to deeply interrogate the sample in order to both identify and quantify a broad range of analytes. Data independent acquisition (DIA) strategies have recently emerged as a key workflow to increase the reproducibility, comprehensiveness of data collection and improve biological conclusions. DIA workflows have been enabled by the good improvement in MS/MS speed, sensitivity and resolution in today's accurate mass LC/MS systems.

Cysteine, a thiol containing amino acid synthesized via the trans-sulfuration pathway, is involved in variety of biological functions in a cell. To understand the mechanism of cysteine induced toxicity in yeast we present a workflow which uses the DIA strategy termed SWATH™ acquisition to quantitate large numbers of proteins across a set of yeast samples.

MATERIALS AND METHODS

Cells of Saccharomyces cerevisiae, both with and without cysteine treatment were lysed, digested with trypsin and divided into two aliquots. Aliquot one was subjected to an in-depth protein identification experiment using strong cation exchange (SCX) fractions to generate a deep fragment ion spectral library of S. cerevisiae. Aliquot 2 was analysed using SWATH™ acquisition where 34 Q1 windows of 25 Da were stepped through the precursor mass range and high resolution MS/MS was acquired (1). MS/MS accumulation time for each SWATH™ cycle was set to 100 ms. In iTRAQ® reagent labelled based relative quantitation experiment cells of S. cerevisiae, both with and without cysteine treatment lysed, digested with trypsin in similar way. The samples (two technical replicates) were then used for labelling with iTRAQ® Reagent 4-plex, and combined at equal ratios. Both in DIA ion library generation method and iTRAQ® experiment an Information Dependent Acquisition (IDA) method was used where maximum 12 most intense multiple charged ions per MS cycle were selected to perform MS/MS fragmentation. A dynamic exclusion criteria was applied to each of the ions for 10 seconds with one occurrence. In iTRAQ® analysis relatively higher collision energy was used during fragmentation in order to get better reporter ion data statistics. The accumulation time for each MS/MS experiment was set to 100 ms. The samples were loaded on the trap (Eksigent ChromXP™ 350 µm x 0.5 mm, 3 µm 120Å) and washed for 30 minutes at 3 µL/min. A 100 min gradient in multiple steps ranging from 5-50% Acetonitrile in water containing 0.1% formic acid was set up to elute the peptides from the ChromXP 3-C18, 0.075 x 150 mm, 3 um 120Å analytical column. Both the iTRAQ® and SWATH™ analysis were performed using Eksigent nanoLC-Ultra™ 2D system coupled with TripleTOF® 5600+ system (AB SCIEX). All data were processed using ProteinPilot™ Software 4.5 beta, SWATH™ Acquisition MicroApp in PeakView® software and MarkerView™ Software.

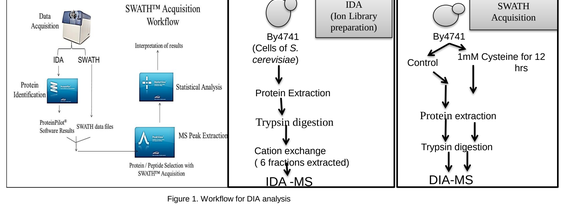

Figure 1. Workflow for DIA analysis

Results

High quality consistent data obtained among analytical replicates from SWATH™ analysis as evident not only from reproducible TIC but also from the correlation analysis of protein areas of technical replicates (Figure 2). Reproducibility among technical replicates also attributed from very low percentage CV of the peak areas of fragment ions obtained from one of the peptides APLDNDIGVSEATR of beta galactosidase digest spiked equally with each sample (Table 1). Reproducibility among technical replicates also reflected from cumulative frequency plot, which shows a presence of considerable proportions of low intense fragment ions with percent CV less than 20% along with relatively higher to moderate intense fragment ions (Figure 5). Moreover about 80% of the ions, peptides, proteins showed percent CV less than 20% among technical replicates (Figure 6). In IDA experiment 1334 proteins were identified with 1% FDR. Among these proteins 976 proteins, which having at least 2 unique peptides with 95% confidence were selected for relative quantification analysis. Where as in iTRAQ® labelled sample analysis yielded identification of 1588 proteins with 1% FDR. Among these proteins 1041 proteins were qualified for quantitation analysis (Table 2). In SWATH™ analysis along with high abundance protein like ENO2 (24556 ppm) very low abundant protein like TOM5 (0.269 ppm) were detected and quantified from on column load of 100 ng of total yeast cell lysate (Figure 4). The intensity range of the fragments ions qualified for quantification analysis was less than 1000 cps to more than 10000000 cps suggest wider dynamic range of detection and quantification (Figure 3). Depth of coverage similar to that obtained by MRM analysis (2, 3) was obtained but with comprehensive acquisition and highest multiplexing .

