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], Maria D. Person[4],
Jeffrey E. Barrick[1,2,3,5], Edward M. Marcotte[1,2,3,5], and Claus O. Wilke[1,2,3,6]

April 30, 2015

[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]College of Pharmacy, The University of Texas at Austin, Austin, TX, USA
[5]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA
[6]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

Abstract

How do the post-translational modifications [PTMs] change over time during microbial growth under nutrient limiting conditions? We sought to answer this question using *E. coli* grown under glucose starvation conditions. We generated mass-spectrometry based proteomics data at 9 different time points ranging from exponential to long stationary phases i.e., upto 2 weeks. We then ran MODa on this data for an unrestricted search of PTMs. There were several interesting observations in this MODa analysis. First, we show that the amount of modified protein seems to be constant, occuring approx. 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, protecting *E. coli* from oxidative damage. Fourth, we found some new n-terminal acetylation targets that were not identified earlier in *E. coli*. Finally, we found some novel post-translational modifications in this data, a frequent one being phosphogluconylation on serine in R6 ribosomal protein.

1 Introduction

Mass-spectrometry based proteomics [21] offers a unique way to characterize different post-translational modifications [PTMs] associated with proteins. PTMs are key in determining protein function, localization and regulation. Hundreds of PTMs are known *in vivo*, however most of the current

1

studies focus on identifying only few PTMs from mass-spec data. This is both because of computational limitations as well as the inefficiency of enrichment techniques to characterize or identify all PTMs. However with the improvement of search algorithms for an unrestricted search of PTMs, identifying multiple PTMs in *tandem* has become a possibility. In this work, we used MODa [26], a naive based spectral alignment algorithm to identify different kinds of PTMs in *E. coli* REL606 strain.

In a typical mass-spectrometry based proteomics experiment, proteins are first digested into peptides and then the sample is enriched for any PTMs of interest. This sample is then analyzed by mass-spectrometry. Peptide search algorithms are then run on this mass-spectrometry data to computationally identify peptides and PTMs associated with it. These search algorithms generall fall into 3 categories: (1) Sequence or database search algorithms, (2) de novo sequence search and (3) Hybrid approach that is a mixture of partial de novo search followed by a database search. Most of the current search algorithms such as Mascot [32], Sequest [12], OMSSA [14], X!Tandem [8] search for only few variable modifications, i.e., oxidation of methionine etc or a small targeted list of PTMs. This technique of restricted search reduces the size of the database to search and hence alleviates lot of computational limitations [24]. However this limited search identifies only PTMs that were used in the target list, limiting the analysis to either the previously known PTMs or target PTMs. However, recently, software for unrestricted search of PTMs is starting to be available. MODa [26] is one such unrestricted search engine, along with others such as TagRecon [9] and Byonic [2]. We used MODa, a naive based multi-blind spectral alignment algorithm, to look for PTMs in our *E. coli* dataset. A few studies that used MODa in the past looked at hand-ful of interesting PTMs in depth such as [18], or looked at wider coverage of PTMs such as [22]. Our interest is the wider coverage of PTMs and look in depth at both the frequently occuring mass-shifts and the well known PTMs mapped from known mass-shifts.

In our current work, we generated mass-spectrometry data for *E. coli* grown under glucose limiting conditions. *E. coli* grown under these nutrient limiting conditions generally behave differently at different phases of their growth i.e., during exponential and stationary phases. During the stationary phase, stress response proteins act because of the conditions casued by limiting nutrients, change in pH value, accumulation of toxic substances in the flasks etc [28]. Most of the studies till now focussed on identifying proteins that were differentially expressed under these different conditions of growth. However, there are very few studies that looked at the PTM level information during such nutrient-limiting growth. In one study that looked at phosphoproteome [36], the number of phosphorylation sites as well as their occupancy levels have been shown to increase from exponential to stationary phase, suggesting likely role of this PTM in stationary phase. In the same study, the authors showed that protein abundance of SspA, a stress protein seemed to be constant during the time course of growth, while the number of phosphorylated sites on the same protein increased. This clearly suggests PTM level quantitation to be key to understand how *E. coli* responds to stress. In a recent study by Soufi et. al., [37], the authors investigated the *E. coli* proteome dynamics at 7 different time points (exponetial to extended stationary phases) of growth. They looked at protein abundances, copy numbers and also investigated briefly on PTMs. In their work, the authors looked at the pair-wise correlation of the PTMs identified in different stages of growth. Even though the authors looked at few PTMs, they did not do PTM level quantitation. Our work is focussed not only on doing an unrestricted search of PTMs, but understanding the PTM levels during the exponential and long-stationary phases of *E. coli* growth.

