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

Viswanadham Sridhara[1], Colin Brown[2], Daniel R. Boutz[2,3], Craig Barnhart[3], Maria Person[2,3], Edward M. Marcotte[1,2,3,4,5], and Claus O. Wilke[1,2,3,5]

December 19, 2014

[1]Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX, USA
[2]Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA
[3]Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX, USA
[4]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA
[5]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

Abstract

How do the post-translational modifications change over time during species growth? We sought to answer this question using E. coli grown under glucose starvation conditions. We gathered mass-spectrometry based proteomics data at 9 different time points during exponential and long stationary phases. We then used MODa to perform an unrestricted search of PTMs in this data. We found new and interesting observations from this analysis. First, we show that modified protein seems to be constant, occuring 30%, at all the 9 time points. Second, we show that acetylations,carboxylations and phosphorylations increase, while nitrosylations seem to be constant from exponential to stationary phases.Third, we show that sulfoxide reductases fix MetSO to Met during stationary phase, when oxidative damage should be more. Finally, we found some novel post-translational modifications using this data, a frequent one being phosphogluconylation on serine in R6 ribosomal protein.

1 Introduction

Most of the times, specific biological function of the proteins is identified by the post-translational modification associated with it. Much work has been done in identifying or quantifying the protein coding genes, for example, using mass-spectrometry based proteomics. However, identification or quantification of the functional PTMs is still at its infancy. Mass-spectrometry (MS) based proteomics is the only technique that can be used to characterize these PTMs in large scale. To identify all the PTMs in the dataset, an unrestricted search of all possible PTMs in the data is necessary.

Here we gathered MS data on E. coli proteome grown under glucose-limiting conditions for 2 weeks. The goal is then to search for most of the PTMs and understand how these modifications change over time i.e., from exponential to long-stationary phases (2 weeks).

Peptide search algorithms are then run on these mass-spectrometry data sets to identify peptides and PTMs associated with it. However, because of computational limitations, most of the current search algorithms such as Mascot [25], Sequest [10], OMSSA [12], X!Tandem [6] search for only few variable modifications, i.e., oxidation of methionine etc or a small targeted list of PTMs. This technique of restricted search not only reduces the size of the database to search but in some cases shown to reduce false positives (Cite?). However recently, software for unrestricted search of PTMs is starting to be available. MODa [21] is one such unrestricted search engine, along with others such as TagRecon [7] and Byonic [2]. We used MODa, a naive based multi-blind spectral alignment algorithm, to look for PTMs in the E. coli dataset.

Most of the studies till date limit to capturing proteome information to 1 or 2 time points during exponential and stationary phases for analysis. Here we collected MS data on E. coli proteome at 9 different time points ranging from exponential to long-stationary phases up to 2 weeks under glucose-limiting conditions. Moreover bioinformatics analyses focussed on few PTMs, but to our knowledge never looked at time course of PTMs. Ours is a first attempt to look into time-course of PTMs over 2 weeks of a E. coli REL606 strain grown under glucose starvation.

2 Results

Post-translational modifications determine the specific function of the protein. Identifying all the possible PTMs in a complex sample is limited by both experimental enrichment techniques for PTMs as well as computational limitations. Here, we used a search algorithm MODa, which uses an unrestricted approach to find the post-translational modifications in the data. The E. coli samples are not enriched for any post-translational modifications resulting in comparatively lower identifications of modified peptides. However our goal is not only to identify the PTMs present in the sample, but also understand the time evolution of these PTMs from exponential phase (3,4,5,6 hours of growth) to long-stationary phases (8,24,48, 168 and 336 hours).

2.1 Running MODa on *E. coli* proteome

The mass-spectrometry data gathered at the 9 time points is then searched for PTMs using MODa, a naive based spectral alignment search algorithm, for peptide identifications. At each of these 9 time points, we gathered data for 3 biological replicates. Unlike most of the other peptide identification search algorithms, MODa requires range of mass for PTM identification. So, for our analysis we used a range of -200 to 200 Da which is typical to this search engine. MODa is a multi-blind algorithm i.e., there is no restriction on the number of PTMs identified on a single peptide. However previous studies have shown better sensitivity of the algorithm with the use of 1 mod per peptide. We used 1 mod per peptide for most of the analysis, however we compared our results with using 2 mods per peptide too.

We asked the following questions using our MODa output on peptide identifications: (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? In this project, we also collected mRNA via RNA-seq, lipid profiles and metabolic fluxes using mass-spec methods. We went back and compared our proteomics data with these other kinds of data for further validation, wherever necessary.

2.2 Modified *E. coli* proteome

First, we are interested in understanding how the modified protein or the total number of modifications changes over time i.e., from exponential to long-stationary phases? To estimate how much of the proteome is modified at each time point, we calculated the percentage of the total number of modified peptide-spectral matches in the search output of MODa. 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. 30% of the peptide-spectral matches seem to be consistently modified at all the 9 time points analyzed. OD600 at these 9 time points is givien in supplementary information of this article (see Suppl Figure S1). The 30% of protein getting modified is in agreement with some of the previous studies (Cite papers which show the same numbers).

2.3 Mapping mass-shifts to post-translational modifications

Since MODa outputs mass-shift, we used UNIMOD database to map the mass-shift to the most probable post-translational modification. In this study, we did not investigate mutations. To understand what kinds of PTMs (or mutations) are present in the sample, we plotted the frequency of the mass-shifts. MODa program outputs the frequency of mass-shifts in intervals of 1Da, along with the amino acid on which this mass-shift occurs. Figure 2 shows the mass-shift frequency matrix. (A) shows the distribution at 3 hours and (B) at 24 hours. We did not plot the error bars as this makes the figure a bit clumsy, with error bars appearing at each bin. These are the means of the 3 biological replicates. It is clear that +1 Da peak is the most frequent one. However this modification seems to be 13C peak-picking as it seems to be random and occur on all the amino acids, as shown in the profile (Suppl Figure S5). Next frequent modification is +16Da. A look at profile (Suppl Figure S6) shows majority of oxidations on methionine which is expected. Likewise we mapped frequent and known mass-shifts and looked at analysis of handful of PTMs i.e., acetylation, phosphorylation, carboxylation and nitrosylation in depth.

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

[4] N-terminal acetylation of ectopic recombinant proteins in Escherichia coli

[9] The mechanism of N-terminal acetylation of proteins

[15] Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome,

[13] Acetylation of L12 increases interactions in the Escherichia coli ribosomal stalk complex

[26] Nat3p and Mdm20p are required for function of yeast NatB Nalpha-terminal acetyltransferase and of actin and tropomyosin

[27] Composition and function of the eukaryotic N-terminal acetyltransferase subunits,

[28] N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins

[32] Protein N-terminal acetyltransferases: when the start matters

[33] Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12

[37] Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12 We started out with analysis of acetylation i.e., 42 Da mass-shift. Acetylation occurs in 2 forms i.e., N-alpha acetylation and N-sigma acetylation. It is known that acetylation occurs co-translationally on n-term (N-alpha) and is not reversible, while post-translationally occurs as a reversible modification, mostly on lysine. Irrespective of the type of acetylation, since MODa allowed us to look at the global acetylation level i.e., all possible shifts of 42Da, we plotted the acetylation frequencies at different time points as a function of time (see Figure 3). The total number of acetylations seem to go up from exponential to stationary phases. In our analysis, most of the acetylations seem to be protein n-term acetylations. Either this acetylation occurs on the 1st methionine amino acid or it happens on the 2nd amino acid after the excision of the 1st methionine. Figure 3 also shows the total n-term acetylations and the total serine acetylations. Later, we show that the serine acetylations seem to happen ¿99% of the times at n-term.

