

Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions over 2 weeks

Viswanadham Sridhara¹, Daniel R. Boutz^{2,3}, Craig Barnhart³, Jeffrey E. Barrick^{1,2,3,4}, Edward M. Marcotte^{1,2,3,4,5}, and Claus O. Wilke^{1,2,3,5}

November 18, 2014

¹Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX, USA

²Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA

³Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX, USA

⁴Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA

⁵Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

Abstract

A

1 Introduction

Research into

2 Results

We wanted to use an unrestricted approach to find the post-translational modifications in *E. coli* sample grown under glucose-limited conditions. This sample is not enriched for any post-translational modifications. So, the coverage of the PTMs identified is limited, however we can get information on most of the PTMs.

2.1 Running MODa on *E. coli* proteome

We did this analysis using MODa, a naive based spectral alignment search algorithm. Unlike other peptide identification search algorithms, MODa requires as an input the range of PTM masses instead of the actual PTMs. So, for our analysis we used a range of -200 to 200 Da which is typical to this search engine. We generated mass-spec raw data at different time points in exponential and

stationary phases of *E. coli* growth. At each of the 9 time points, we have 3 biological replicates. For each of the replicates, we ran MODa. We asked some of the following questions while analyzing the data: (a) How much of the proteome is modified? (b) How do these modifications change over time? (c) Can we find any novel modifications? (d) Can we explain the pattern of the PTMs over time using other diverse kinds of data such as RNA-seq?

2.2 Modified *E. coli* proteome

To estimate how much of the proteome is modified, we counted the number of peptide-spectral matches (PSMs) for each MODa run. We used standard error to plot the variation within 3 biological replicates. Figure 1 shows the total percent of the peptide-spectral matches on the y-axis and the growth curve time on the x-axis. Consistently 25% of the peptide-spectral matches seem to be modified at all the 9 time points analyzed. This is in agreement with some of the previous studies (Cite papers which show the same numbers). However note that a modification could be a mutation or a post-translational modification.

2.3 N-term protein acetylations are frequent in *E. coli*

Since MODa outputs mass-shift, we used UNIMOD database to map the mass-shift to the probable PTM. MODa outputs not only the mass-shifts, but also the amino-acids on which this mass-shift occurs. We focussed not only on the most common PTM-amino acid combination, but also on the PTMs irrespective of the amino acid. For example, we know acetylation is most common on n-term and lysine, however we also looked at the time-course of all acetylations i.e., sum of acetylations on all amino acids. (Add other citations and results from literature)

2.4 *E. coli* carboxylations, nitrosylations and other PTMs

To understand what kinds of PTMs or mutations are present in the sample, we plotted the frequency of the mass-shifts from the MODa output. MODa program calculates the mass-shifts in intervals of 1Da, along with the amino acid on which it finds the mass-shift. Along with the amino acids, it also calculates the N-term and C-term mass-shifts. Figure 2 shows the mass-shift frequency matrix. A zoomed-in version (b) shows that +1 Da peak is the most frequent one. However this modification seems to be ¹³C peak-picking as it seems to occur on all the amino acids, when we looked at the profile (Suppl Figure 1). Next frequent modification is +16Da. A look at profile (Suppl Figure 2) shows majority of oxidations on methionine which is expected.

Next we looked at the time course evolution of few post-translational modifications. In particular, we looked at acetylations, carboxylations, nitrosylations, phosphorylations and oxidations. Figure 3 shows the time course evolution of all acetylations, along with N-term and serine acetylations which seem the dominant ones. A close look at the serine acetylations show that they fall at the N-term x% of the time. During the exponential phase, the acetylations seem not to change much over time. However during the stationary and the long-stationary phases, the acetylations seem to go up. *E. coli* grown on glucose have shown to accumulate acetate (Cite proper references). Our results might suggest the same given the increase in the number of acetylated proteins from exponential to stationary phases. Table 1 shows the acetylated proteins found under these 2 different phases under *E. coli* growth under glucose limiting conditions.

Next, we investigated the phosphorylations found in this study. Note that this study is not enriched for phosphorylation or other modifications. Typically studies that identify phosphopeptides are generally enriched for phosphorylations using IMAC or TiO₂. However as we mentioned early, our goal is to identify different kinds of PTMs in this proteomics data. Figure 4 shows the phosphorylations at different points of the growth curve. The number of phosphorylations seem to be small and there is an increase of the number of phosphorylations from exponential to stationary phases. Similarly we looked at the carboxylations and nitrosylations (Suppl Figure 3), as they are shown in the past to be important in *E. coli*. Suppl Figure 4 shows the Na and K adduct profiles. They seem to occur on all the amino acids, except basic charge amino acids.

Pyroglutamate conversion and Succinylation: Glutamine conversion at n-term of the peptide to pyroglutamate is a well known PTM. Suppl Figure 5 shows that this conversion seems to be more or less same at different phases of the growth. (Add data about succinylation)

2.5 Oxidative damage and repair in *E. coli*

Finally, we investigated oxidative damage by looking at the oxidations identified by MODa program i.e., mass-shifts of +16Da. The oxidations seem to go down from exponential to stationary phases. This result seemed strange as we expected oxidative damage and hence the oxidations to go up as the *E. coli* reaches stationary and long-stationary phases. However may be the sulfoxide reductases MsrA and MsrB found in *e. coli* seem to fix methionine sulfoxide back to methionine. Since we also gather RNA-seq data on the same sample, we were able to retrieve the transcriptomic abundances of MsrA and MsrB along with proteomic abundances. There is not much change in the profiles of transcripts of these reductases i.e., the abundance seem to go down from exponential to stationary and stayed flat during long stationary phase. However the proteomic abundance seem to go down initially from exponential to stationary phase, however it surged up during the long-stationary phase indicating the probable fixing of MetSO to Met. Previously it was shown that these reductases play a role, however in this study for the first time we showed the time course fixation of MetSO to Met.