Few other studies of PTMs focussed on multiple PTMs (citation?), but they looked at only

one "snapshot" of interest i.e., a single time point during the growth. Since it has already been shown that during different phases of growth, different proteins will be expressed, a "snapshot" of PTM analysis does not provide a clearer picture on the significance or the function associated by these PTMs. Our current work of using multiple time points during the growth curve or looking at multiple "snapshots" along with an unrestricted search of PTMs provides a holistic view of proteome function.

2 Results

2.1 Running MODa on *E. coli* proteome

Post-translational modifications, in most cases, determine the specific function of the protein. Identifying all the PTM combinations or the set of PTMs associated with all the proteins (i.e., entire proteome) is limited by experimental enrichment techniques for PTMs as well as computational limitations. The current analysis of *E. coli* dataset is not enriched for any PTMs. This means that this dataset is suitable for hunting for different kinds of PTMs, instead of focussing on particular PTMs. Most of the current studies focus on a single PTM coverage, such as generally done in phosphoproteomics studies, glycosylation studies or acetylation studies, where the sample is enriched for a particular PTM.

We have mass-spectrometry data at 9 different points of the OD600 curve (see Suppl Figure S1). At each time point, there were 3 biological replicates. We then used MODa on each of these data sets to identify the PTM level information. MODa is a sequence search algorithm that takes as an input mass-spectrometry data and identifies peptides and PTMs associated with the proteins. MODa is a naive based search algorithm that uses an unrestricted approach to find the post-translational modifications in the data. The parameters used with MODa is described in detail in the methods section of the paper. In brief, MODa uses multiple short sequence tag information in the spectra along with a dynamic programming approach to identify the peptide hits along with their protein identifiers. The use of multiple tags reduces both the database size and the false positives. MODa outputs 2 data files that we used in our analyses (1) it outputs the frequency of the mass-shifts on different amino-acids along with N- and C- term of the peptides, (2) it outputs peptide level information like most of other search engines. Once we have the MODa search results, we asked the following questions: (a) How much of the proteome is modified? (b) How do these modifications change over time? (c) Is there any biological relevance associated with the temporal change of a particular PTM? (c) Can we find any novel modifications in this *E. coli* data?

2.2 Modified *E. coli* proteome

First, we are interested in identifying the amount of proteome modified. Since we have data at each of the 9 time points i.e., from exponential to long-stationary phases, we plotted the total frequency of all PTMs at each of these time points. For this, we simply calculated the percentage of the total number of modified peptide spectral matches (PSMs) at each time point. We used 1 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 time from OD600 plot on the x-axis. It is clear that the amount of spectra that is modified stays approx.

at 30% during all the phases of the growth. Most of the previous studies on identifying multiple tandem PTMs looked at a "snapshot" of the proteome, instead of looking at different phases during the growth. We then looked at the coverage of the proteome at these time points. The number of proteins identified at each of these 9 time points are 2227, 2246 ,2278, 2251, 2306, 2342, 2298, 2234 and 2193 respectively. We considered all the peptide spectral matches i.e., one-hit wonders too. Later, when we use the data to identify the protein targets for different PTMs, we included a strict criteria of including proteins that have at least 2 peptide hits.

2.3   Mapping mass-shifts to post-translational modifications

We next looked at the distribution of the mass-shifts or PTMs at each time point of the growth curve i.e., what PTMs are present at a given time point and how does it change over time? MODa outputs mass-shifts in integer masses in the peptide hits. A mass-shift indicates either a PTM or a substitution compared to the reference database. The mass-shifts are then converted to PTMs using UNIMOD database, as described in methods section. For example, we know that +16 Dalton shift on any amino acid is most likely the oxidation. When we looked at the UNIMOD PTM list, we did see that this is oxidation, generally shown to occur on few amino acids such as cysteine, methionine etc. We did not limit our analysis to frequently occuring mass-shifts, but also focussed on well known PTMs, and novel mass-shifts that were not characterized earlier. However, we did not investigate into mutations and this will be a topic we will look in the future as this is a possibility with MODa output.