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.

Out of 3919 total acetylations at all 9 time points combined, 3373 (86%) of them seem to occur at n-term of the protein. A look-up of the amino acid position for n-term acetylations on ID'ed peptides showed that these were the 2nd amino acid in most (quote?) of the cases. May be add about the signal peptides here. Also, most of the acetylations seem to occur on serine (2358 out of 3919). Among n-term protein acetylations, 2350 (70%) happened on serine, 483 (14.3%) happened on alanine and 260 (7.7%) were on threonine. These were the top 3 frequent n-term acetylations. A literature search also showed us the same frequently occuring n-term acetylated amino acids in E. coli cite.

Colin Brown's analysis could go here

2.5 *E. coli* Phosphorylations, carboxylations, nitrosylations and other PTMs

E. coli phosphoproteomics [20] Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation

E. coli phosphoproteome dynamics [30] Global dynamics of the Escherichia coli proteome and phosphoproteome during growth in minimal medium,

E. coli nitrosylation [29] Endogenous protein S-Nitrosylation in E. coli: regulation by OxyR,

Next we analyzed other PTMs that were previously shown to be key in E. coli. In particular, we looked at phosphorylation, carboxylation, nitrosylation and oxidation. As mentioned earlier, a key point to note in this analysis is that there is no enrichment done for any of these PTMs, hence the coverage will be smaller than expected. Typical 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 and look at how they change over time. Figure 4 shows the phosphorylations at different points of the growth curve. There are 2 interesting observations seen in this curve. First, the number of phosphorylations seem to

be small as expected, as these are not enriched for phosphorylations and moreover, the number of phosphorylations seem to be comparatively lower in E. coli. Second, and probably an interesting observation is that there is an increase of the number of phosphorylations from exponential to stationary phases. Such pattern is also previously shown in phosphoproteomic studies (cite?).

Similarly we looked at the carboxylations (See Suppl Figure S2)and nitrosylations (See Suppl Figure S3), as they are shown in the past to be important in E. coli. Finally, we also plotted the profiles of Na and K adducts (see Suppl Figure S4). Most of the software have an easy option to include these adducts and hence we recommend this giving a 3-4% increase in identifications. They seem to occur on all the amino acids, except basic charge amino acids, as expected.

Pyroglutamate conversion and Succinylation: Glutamine conversion at n-term of the peptide to pyroglutamate is a well known PTM. Suppl Figure 7 shows that this conversion seems to be more or less same at different phases of the growth. (Add data about succinylation)

2.6 Oxidative damage and repair in *E. coli*

[31] Protein oxidation and aging

[36] Oxidation of methionyl residues in proteins: tools, targets, and reversal

[14] Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase, Msr fixing MetSO to Met [3] Enzymatic reduction of protein-bound methionine sulfoxide

[11] Methionine sulfoxide reductases protect Ffh from oxidative damages in Escherichia coli

[34] Crystal structure of the Escherichia coli peptide methionine sulphoxide reductase at 1.9 A resolution

[35] Crystallization and preliminary X-ray diffraction studies of the peptide methionine sulfoxide reductase from Escherichia coli

[38] Origin and evolution of the protein-repairing enzymes methionine sulphoxide reductases This fixation of MsrA is helpful [1] Enzymatic reduction of oxidized alpha-1-proteinase inhibitor restores biological activity

[8] HIV-2 protease is inactivated after oxidation at the dimer interface and activity can be partly restored with methionine sulphoxide reductase

Finally, we investigated oxidative damage by looking at the oxidations identified by MODa program i.e.., mass-shifts of +16Da. Since we capped the cysteine to avoid formation of disulphide bonds and oxidations, we did not expect these residues to be oxidized. There are at least 10 amino acids that get oxidized and other than cysteine, the next expected amino acid is methionine. 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. There is a class of sulfoxide reductases previously characterized that fix MetSO back to Met. In *E. coli*, these are MsrA and yeaA. When we plotted the protein abundances of these reductases, we see some increase in their protein frequences from exponential to stationary phases, as shown in Figure 6. To see if there is an increase in transcript level of these genes, we also plotted the mRNA abundances using the RNA-seq data on the same sample. There is not much change in the profiles of transcripts indicating some kind of post-translational regulation. Using the protein abundances, it is clear that the fixation of MetSO to Met increased over time.

2.7 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 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.

3 Discussion

Limitations: Large-scale analysis of PTMs [23]

PTM network motif [22] Global, in vivo, and site-specific phosphorylation dynamics in signaling networks

PTM cross talk [24] Identification of enriched PTM crosstalk motifs from large-scale experimental data sets

MODa application [16] ROSics: Chemistry and proteomics of cysteine modifications in redox biology

N-terminal processing [17] N-Terminal modifications of the 19S regulatory particle subunits of the yeast proteasome

Mass-spec E. coli proteome [18] Deep coverage of the Escherichia coli proteome enables the assessment of false discovery rates in simple proteogenomic experiments,

MODa application in urine proteomics [19] Unrestrictive identification of post-translational modifications in the urine proteome without enrichment,

Large-scale PTM analysis is still at its infancy because of limitations in both experimental enrichment techniques of PTMs as well as the computational search algorithms that have to look at different combinations to identify the PTMs. Here we used MODa, a naive based multi-blind spectral alignment search algorithm for an unrestricted identification of PTMs. For our analysis, we used E. coli proteome data obtained under glucose-limiting conditions for over 2 weeks. This is a first study to look at E. coli growth for a period of 2 weeks, considered as long-stationary phase in this work. These are some of the key findings in this study. Our analysis showed that 1/3rd of the peptide spectral matches found are modified consistently at all the 9 time points analyzed (from 3 hours to 2 weeks). An increase in acetylation level from exponential to stationary phase is shown in this work. Likewise, even though phosphorylation levels are low, they seem to increase from exponential to stationary phases. We also saw a decrease in oxidation levels from exponential to stationary phase, with sulfoxide reductases playing a role in fixing Methionine oxide back to methionine. n-term acetylations are frequent, with serine n-term acetylation happening 75% of all n-term acetylations, all the times on the 2nd amino acids of the protein, after methionine excision. Phosphorylation increased from exponential to stationary phases, while carboxylations and nitrosylations remained constant throughout the time course of E. coli growth. Na and K adducts seem to happen 3-4% of the time. Finally, a novel serine modification i.e., phosphogluocnylation of serine seems to happen on ribosomal protein S6 frequently.

Most of the peptide identification search algorithms require a list of PTMs to search for and generally this list is limited to 6. However there are hundreds of PTMs known to date, and well

documented in databases like UNIMOD, RESID. So, we used a naive based search algorithm that outputs mass-shifts, instead of the PTMs in the peptide-spectral match. Then we used UNIMOD to identify the PTM that most probably matches the mass-shift. However we did not try and match all the mass-shifts, but investigated in detail the frequent and well known mass-shifts identified by MODa. This resulted in analysis of +1 Da mass-shift (C13 peak detection), oxidation, acetylation, Na and K adducts, along with some widely studied PTMs in phosphorylation, carboxylation and nitrosylation. This program is previously used for similar large-scale analysis with urinary proteomics and they identified novel PTMs. However the study used 2 programs and considered the overlap of PTMs as highly confident. Instead here, our focus is not to identify highly-confident PTMs later used as biomarkers, but to get a wider coverage at a lower FDR of 1% and look at the time course or evolution of these PTMs during the entire 2 weeks of E. coli growth.