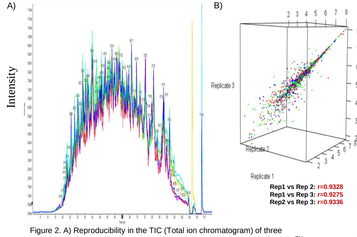

Figure 2. A) Reproducibility in the TIC (Total ion chromatogram) of three technical replicates of both control and cysteine trated cells used in SWATH™ analysis.B) Correlation between protein area (Log_{10} transformed) of three technical replicates of control sample

Control sample

Fragments	%CV
1176.55	2.490304
1289.63	8.605678
1061.52	7.481788
832.45	8.42439
563.27	10.82792

Cysteine treated sample

Fragments	%CV
1176.55	8.551965
1289.63	9.446172
1061.52	8.503135
832.45	7.719499
563.27	6.146572

Table 1. Reproducibility among technical replicates as evident from the low percentage CV calculated from the areas of the fragments of APLDNDIGVSEATR of beta galactosidase spiked equally with control and cysteine treated samples.

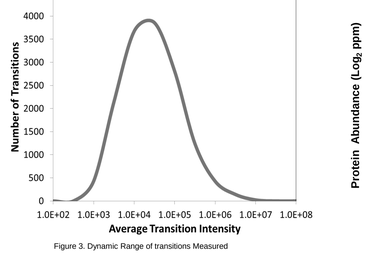

Figure 3. Dynamic Range of transitions Measured

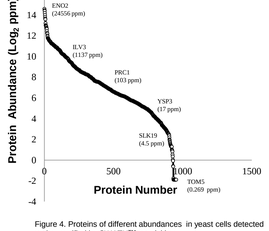

Figure 4. Proteins of different abundances in yeast cells detected and quantified by SWATH™ acquisition.

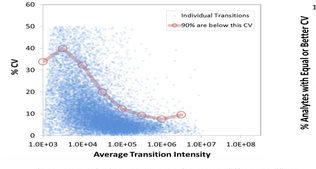

Figure 5. Cumulative frequency plots showing reproducibility at the different XIC peak areas

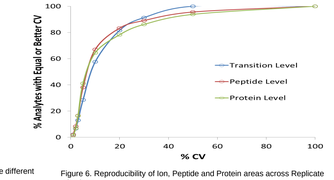

Figure 6. Reproducibility of Ion, Peptide and Protein areas across Replicate Injections

Experiment Type	Number of Proteins Identified with 1% FDR	Number of Proteins selected with high stringency for quantification	Percentage of quantifiable proteins
iTRAQ®	1588	1041	65.55
SWATH™	1334	963	72.22

Table 2. Identifiable and quantifiable proteins in SWATH™ and iTRAQ reagent

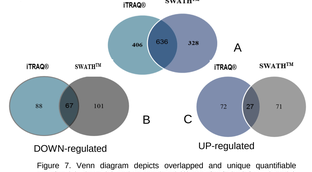

Figure 7. Venn diagram depicts overlapped and unique quantifiable proteins (A), down (B) and up regulated (C) proteins in both iTRAQ and S SWATH™ analysis.