Even though the total amount of modified proteome is mostly constant 30%, the content of PTMs at each time point vary during different phases of growth i.e., a particular modification can either stay constant, decrease or increase from expoential to long-stationary phases. To look at this in detail, we first plotted the frequency of different mass-shifts from the MODa output at 2 different time points. Figure 2 shows this mass-shift frequency at 3 hrs and 24 hrs. Figure 2 (A) shows the mass-shift frequency at 3 hours, while (B) shows the frequency at 24 hours. The height of the bins is the percentage of the mass-shifts in the entire peptide-spectral match list. Since we used 3 biological replicates, the actual % is the mean of the 3 replicates. We tabulated the frequently occuring mass-shifts at each of these time points (Table 1). +1 Da is the most frequent mass-shift at all time points. However this +1 Da modification seems to be 13C peak-picking as it seems to be randomly occuring on all the amino acids, as shown in the profile (Suppl Figure S6). Next frequently occuring modification is +16Da that represents to oxidation. Oxidation seems to clearly go down from 3 (early exponential phase) to 24 hours (stationary phsae). We investigated this in detail and presented as a separate subsection below. Other than 42Da, -2Da most other frequently occuring mass-shifts are either artifacts due to sample preparation or a combination of 2 mass-shifts i.e., +1 Da and +57 Da, 2 +57Da's = 114 Da etc. +42Da corresponds to acetylation and is a PTM that is of biological interest. Hence, we investigated this PTM in detail, especially n-terminal acetylated proteins. -2 Da seems to be methionine to glutamine substitution. Other important PTMs in E. coli reported in literature are n-terminal methionine excision, phosphorylation, carboxylation, nitrosylation, methylation, glycosylation etc. So, we investigated few of these even though these appear at lower levels in our data set.

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

N-terminal Acetylation occurs in 2 forms i.e., N-alpha acetylation and N-sigma acetylation. Although cotranslational N-terminal N-alpha acetylation (NtAc) is widespread in eukaryotic proteins, the prevalence and physiological significance of this modification in prokaryotes is poorly understood. In E. coli, only five native proteins are known to possess an NtAc modification: the ribosomal proteins S5, S18, and L12/7 [27]; elongation factor Tu (EFTu) [[1]; and the chaperone SecB [35]. In addition, a number of heterologous eukaryotic proteins are modified with an NtAc when overexpressed in E. coli [3] [5] [42] [25]. Surprisingly, we identified 4335 NtAc-modified peptides from 55 proteins across the 9 time points and 3 biological replicates of our glucose starvation data. Although the total number of N-terminal peptides recovered from these proteins remains constant across all 9 time points, there is a clear enrichment of NtAc-modified peptides beginning in early stationary phase (8h), continuing through late stationary phase (336h) (Suppl Figure S2)

No N-terminal peptides were recovered from known targets S5, S18, L12/7, or EFTu, so we were unable to confirm the acetylation status of these proteins. However, among the NtAC peptides that were recovered, the Nt fragment from SecB was by far the most frequently observed, representing 15-41% of the total Nt-Acetylated peptides across the nine time points[Fig. 2]. In addition, six other proteins from our dataset were identified as Nt-acetylation targets in an enrichment-based analysis of N-terminal modifications in Pseudomonas aeruginosa [8] [Table 1]. The 49 novel Nt-acetylated proteins we identified span a variety of functional categories, including transcription (RpsH, YfgB, YjfH, LysS, CysS); amino acid metabolism (IlvA, SerC, CysM, AvtA); DNA repair and replication (FtsK, ParE, MutS, XerD) and membrane transport (YtfT, SecF, CopA, PotH, DppF, OppD, XanP, AtpH) [see Table 2].