There are some limitations with respect to how we did MODa searches. One of the limitations in the searches performed in the current work is that we used the default mass-range search between -200 and 200 Da (and one other search with -100 and 300 Da range). So, we miss out on larger PTMs i.e., polyubiquitination tails etc. It was shown in the past that looking for multiple PTMs on the same peptide leads to higher false positives. So, we ran our searches looking for only 1 possible modification on the peptide. After the peptides are ID'ed, generally the PTMs are validated by a 2nd round by using programs like Ascore etc. However here, we did not do any 2nd round of validating peptides, as MODa is shown to ID high-confidence PTM identications in its search. We then used MODa output to look at the global level of PTMs, instead of looking at the peptide or at the protein level. We did not do any PTM level quantitation or try to understand the specific PTM stoichiometry, as these require sophisticated experimental instrumentation, protocol and the algorithms to characterize the PTMs associated with proteins.

Nice review article on E. coli proteomics by different technologies including MS can be found here: Cite this paper: The Escherichia coli Proteome: Past, Present, and Future Prospects† Mee-Jung Han1 and Sang Yup Lee1,2,*

Directly taken from the above paper for my future analyses: For example, SspA expression increased with decreasing growth rate and was induced by glucose, nitrogen, phosphate, or amino acid starvation. Furthermore, the proteome profiles during the exponential growth phase showed that the expression levels of at least 11 proteins were altered in sspA mutant strains (314). These findings indicate that SspA acts as a transcription factor and is essential for starvation stress-induced tolerance (e.g., stationary phase) in E. coli.

Copied from the same paper At the onset of glucose starvation, cyclic AMP and its receptor protein (cAMP-CRP) were found to play important roles in the expression of a number of genes. An early 2-DE study identified five glucose-responsive outer membrane proteins (four upregulated and one downregulated) (186). A comparison with membrane proteins from mutant strains revealed that two of the upregulated proteins were the receptors for lambda and T6, and coelectrophoresis of the outer membrane fraction identified the downregulated protein as OmpA. The glucose starvation stimulon was further examined using 2-DE followed by comparison to the E. coli gene-protein database (218). Members of this stimulon were found to include enzymes of the Embden-Meyerhof-Parnas pathway, phosphotransacetylase (Pta) and acetate kinase (AckA) in the acetic acid pathway, and formate transacetylase. Trichloroacetic acid cycle enzymes were repressed, whereas enzymes involved in acetate and formate production and the Embden-Meyerhof-Parnas pathway were induced. These modulations suggest that a glucose-starved cell increases the relative flow of carbon through the Pta-AckA pathway. Indeed, pta and pta-ackA mutants were

found to be impaired in their abilities to survive glucose starvation, indicating that the capacity to synthesize acetyl phosphate, an intermediate of this pathway, is indispensable for glucose-starved cells. The pta mutant failed to induce several proteins of the glucose starvation stimulon. More recently, proteome studies revealed that glucose limitation upregulates the levels of proteins such as AceA, AldA, ArgT, AtpA, DppA, GatY, LivJ, MalE, MglB, RbsB, UgpB, and YdcS (311). Of these, ArgT, DppA, LivJ, MalE, MglB, RbsB, UgpB, and YdcS are periplasmic binding proteins of the ABC transporters, suggesting that in addition to the central metabolism proteins, periplasmic binding proteins are involved in the carbohydrate and amino acid uptakes that are important during glucose limitation.

Also, A functional relA gene is required for sspA to affect protein synthesis. (taken from another paper) Interesting to find PTMs on these proteins?

Most of the times, proteins act in complexes to perform a specific function. In such process, 1 or few amino acids of a protein interact with other residues of another protein, generally through the PTMs. So, large-scale analysis of PTMs such as this work would help us better understand the fine granularity at PTM level that is responsible for a particular mechanism, such as multiple phosphorylations in the case of signalling cascades.

Mass-spec analysis of human soluble protein complexes has revealed a large number of conserved complexes along with thousands of protein-protein interactions. Large-scale analysis of PTMs on these kinds of proteomics data would help to understand these interactions at PTM level. Such annotations could then be integrated with databases such as NCBI CDD Across diverse species (cite Emili/Marcotte collaboration cell papers) and talk about PTM annotation in databases like CDD? Cite the paper on conservation of phosphorylations etc on CDDs, but extend to other PTMs?

Current whole-cell models [5] are trying to integrate diverse kinds of OMICS data i.e., transcriptomics, proteomics not only to refine the existing models but also get the response of the models close to the experimental metabolic flux measurements. Here, we argue that including the modification information (i.e., number of modified proteins to that of the unmodified version) will improve the existing methodologies.

4 Conclusions

The modified protein seems to be 30% during exponential as well as the long stationary phases. Acetylation, in particular n-term acetylation seems to go up from exponential to stationary phases. Surprisingly oxidation seemed to go down from exponential to stationary phases, owing to sulfoxide reductases playing a role in fixing methionine sulfoxide back to methionine. A novel phosphoserinegluconylaiton on ribosomal protein seem to happen frequently. Finally, we would like to conclude that unrestricted search engines can be used to identify frequently occuring PTMs, which can then be used with restricted search algorithms to improve the sensitivity of the ID'ed peptides.

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 where 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 where 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 where 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 where then ultrafiltrated to remove insoluble and undigested material using Amicon Ultra MWCO 10kD spin-caps and finally concentrated and purified by C18 filtration. Liquid chormotography 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.

We manually mapped mass-shifts outputted by MODa to those of UNIMOD mappings. So, we limited our focus to the well known PTMs and the frequent mass-shifts obtained from MODa. For example, even though carboxylations and nitrosylations seemed rare (from frequencies of mass-shifts), since we know the mass-shift and the expected amino acids on which this mod happens, we considered that. Likewise, frequently occuring mass-shifts were mapped to NA and K adducts and hence we considered those too in our analysis. We looked at the amino acid profile for each of the mass-shifts considered, to see which amino acid gets modified frequently. These seem to agree well with the literature i.e., oxidation happens on M frequently, when cysteine is capped with carbamidomethylation etc.

(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.W.B, M.D.P, C.O.W and E.M.M. Wrote the paper: V.S, C.W.B, D.R.B, C.B, M.D.P, 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 the Bioinformatics Consulting Group (BCG) and the Texas Advanced Computing Center (TACC) at UT for high-performance computing resources.

References

1. W. R. Abrams, G. Weinbaum, L. Weissbach, H. Weissbach, and N. Brot. Enzymatic reduction of oxidized alpha-1-proteinase inhibitor restores biological activity. *Proc Natl Acad Sci U S A*, 78(12):7483–6, 1981.

2. M. Bern, Y. J. Kil, and C. Becker. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics*, Chapter 13:Unit13 20, 2012.

3. N. Brot, L. Weissbach, J. Werth, and H. Weissbach. Enzymatic reduction of protein-bound methionine sulfoxide. *Proc Natl Acad Sci U S A*, 78(4):2155–8, 1981.