2.6 Novel phosphoserinegluconylation on ribosomal protein S6

To see if the results and trends of the PTMs over time holds, we changed the MODa range search from -300 to 300Da. The trends and the results were the same as previously shown. However, a +258 Da shift on serine is found consistently at all time points of the growth curve. When we looked into literature, this seemed to be a phosphorylation+gluconylation on Serine, which is a known modification. However it was not shown in literature to occur frequently on the ribosomal protein S6 in *E. coli*.

2.7 in *E. coli*

3 Discussion

We have developed

4 Conclusions

We have found

5 Materials and Methods

5.1 *E. coli* growth

Craig Barnhart part (Copied from John's paper): *E. coli* B REL606 was inoculated from freezer stock in 50mL DM500 and incubated at 37C overnight. 500 UL *change to micro* of the overnight culture was diluted in 50mL of DM500 at 37C and grown for 24hrs. On the day of the experiment, 500 UL *change to micro* of the 24hr culture was added to 10 flasks containing 50mL DM500 each, grown at 37C. At each time point 1ml was removed, washed with 0.7% NaCl, spun down, the supernatant was removed, and the remaining cell pellet was flash frozen using liquid nitrogen and stored at -80C. Samples for each experiment were taken from the same batch of culture.

To measure colony forming units (CFU) the OD600, at each time point, was taken relative to sterile DM500 glucose, cultures were diluted in sterile saline, and finally plated on DM agar supplemented with 0.2g/L glucose. Colonies were counted, after incubation at 37C for 24hr. Cultures for measuring CFU were grown separately from the main culture but in identical conditions.

5.2 Mass-spectrometry of *E. coli* proteome

Dan Boutz part (Copied from John's paper): Frozen pellets were re-suspended in 300 uL of buffer (50mM Tris-HCL pH 8.1, 100mM KCL, 5mM MgCl₂). 50uL of sample was removed for preparation, treated with 50uL trifluoroethanol (TFE), and placed on ice for 15min. DTT was then added to a final concentration of 5mM and incubated at 55C for 45 min. Samples were alkylated by addition of Iodoacetamide (IAM) to a final concentration of 15mM. Trypsin digest was performed by addition of 800uL of 50mM Tris pH 8.0, 2mM CaCl₂ followed by 2ug of trypsin. Digestion took place at 37C for 4-5 hrs and was stopped by 10uL of formic acid (1% vol./vol). Samples were then ultrafiltrated to remove insoluble and undigested material using Amicon Ultra MWCO 10kD spin-caps and finally concentrated and purified by C18 filtration. Liquid chromatography and mass spectrometry (LC/MS) was carried out on a LTQ-Orbitrap (Thermo Fisher). particular settings need to be filled in.

5.3 Post-translational modification identification and analysis

We used MODa (cite) to identify peptides and characterize novel PTMs. MODa outputs mass-shifts on the amino acids instead of the post-translational modifications. So, we used UNIMOD database to map the mass-shift to the appropriate PTM or a mutation. We ran separate searches for each of the 9 time points and at each time point, there were 3 biological replicates. So, there were in total 27 MODa searches. We used TACC for computing resources. The enzyme used in the searches is trypsin with fully-tryptic and no proline rule. The missed cleavages allowed are 2. Since the fragmentation technique used is CID, we looked for b/y ions. The mass-tolerance

of the precursor is 10ppm, while the mass-tolerance of the product ion is 0.5 Da. We used carbamidomethylation of cysteine as a static or fixed modification. As mentioned earlier, MODa requires a mass range to search for variable modifications, so we tried 2 scenarios: (a) mass range between -200 to 200Da and (b) second search with mass range between -100 to 300Da. We used REL606 sequence library from NCBI sequence database. To identify high-confidence hits, we used target-decoy approach (cite). In this approach, we reverse the original REL sequences and concatenate to the original sequence database to form a database that is twice as much as of the original sequence database. The idea is that there are as many false positive hits to that of the original database as that of the decoy database. We used a 1% FDR in this approach which is a general norm in mass-spectrometry based proteomics searches.

(Add how MODa uses score and other features to calculate the probability?), (Also add how MODa can distribute the hits with the charge state?)

5.4 Raw data and analysis scripts

All raw data and analysis scripts are available online in the form of a git repository at <https://github.com/clauswilke/PTMs>.

6 Author Contributions

Conceived and designed the experiments: V.S., C.O.W and E.M.M. Performed the experiments: V.S. Analyzed the data: V.S, C.O.W and E.M.M. Wrote the paper: V.S, D.R.B, C.B, J.E.B, C.O.W and E.M.M.

7 Acknowledgments

This project was funded by ARO Grant W911NF-12-1-0390. We thank members of the Colin Brown for helpful discussions on post-translational modifications. The Bioinformatics Consulting Group and the Texas Advanced Computing Center (TACC) at UT provided high-performance computing resources.

Figures

Figure 1. Flowchart. We use .

Figure 2. Misclassification . For j.

Figure 3. Probability . For .

Figure 4. Predictive . Reactions .

Figure 5. Predictive . Reactions .

Supplementary Figures

Figure S1: Scatter . We .

Figure S2: . We .

Figure S3: . For .

Supplementary Tables

Table S1: . These .

Table S2: . These .