Three N-terminal N-alpha acetyltransferases (NATs) are known to regulate NtAc in E. coli: RimI, which acetylates the N-terminal Ala residue of S18; RimJ, which targets the Nt Ala residue of S5 [44]; and RimL, which targets the N-terminal Ser of L12 [40]. The E. coli genome also encodes four putative N-acetyltransferases (yjaB, yhhY, yjhQ, and yjgM) that have been suggested as candidate NATs [[40]. The acetyltransferases responsible for Nt-acetylation of non-ribosomal proteins have not been identified, although RimJ [3] [40] and RimL [25] have been shown to be responsible for NtAc of ectopically expressed eukaryotic proteins. Little is known about the features governing target specificity of the Rim proteins; although specificity of eukaryotic NATs is largely dependent on the first AA residue following the iMet, mutation studies have shown that prokaryotic Rim protein specificity is largely unaffected by AA-1 mutations [25], but can be drastically changed by substitutions in downstream AA positions [5]. The AA-1 position in our 56 E. coli NtAc proteins are largely made up of Ser (34), Thr (10), and Ala (5) residues, and all but 3 have undergone iMet cleavage. Although there is a slight bias for Ser, Thr, Asp, Glu, and Gln at the AA2 position, sequence alignment and motif enrichment analysis were unable to identify any downstream targeting signal (Data not shown). Alternatively, some of the observed Nt-acetylation may take place via a non-enzymatic mechanism, as has recently been shown for Lys N-epsilon acetylation in E. coli [20] [41]. Nonenzymatic Lys acetylation increases in stationary phase similarly to our observations for NtAc levels, and is associated with accumulation of the high-energy intermediate acetyl phosphate. Nonenzymatic Lys acetylation is thought to act as a sensing system for high levels of cellular acetate metabolism, and interestingly, a similar relationship between Acetyl-CoA levels and Nt-acetylation has recently been shown for eukaryotic proteins [43].

NtAc has a variety of functions in eukaryotic cells, including regulating protein stability, ER

trafficking, protein complex formation, and membrane attachment [39], but there is no evidence for a similar role in prokaryotic cells. Nt acetylation of E. coli 30S ribosomal subunits S5 and S18 is thought to affect 30S ribosomal assembly by governing direct contacts with the rRNA [6], but no function for prokaryotic Nt acetylation outside of the ribosome has been proposed. Our observation of widespread NtAc in E. coli proteins suggests that this PTM is broadly important for protein homeostasis, but further work will be necessary to determine its effects on target proteins and the mechanisms of its regulation.

Acetylation in *E. coli* has been studied well in the past [5] [15]. 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 it occurs as a reversible modification, mostly on lysine [45]. In the current work, we focus on n-terminal acetylation. It has been known that almost 2/3rds of the proteins are substrates for n-terminal methionine excision. Following this methionine excision, n-terminal processing, especially n-terminal acetylation [10] has been shown in the past to be one of the most frequently occuring PTMs. Previous data suggests that this modification is more frequent in eukaryotes than bacteria. However our results indicate that there are quite a few n-term protein acetylations identified. We summarize our results below.

N-terminal acetyltransferases are responsible for the n-terminal acetylation [39]. 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 4). 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 4 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 [20]. Our results might suggest the same given the increase in the number of acetylated proteins from exponential to stationary phases. Table 1 (not seen, add this and other tables that show the ranked list of frequently occuring 10 PTMs at each of 9 or few time points) shows the acetylated proteins found under these 2 different phases under E. coli growth under glucose limiting conditions.

2.5 *E. coli* Phosphorylation

Even though some of the mass-shifts that map to phosphorylation (+80Da), carboxylation (+44Da), nitrosylation (+29Da) were less abundant, we did a brief survey on these are some PTMs that were shown to be part of bacterial proteins. Most of the studies that looked at phosphoproteome involved enriching the sample for this PTM using techniques such as IMAC, TiO2 and then running mass-spectrometry on the sample. Figure 6 shows the phosphorylations at different points of the growth curve. We considered only the phosphorylations on serine, threonine and tyrosine, as these are well known amino acids that get phosphorylated in most species including *E. coli* [23]. The number of phosphorylations we identified during the time course seem to be small. This is expected, as the sample is not enriched for phosphorylations. However, we observed that the phosphorylations increase from exponential to stationary phases. Such pattern is also previously shown in phosphoproteomic studies [36] probably pointing to a role of phosphorylation in later

stages of the growth cycle.

We then looked at carboxylations (See Suppl Figure S3)and nitrosylations (See Suppl Figure S4). Like, phosphorylations, carboxylations seem to increase during the later stages of the growth indicating a likely role of this PTM in late stationary phase.