4. E. Charbaut, V. Redeker, J. Rossier, and A. Sobel. N-terminal acetylation of ectopic recombinant proteins in escherichia coli. *FEBS Lett*, 529(2-3):341–5, 2002.

5. M. W. Covert, N. Xiao, T. J. Chen, and J. R. Karr. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 24:2044–50, 2008.

6. R. Craig and R. C. Beavis. Tandem: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9):1466–7, 2004.

7. S. Dasari, M. C. Chambers, R. J. Slebos, L. J. Zimmerman, A. J. Ham, and D. L. Tabb. Tagrecon: high-throughput mutation identification through sequence tagging. *J Proteome Res*, 9(4):1716–26, 2010.

8. D. A. Davis, F. M. Newcomb, J. Moskovitz, P. T. Wingfield, S. J. Stahl, J. Kaufman, H. M. Fales, R. L. Levine, and R. Yarchoan. Hiv-2 protease is inactivated after oxidation at the dimer interface and activity can be partly restored with methionine sulphoxide reductase. *Biochem J*, 346 Pt 2:305–11, 2000.

9. H. P. Driessen, W. W. de Jong, G. I. Tesser, and H. Bloemendal. The mechanism of n-terminal acetylation of proteins. *CRC Crit Rev Biochem*, 18(4):281–325, 1985.

10. J. K. Eng, A. L. McCormack, and J. R. Yates. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*, 5(11):976–89, 1994.

11. B. Ezraty, R. Grimaud, M. El Hassouni, D. Moinier, and F. Barras. Methionine sulfoxide reductases protect ffh from oxidative damages in escherichia coli. *EMBO J*, 23(8):1868–77, 2004.

12. L. Y. Geer, S. P. Markey, J. A. Kowalak, L. Wagner, M. Xu, D. M. Maynard, X. Yang, W. Shi, and S. H. Bryant. Open mass spectrometry search algorithm. *J Proteome Res*, 3(5):958–64, 2004.

13. Y. Gordiyenko, S. Deroo, M. Zhou, H. Videler, and C. V. Robinson. Acetylation of l12 increases interactions in the escherichia coli ribosomal stalk complex. *J Mol Biol*, 380(2):404–14, 2008.

14. R. Grimaud, B. Ezraty, J. K. Mitchell, D. Lafitte, C. Briand, P. J. Derrick, and F. Barras. Repair of oxidized proteins. identification of a new methionine sulfoxide reductase. *J Biol Chem*, 276(52):48915–20, 2001.

15. A. O. Helbig, S. Gauci, R. Raijmakers, B. van Breukelen, M. Slijper, S. Mohammed, and A. J. Heck. Profiling of n-acetylated protein termini provides in-depth insights into the n-terminal nature of the proteome. *Mol Cell Proteomics*, 9(5):928–39, 2010.

16. H. J. Kim, S. Ha, H. Y. Lee, and K. J. Lee. Rosics: Chemistry and proteomics of cysteine modifications in redox biology. *Mass Spectrom Rev*, 2014.

17. Y. Kimura, Y. Saeki, H. Yokosawa, B. Polevoda, F. Sherman, and H. Hirano. N-terminal modifications of the 19s regulatory particle subunits of the yeast proteasome. *Arch Biochem Biophys*, 409(2):341–8, 2003.

18. K. Krug, A. Carpy, G. Behrends, K. Matic, N. C. Soares, and B. Macek. Deep coverage of the escherichia coli proteome enables the assessment of false discovery rates in simple proteogenomic experiments. *Mol Cell Proteomics*, 12(11):3420–30, 2013.

19. L. Liu, X. Liu, W. Sun, M. Li, and Y. Gao. Unrestrictive identification of post-translational modifications in the urine proteome without enrichment. *Proteome Sci*, 11(1):1, 2013.

20. B. Macek, F. Gnad, B. Soufi, C. Kumar, J. V. Olsen, I. Mijakovic, and M. Mann. Phosphoproteome analysis of e. coli reveals evolutionary conservation of bacterial ser/thr/tyr phosphorylation. *Mol Cell Proteomics*, 7(2):299–307, 2008.

21. S. Na, N. Bandeira, and E. Paek. Fast multi-blind modification search through tandem mass spectrometry. *Mol Cell Proteomics*, 11(4):M111 010199, 2012.

22. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3):635–48, 2006.

23. J. V. Olsen and M. Mann. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics*, 12(12):3444–52, 2013.

24. M. Peng, A. Scholten, A. J. Heck, and B. van Breukelen. Identification of enriched ptm crosstalk motifs from large-scale experimental data sets. *J Proteome Res*, 13(1):249–59, 2014.

25. D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–67, 1999.

26. B. Polevoda, T. S. Cardillo, T. C. Doyle, G. S. Bedi, and F. Sherman. Nat3p and mdm20p are required for function of yeast natb nalpha-terminal acetyltransferase and of actin and tropomyosin. *J Biol Chem*, 278(33):30686–97, 2003.

27. B. Polevoda and F. Sherman. Composition and function of the eukaryotic n-terminal acetyltransferase subunits. *Biochem Biophys Res Commun*, 308(1):1–11, 2003.

28. B. Polevoda and F. Sherman. N-terminal acetyltransferases and sequence requirements for n-terminal acetylation of eukaryotic proteins. *J Mol Biol*, 325(4):595–622, 2003.

29. D. Seth, A. Hausladen, Y. J. Wang, and J. S. Stamler. Endogenous protein s-nitrosylation in e. coli: regulation by oxyr. *Science*, 336(6080):470–3, 2012.

30. N. C. Soares, P. Spat, K. Krug, and B. Macek. Global dynamics of the escherichia coli proteome and phosphoproteome during growth in minimal medium. *J Proteome Res*, 12(6):2611–21, 2013.

31. E. R. Stadtman. Protein oxidation and aging. *Science*, 257(5074):1220–4, 1992.

32. K. K. Starheim, K. Gevaert, and T. Arnesen. Protein n-terminal acetyltransferases: when the start matters. *Trends Biochem Sci*, 37(4):152–61, 2012.

33. S. Tanaka, Y. Matsushita, A. Yoshikawa, and K. Isono. Cloning and molecular characterization of the gene riml which encodes an enzyme acetylating ribosomal protein l12 of escherichia coli k12. *Mol Gen Genet*, 217(2-3):289–93, 1989.

34. F. Tete-Favier, D. Cobessi, S. Boschi-Muller, S. Azza, G. Branlant, and A. Aubry. Crystal structure of the escherichia coli peptide methionine sulphoxide reductase at 1.9 a resolution. *Structure*, 8(11):1167–78, 2000.

35. F. Tete-Favier, D. Cobessi, G. A. Leonard, S. Azza, F. Talfournier, S. Boschi-Muller, G. Branlant, and A. Aubry. Crystallization and preliminary x-ray diffraction studies of the peptide methionine sulfoxide reductase from escherichia coli. *Acta Crystallogr D Biol Crystallogr*, 56(Pt 9):1194–7, 2000.

36. W. Vogt. Oxidation of methionyl residues in proteins: tools, targets, and reversal. *Free Radic Biol Med*, 18(1):93–105, 1995.