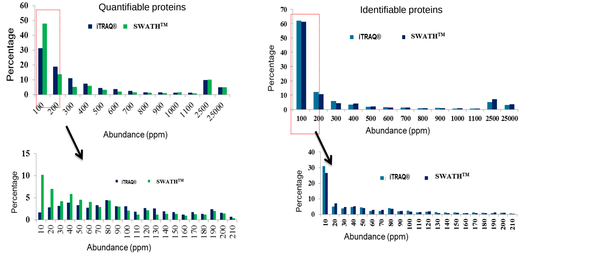

Figure 8. Abundance analysis of identifiable and quantifiable proteins by iTRAQ® and SWATH™ methods : Abundance values were taken from PaxDB integrated dataset, histogram represents the percentage of total identified or quantified proteins at different abundance in iTRAQ® and SWATH analysis.

Number of down regulated proteins (fold change less than 0.5 in comparison to control) in iTRAQ® and SWATH™ analysis were 155 and 168 respectively, in which 67 proteins found to be down regulated in both type of analysis. Where as number of up regulated proteins (fold change more than 2 in comparison to control) were 99 and 98 respectively for iTRAQ® and SWATH™ analysis, where 27 proteins found to be present common as up regulated proteins in both of the analysis (Figure 7). Cysteine treatment for 12 h to the yeast cells yielded significant numbers of deregulated proteins involve in Glycolysis and TCA cycle pathways. Deregulated proteins like Aconitase 1, Malate dehydogenase, Enolase 1 etc identified in both of iTRAQ® and SWATH™ analysis.

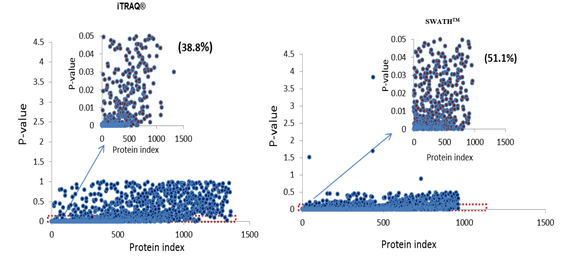

Figure 9. SWATH™ analysis showed more consistencies in the peptide level ratios belong to a specific protein as evident from lower p values.

Abundance analysis of various proteins showed that both SWATH™ and iTRAQ® analysis showed similar trend where a significant proportions of identified proteins are present in relatively low abundant region. Although in comparison to iTRAQ® analysis, SWATH™ showed significant numbers of quantifiable proteins specially at very low abundant region (Figure 8). In order to find out the consistencies between the peptide ratios belong to each of the quantifiable proteins, p values were calculated in both iTRAQ® and SWATH™ analysis. In SWATH™ analysis comparatively more number of proteins were quantified keeping more consistencies in the peptide ratios for a each of the proteins (p<0.05 for 51.1% of total proteins) in comparison to iTRAQ® analysis (p<0.05 for 38.8% of total proteins). Moreover the same trend was more prominent at low abundant proteins segment (Figure 9).

Conclusions

1. **The data independent acquisition strategy, MS/MS^{ALL} with SWATH™ Acquisition, provides a comprehensive analysis of yeast proteomes with good quantitative robustness.**
2. **Both SCIEX iTRAQ® and SWATH™ workflow identified a significant numbers of deregulated proteins.**
3. **There are significant numbers of unique deregulated proteins also found indicate the complementarity of two analysis methods.**
4. **SWATH™ analysis showed more consistencies in the quantitation specially for low abundant proteins. This is more likely due to the fact that in SWATH™ analysis the ions being collected in time, where an extracted ion chromatogram can be generated. This can help in getting a lot more ion statistics rather than iTRAQ analysis where the quantification is based on a single spectrum.**
5. **Significant number of deregulated proteins identified using both of the analysis type involve in various important biochemical pathways demand detailed investigation of those proteins.**
6. **This study revealed that SWATH™ can be used as a verification tool in the biomarker study.**

References

1. Gillet LC *et al (2012) Mol. Cell. Prot. E-pub.*
2. Picotti P *et al. (2009). Cell* **138 (795-806).**
3. Ghaemmaghami S, *et al. (2003). Nature* **425(6959), 737-41.**

Acknowledgement

This work was supported by CARDIOMED-Council of scientific and industrial Research(CARDIOMED-CSIR)

TRADEMARKS/LICENSING