2.6 Oxidative damage and repair in *E. coli*

E. coli grown aerobically react with oxygen in the atmosphere producing reactive oxygen species such as H2O2 CiteGonzalez-Flecha B, Demple B (1995) Metabolic sources of hydrogen peroxide in aerobically growing Escherichia coli. J Biol Chem 270: 13681–13687. Targets for these ROS are generally proteins and lipids, that get oxidized. This oxidized state affects the protein structure and hence results in changes in its function leading to disturbances in the metabolism. Bacteria have genetic systems that responds to this oxidative stress through oxyR, SoxXY etc or through reductases that reduce protein to its original non-oxidized state. Here, we investigated the levels of oxidation at different phases of growth.

Oxidation frequently occurs on cysteine and methionine and also has been shown to occur relatively less frequently on few other amino acids. However in our sample, since cysteine is capped to avoid disulphide bond formation, we expect +16 Dalton oxidation modification to occur primarily on methionone (see Suppl Figure 6), which is the case. We plotted both the global oxidation levels and the methionine oxidation levels. Figure 7 (A) shows the global oxidation levels occuring on all amino acids. The oxidations seem to go down from exponential to stationary to long stationary phases. Figure 7 (B) shows the oxidation occuring on methionine only. The pattern seems to be same as the global oxidation level i.e., goes down from exponential to stationary phases. Since oxidative stress during stationary phase leads to accumulation of ROS, we expected the oxidation level to go up from exponential to stationary phases. However, the observed trend is opposite to what we expected. There could be 2 possible explanations (1) Either the level of methionines went up from exponential to stationary phases and/or (2) *E. coli's* bacterial genetic system is acting to bring the oxidation levels down to protect against this oxidative damage. The number of methionines seemed more or less constant during the entire time course (data not shown). To see if *E. coli* machinery is responding, we looked into literature to find any such reductases that fix methionine sulfoxide (MetSO) back to methionine (Met). There seems to be 2 such reductases, MsrA and yeaA, In *E. coli* REL606 strain.Since MODa also outputs the peptide and hence the protein level information, we plotted the protein abundances of these reductases in Figure 8. There seems to be a sudden increase in these levels during stationary and long-stationary phases that might explain the lower levels of methionine oxidation levels. We also validated this result using our previous work (Cite John's paper) on protein level analysis on this data using Sequest, a different search engine. A point to note is that we could not use the previous search results for our multiple PTM analysis here, as the only variable modification used in that study is oxidation methionine. The Sequest result is shown as a fig:SequestMODaFig (see Suppl Figure 7).

Then we looked at the protein oxidation targets at these different points of time curve Alkyl hydroperoxide reductase subunit, elongation factor Tu, chaperonin GroEL, S-methyltransferase, cytochrome d ubiquinol oxidase, elongation factor G, phosphoribosylaminoimidazole-succinocarboxamide synthase were found to be the major oxidation targets. Some of these were not shown previously to be major oxidation targets. Most of the proteins were also seem to be oxidized at later stages of growth i.e., during stationary and late-stationary phases.

One interesting modification that might be preventing the oxidation of methionine is the substitution of methionine to glutamine that seem to decrease from exponential to stationary phase.

2.7   Other PTMs and artifacts

Some of the frequently occuring mass-shifts (add 22 Da and 38 Da from Table X) were annotated as artifacts in UNIMOD database. These were annotated as Na and K adduct formation that were also shown previously [?]. Still, most of the searches performed in the literature ignore these modifications. The frequency of these mass-shift seems to be 3-4%, combined. To validate if these are really Na and K adducts, or if it is the PTM on any particular amino acid, we plotted the profiles of Na and K adducts (see Suppl Figure S5). This mass-shift seems not to occur on H, K and R but seemed to occur randomly on all the other amino aicds. Histidine, lysine and arginine are amino acids that have side chains that are basic and carry a small positive charge at physiological pH. This explains that this mass-shift is indeed adducts from Na and K, as these adducts don't form on basic amino acids at physiological pH. Since including these Na and K adducts seem to increase the sensitivity of the search, we give a general recommendation to include these adducts as variable modifications in search algorithms that limit the number of PTMs as variable modifications.

Most of the PTMs seem to occur less frequently, but sometimes seen on the same peptide or the protein. So, we calculated the peptide-spectrum level information to identify these low-abundant modifications. One of the mass-shifts (-17Da) seem to occur on n-terminal glutamine of the protein. This represents pyroglutamate formation. This is a well known PTM that stabilizes the protein to avoid any further n-terminal processing. This PTM seemed to be constant during the entire time course (see Suppl Figure 8). We provided the list of all the modifications that occur on the same position of the peptide as supplementary information (See Suppl Table X).