37. A. Yoshikawa, S. Isono, A. Sheback, and K. Isono. Cloning and nucleotide sequencing of the genes rimi and rimj which encode enzymes acetylating ribosomal proteins s18 and s5 of escherichia coli k12. *Mol Gen Genet*, 209(3):481–8, 1987.

38. X. H. Zhang and H. Weissbach. Origin and evolution of the protein-repairing enzymes methionine sulphoxide reductases. *Biol Rev Camb Philos Soc*, 83(3):249–57, 2008.

Figures

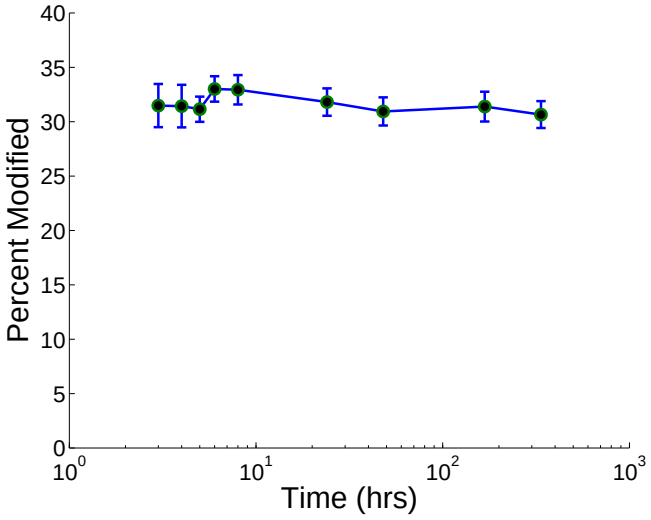

Figure 1. *E. coli* modified proteome. The total number of modified peptide-spectral matches seem to be constant at 30% for all 9 time points. .

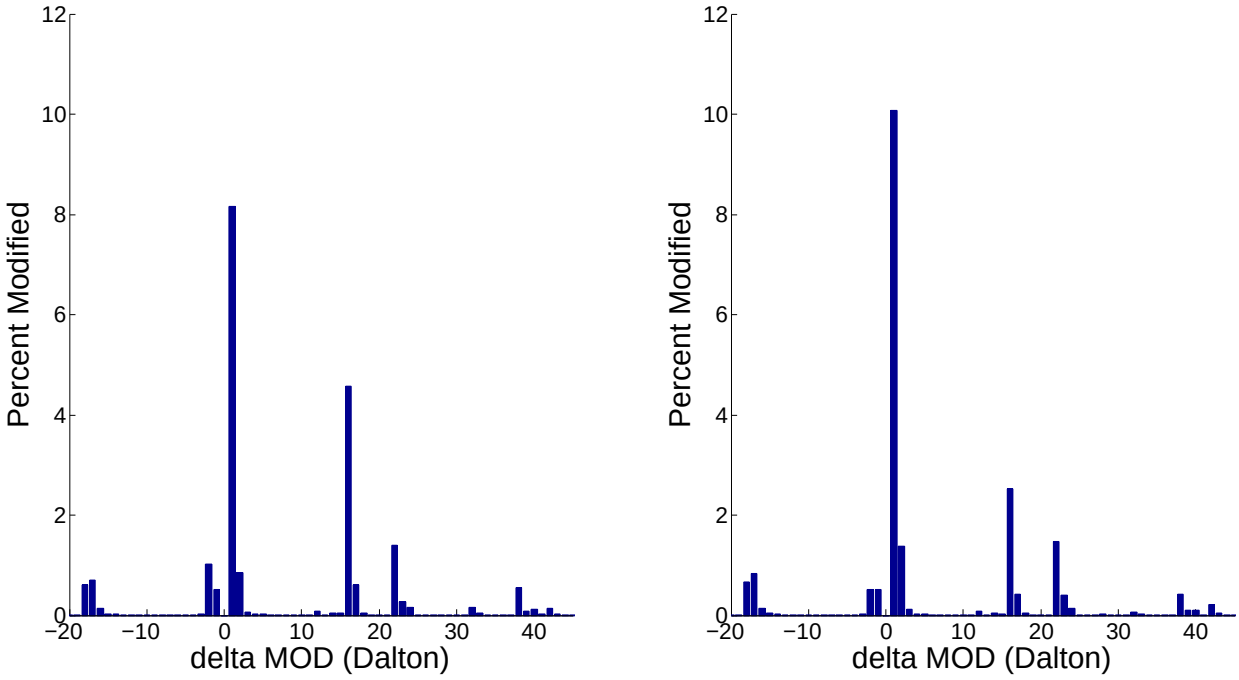

Figure 2. MODa outputs mass-shifts. A naïve based algorithm like MODa can alleviate the requirement of guessing PTMs beforehand. However MODa outputs mass-shifts on the amino acids. We can then use PTM databases like UNIMOD to map the mass-shift to the most probable PTM. (A) and (B) are the frequencies of the mass-shifts observed at 3 hours and 2 weeks of the E. coli growth..

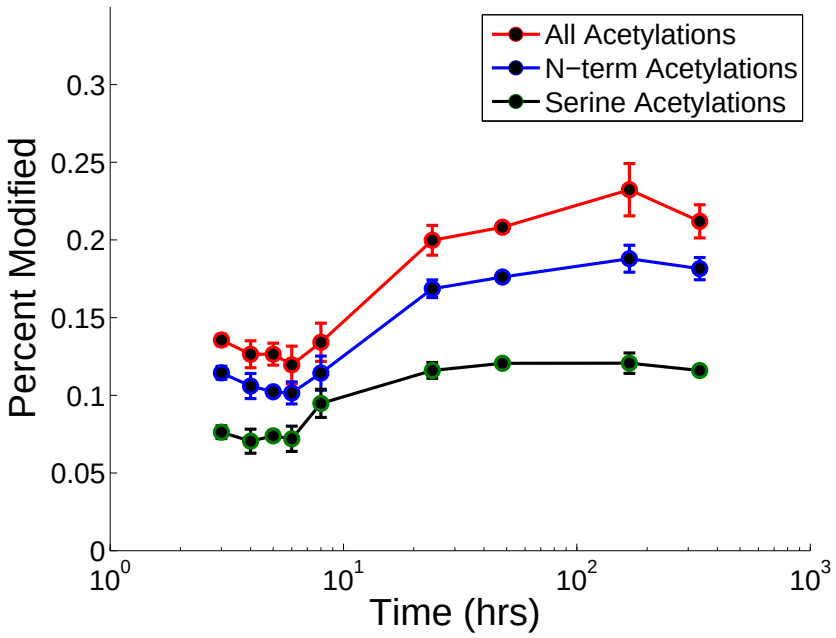

Figure 3. Protein n-term acetylations are dominant. Total number of acetylations as well as the n-term/serine acetylations seem to go up over 2 weeks. E. coli grown on glucose generally tend to accumulate acetate, perhaps this resulted in increase in acetylations..