2.8   Large PTM analysis

The recommended mass-shift for MODa program is finding mass-shifts between -200 and 200Da. However it is known that some PTMs are larger and would miss out using the recommended mass-range. So, to see if we can identify any frequently occuring peptide-spectrum matches with a higher mass-shift we changed this mass-range to -200 and 300. We did not identify any PSMs (peptide spectral matches) that occur frequently that has higher PTMs.

3   Discussion

We looked at temporal changes of the post-translational modifications in *E. coli* REL606 strain. For this, we used mass-spectrometry based proteomics data. This data is collected under glucose limiting conditions for different phases of the growth i.e., exponetial to long-stationary phases upto 2 weeks under these conditions. For the 9 time points analyzed (3 hours to 2 weeks), several interesting insights into the frequency and the temporal changes of post-translational modifications seen in this analysis. First, we show that consistently 30% of the proteome is modified at all the 9 time points analyzed. Second, we show that n-term protein acetylation goes up from exponential to stationary phases pointing to a likely role of this modification during late stages of the growth. Similar results were found for phosphorylation, carboxylation too. However, oxidative stress seem to go

down from exponetial to stationary phases. Nutrient limitation, accumulation of toxic substances and change in pH were supposed to increase the oxidative stress, however *E. coli* has a mechanism to protect the cell from oxidative damage using the sulfoxide reductases fixing MetSO back to Met. Finally, the computational techniques used in this study can be used with any mass-spectrometry based proteomics data to identify or characterize novel post-translational modifications.

In one study [37], variant of MaxQuant algorithm is used to do an unrestricted search in *E. coli* during exponential and extended stationary phases of growth, however our study used a more advanced algorithm, MODa, suited to do a larger window of mass-shifts. In [37], these mass-shifts were not seen, either probably because of the sample processing or more likely because of using a general search engine, instead of MODa.

Recently, the focus has shifted from identifying peptides/proteins to characterizing PTMs as shown in this recent review article [30]. However identifying all the PTMs present in the sample is limited both by experimental enrichment techniques that cannot be applied to all kinds of PTMs at once and also from the computational limitations, although there has been an improvement in the latter limitation lately. Some of the future applications of current methodology of identifying PTMs at once is to help understand the PTM crosstalk under each condition analyzed. This in turn helps us to understand the protein regulation or function specifically generally caused by many PTMs acting in concert, as earlier witnessed [31]. So, analysis of PTMs in *tandem* is a necessity to get accurate reflections of the underlying interactions of the PTMs. Previous work in this area used the PTMs deposited in the database, however care must be taken, as these PTMs were identified at different growth phases under differing conditions [31].

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, that are identified *in vivo* and well documented in databases like UNIMOD, RESID. MODa uses the prefix residue mass to comput the score. The probability is calculated using different features such as PRM score, mass errors, number of fragments matched. The program outputs mass-shifts and we can use databases such as UNIMOD, RESID to map these mass-shifts to the relevant PTMs. Here, we looked at amino acid profiles to determine the exact PTM type. 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 such as 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 that can be used later 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.

N-terminal processing of proteins is a well-known PTM in diverse kinds of species ranging from eukaryotes to bacteria [19]. N-terminal acetylation has been shown in the past to provide insights into the nature of the n-terminal of the proteome [16]. In eukaryotes, n-terminal acetyl-transferases are well characterized and studied, for example, as shown in these studies [33], [34]. In *E. coli*, NATs are primarily characterized into mainly rim types [40], [44]. There is lot of work done on lysine acetylation in *E. coli*, however there are not many n-terminal acetylation sites identified or investigated in detail in the past. Lysine acetylation E. coli, acetylation in exponential and stationary phases ***Colin's discussion***.