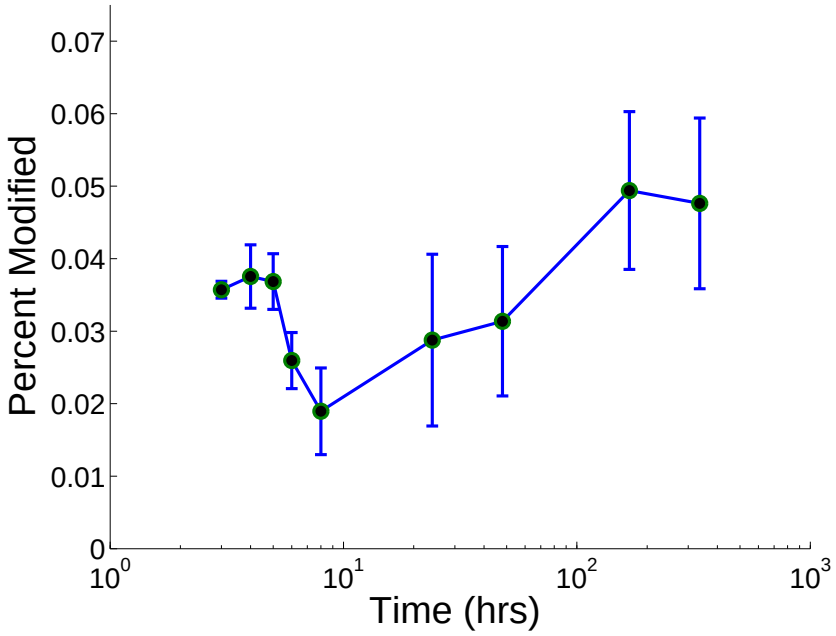

Figure 4. Phosphorylations are rare. Phosphorylations seem to be low and tend to increase during last week of growth. 2 frequently phosphorylated proteins in MODa search output are phosphoglucomutase and elongation factor Tu.

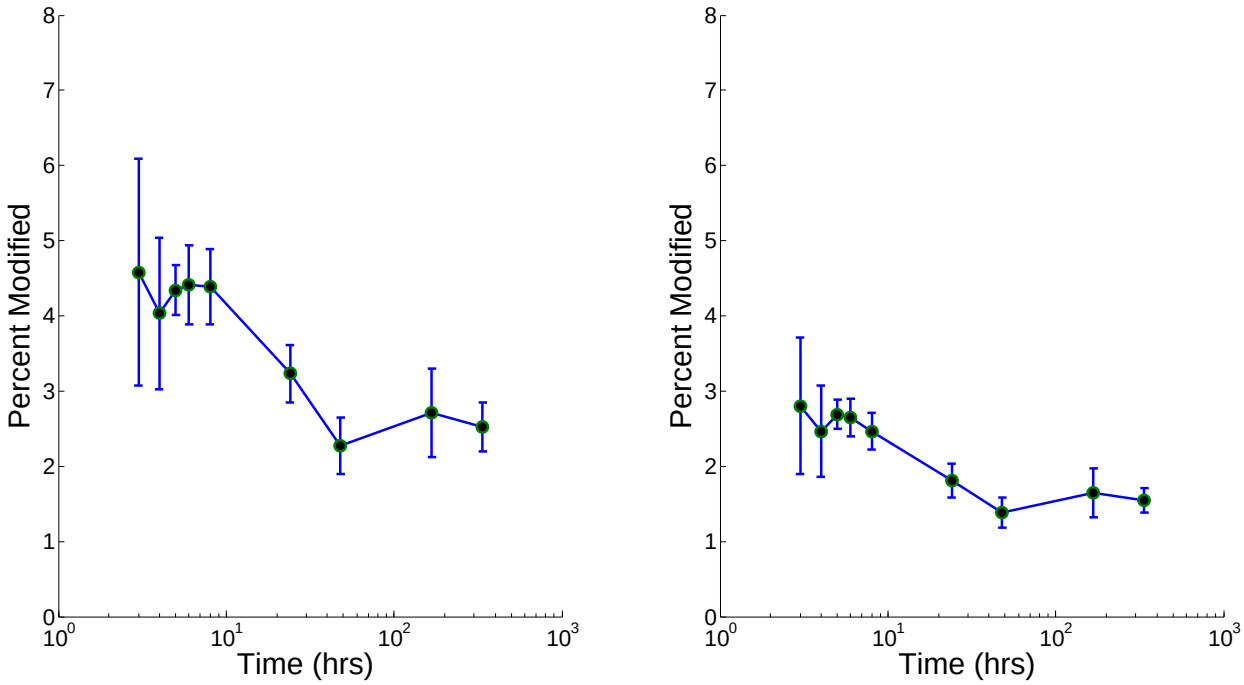

Figure 5. Oxidations go down over 2 weeks. (A) Total number of oxidations seem to go down from exponential to stationary phases. (B) The same trend follows for methionine oxidations too.

Supplementary Figures

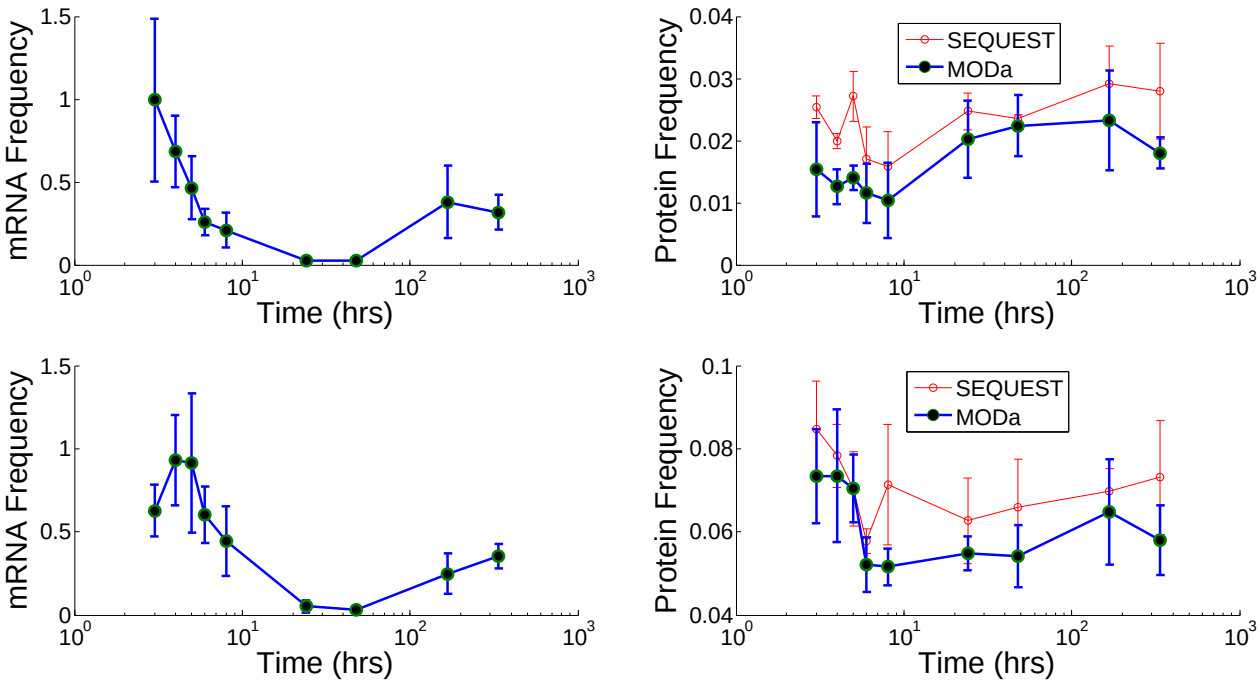

Figure 6. Relative mRNA and protein abundances of methionine sulfoxide reductases MsrA and MsrB. (A) and (C) mRNA abundances of MsrA and MsrB. The increase of mRNA abundance in stationary phase is not prominent. (B) Protein abundances from 2 different programs seem to agree that MsrA and MsrB are probably fixing methionine sulfoxide to methionine. One of the sites oxidized on MsrA is FQAA[M+16]LAADDDR.