Protein oxidation by reactive oxygen species (ROS) is important and linked to various diseases including aging [38]. ROS generally seem to accumulate at later stages of growth, because of possible accumulation of toxic substances, nutrient limiting conditions, changes in pH etc. *E. coli's* internal machinery to tackle this oxidative stress comes in many forms at different levels and in this analysis, we were interested at the protein level. *E. coli* has a class of sulfoxide reductases [4] [46] that fix methionine sulfoxide back to methionine. MsrA and yeaA are the proteins found in this reductase class of enzymes. These reductases has been shown to protect *E. coli* from oxidative damage [13]. Here, we looked at the abundances of these proteins during these late stationary phases, where oxidative stress seems to be more. Our earlier work has already shown the induction of these stress proteins at stationary phases using the same mass-spec data set [17]. Even though it is common for the oxidation to happen during sample preparation, we expect this oxidation level to be random and more or less constant at different time points of the growth curve. However, in our current data set, we clearly see that the levels of oxidation go down from exponential to stationary phase, possibly correlating with the oxidation not caused by sample preparation.

One of the PTM networks that is well studied in the past is the phosphorylation signaling cascade [29] (also look for phosphorylation papers that you used in the past at NCBI). These are well known signalling mechansims and have been shown to respond quickly to stress during the late-stationary phases [36]. Our analysis also supports this behavior and would like to emphasize that while the previous paper investigated only 1 PTM, our analysis investigated most of the PTMs in tandem.

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 -200 and 300 Da range). So, we miss out on larger PTMs. Another computational limitation is that we looked for only 1 possible modification on each of the peptides identified. This is because MODa was shown in the past to generate many false positives if searched for multiple PTMs on the same 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 identifications in its search. Another, probably important limitation is that we did not do any PTM level quantitation or try to understand the specific PTM stoichiometry, as these require sophisticated experimental instrumentation, good enrichment protocol and the algorithms to characterize the PTMs associated with proteins. However MODa has been a software suite that has been recently used in both in bacterial proteomics as well as biomedical applications such as urinary proteomics to find biomarkers [22].

Current whole-cell models [7] integrate diverse kinds of OMICS data i.e., transcriptomics, proteomics not only to refine the existing models, but also make reasonable predictions on which genes/enzymes are key for example, for a particular metabolite production etc. Here, we argue that including the modification information (i.e., number of modified proteins to that of the unmodified version) would improve the protein level abundances and thus will improve the computational predictions.

4 Conclusions

The modified protein seems to be at a constant 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 protecting *E. coli* from oxidative damage by fixing methionine sulfoxide back to methionine. We found some novel targets for n-term acetylation that were not previously identified in *E. coli*. Phosphorylations and carboxylations also seem to increase indicating a likely role of these PTMs in late-stationary phases of the growth curve. A novel phosphoserinegluconylation on ribosomal protein along with few others is identified. 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

The details of *E. coli* growth is provided in our manuscript that described the initial analysis of this data [17]. In brief, *E. coli* was grown in glucose minimal media and the samples are collected at 8 different time points. The time points considered were 3, 4, 5, 8, 24, 48, 168 and 336 hours of *E. coli* growth. At each time point, there were 3 biological replicates i.e., the cultures grown at different time. Details on the mass-spectrometry is provided in the same paper as well. Trypsin was used to digest the proteins and then the sample is analyzed using liquid chromatography mass spectrometry (LC/MS) on a LTQ-Orbitrap (Thermo Fisher). For each time point, there were 3 biological replicates that were analyzed.

5.2 Post-translational modification identification and analysis

Mass-spectrometry raw data was then converted into mzXML files to input into MODa [26]. MODa is a naive based spectral alignment algorithm that identifies peptides and their associated PTMs from the input spectral files. MODa generates sequences tags from experimental MS/MS spectra and using a dynamic programming approach, the program identifies the mass-shifts between the sequence tags. Identifying sequence tags reduces the database search significantly, thus reducing the total run time. For each PSM, if there is mass-shift present, it outputs it as a frequency table, listing the total number of mass-shifts found on each of the amino acids, along with c-term and n-term for all the PSMs involved in the dataset. In a separate file, it also lists all the peptide spectral matches. First, we used the mass-shifts to investigate at PTM level. So, for each mass-shift, we identified the possible PTM it matches to, using UNIMOD database. We did this task manually i.e., if we see +16Da, from UNIMOD we know that it is oxidation. We then looked at the amino-acid profile for the mass-shift we consider, to reduce the ambiguity. If we see +16Da on Met, it is clearly Methionine oxidation than a carboxy to thiocarboxy conversion, as this carboxy conversion happens on glutamine only. At least for all the mass-shifts considered in this analysis, there was no ambiguity in terms of a mass-shift matching to multiple possible PTMs, except +1Da.