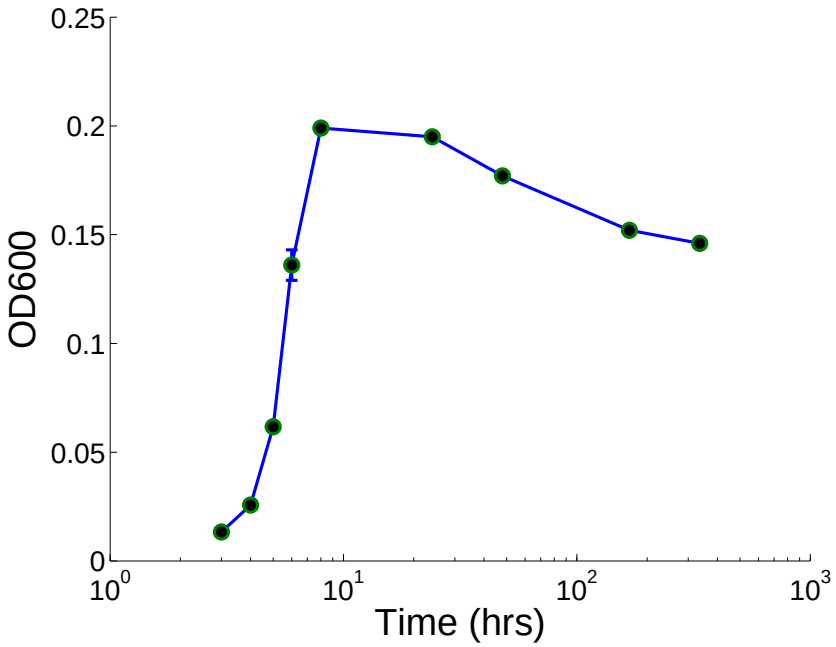

Figure S1: OD600 curve . Growth curve (OD600) of REL606 under glucose starvation conditions.

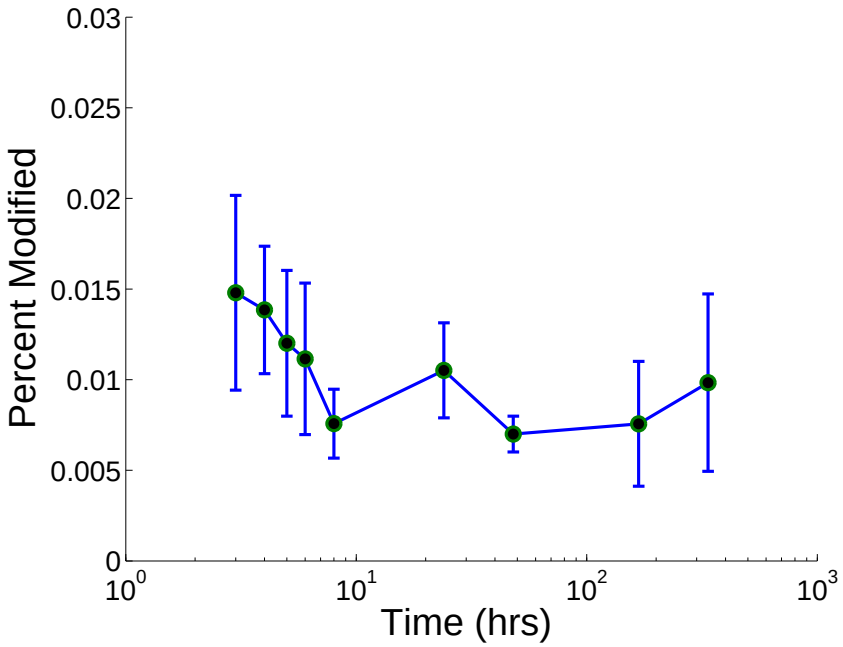

Figure S2: E. coli Nitrosylations. Nitrosylations seem to go down.

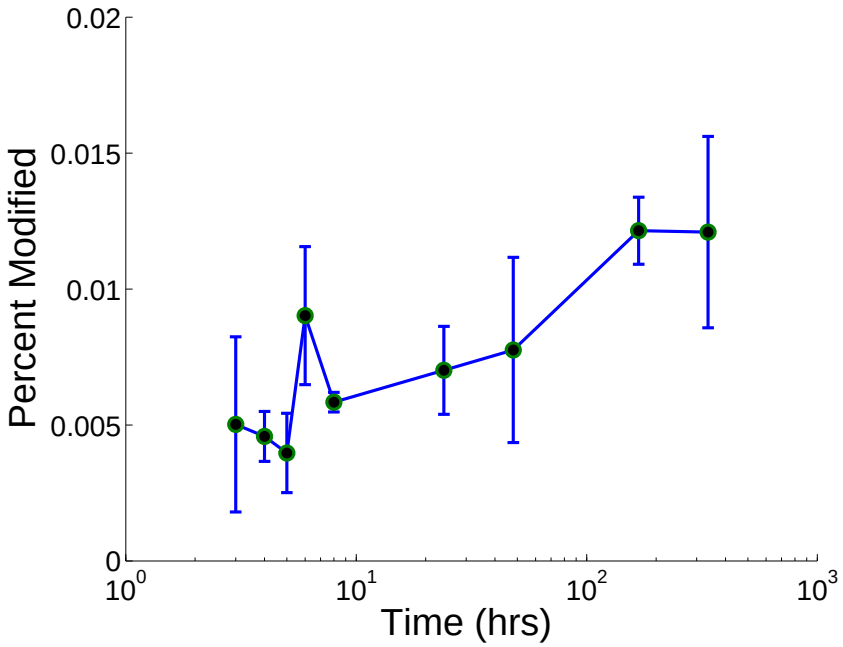

Figure S3: E. coli Carboxylations . Carboxylations seem to go up.

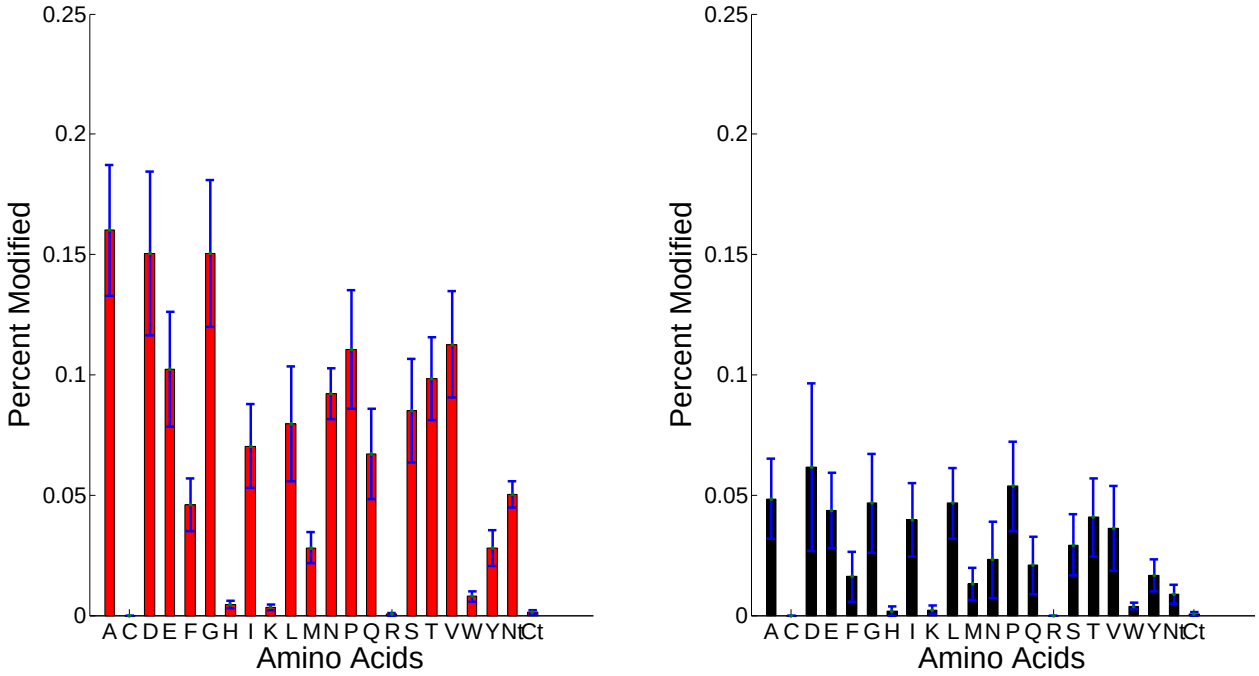

Figure S4: Na and K adducts. Na and K adducts seem to happen on all amino acids except those that are basic and carry some charge, as expected.

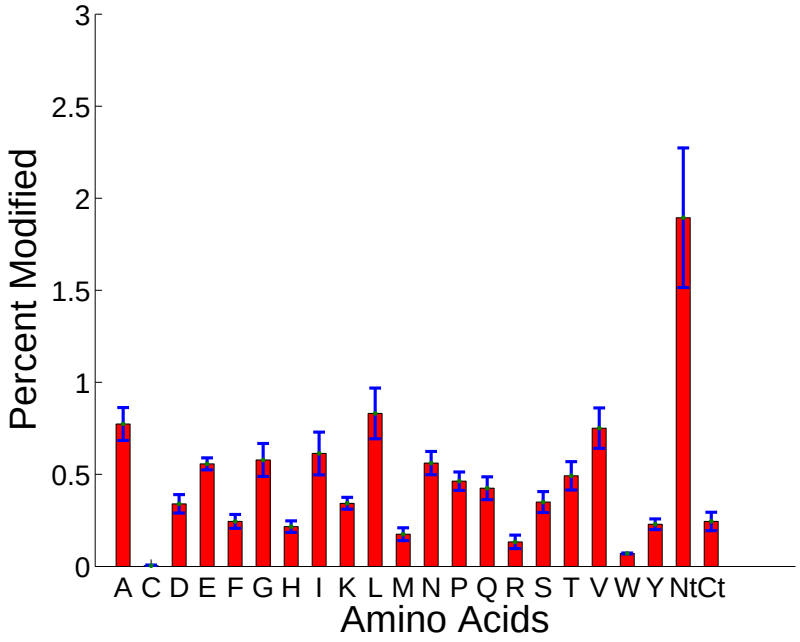

Figure S5: +1Da shift occurs randomly. We cannot infer that this is deamidation as it occurs randomly on all amino acids, inferring it is mostly 13C peak picking as previously shown in many studies.

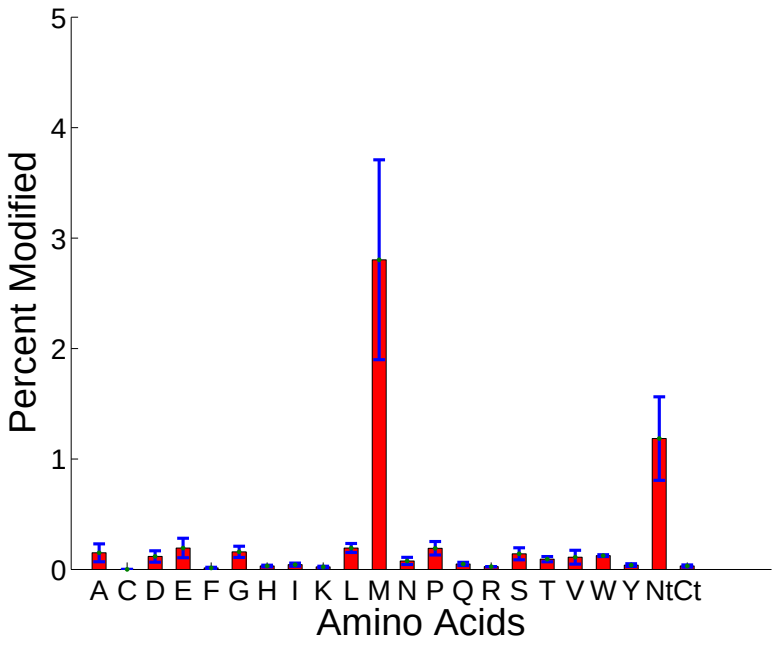

Figure S6: Oxidation is dominant on methionine. Even though many amino acids could be oxidized, in this data set, oxidation seems to occur primarily on methionine, as expected.

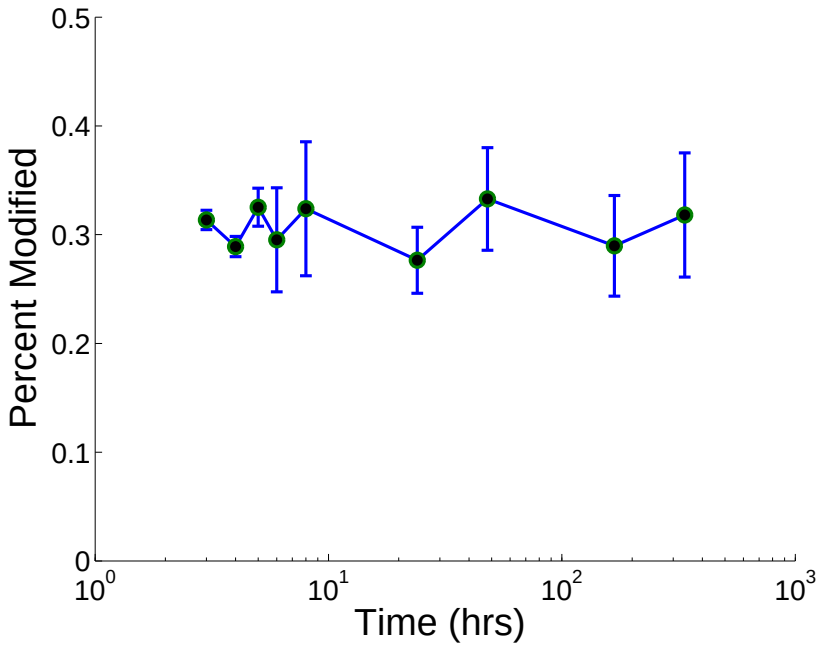

Figure S7: Glutamine to pyroglutamate conversion. Glutamine to pyroglutamate happens to stabilize the protein. This conversion seems to be consistent across both the exponential and stationary phases.