11

We looked at the aminWe did not do any in-depth analysis for +1Da, as it can result both from precursor peak picking deamidation or combination of PTMs (i.e., 43Da, could be +1Da and 42Da). Since we used a restriction to identify only one PTM, to reduce false positives, this scenario is possible. For some mass-shifts, we looked at the amino-acid profiles i.e., which amino acid has the mass-shift frequently to confirm the UNIMOD PTM list i.e., from oxidation profile, it is clear that Met is the most frequent and from UNIMOD, it is obvious that this corresponds to methionine oxidation. We also looked at well known PTMs that were observed in previous studies in *E. coli*. 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 modification happens, we were able to plot the temporal changes. Some frequently occuring mass-shifts did not map to PTMs, but they did map to Na and K adducts, as evident from UNIMOD. In this analysis, we did not consider any mutations. In future, we plan to look at these separately.

We ran separate MODa searches for each of the 9 time points. Since there were 3 biological replicates, this resulted in total 27 MODa searches. To speed up the searches, we used UT TACC 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 used for the precursor ion is 10 ppm, while the mass-tolerance used for the product ion is 0.5 Da. We set carbamidomethylation of cysteine as a static or fixed modification. As mentioned earlier, MODa requires a mass range to search for variable modifications, so we run MODa searches for 2 scenarios: (1) mass range between -200 to 200Da and (2) second search with mass range between -100 to 300Da. Since we were interested in PTMs, there was no need to include negative mass-shifts, but we included these as mentioned in the MODa protocol. We used REL606 NCBI *E. coli* sequence library.

Once we have the peptide hits from MODa output, we used target-decoy approach [11] to identify high-confidence hits. In this approach, we reverse the original REL sequences and concatenate these to the original sequence database to form a target-decoy database. This database is twice the size 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 that is a general norm in mass-spectrometry based proteomics searches. The protein levels are plotted by using simple peptide spectral counting, that seems reasonable in the case of simple model organism such as *E. coli*.

5.3   Protein target analysis

For n-term acetylation and oxidation, we did additional analyses at protein level. We identified all the proteins that were acetylated at n-term, or oxidized at methionine and compared these targets with the available literature to see if we identify additional novel protein targets. The scripts were written in ipython. The GO analysis was done using DAVID.

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/wilkelab/Ecoli_PTMs`.

6 Author Contributions

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

7 Acknowledgments

References

1. K. Arai, B. F. Clark, L. Duffy, M. D. Jones, Y. Kaziro, R. A. Laursen, J. L'Italien, D. L. Miller, S. Nagarkatti, S. Nakamura, K. M. Nielsen, T. E. Petersen, K. Takahashi, and M. Wade. Primary structure of elongation factor tu from escherichia coli. *Proc Natl Acad Sci U S A*, 77(3):1326–30, 1980.

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. L. F. Bernal-Perez, F. Sahyouni, L. Prokai, and Y. Ryu. Rimj-mediated context-dependent n-terminal acetylation of the recombinant z-domain protein in escherichia coli. *Mol Biosyst*, 8(4):1128–30, 2012.

4. 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.

5. 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.

6. S. F. Clatterbuck Soper, R. P. Dator, P. A. Limbach, and S. A. Woodson. In vivo x-ray footprinting of pre-30s ribosomes reveals chaperone-dependent remodeling of late assembly intermediates. *Mol Cell*, 52(4):506–16, 2013.

7. 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.

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

9. 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.

10. 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.

11. J. E. Elias and S. P. Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, 4(3):207–14, 2007.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. John R. Houser, Craig Barnhart, Daniel R. Boutz, Sean M. Carroll, Aurko Dasgupta, Joshua K. Michener, Brittany D. Needham, Ophelia Papoulas, Viswanadham Sridhara, J. Marx Dariya Sadykova, Christopher, M. Stephen Trent, Jeffrey E. Barrick, Edward M. Marcotte, and Claus O. Wilke. Controlled measurement and comparative analysis of cellular components in e. coli reveals broad regulatory changes under long-term starvation. *PLoS Computational Biology*, 0:0–0, 2015.

18. 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.

19. 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.

20. M. L. Kuhn, B. Zemaitaitis, L. I. Hu, A. Sahu, D. Sorensen, G. Minasov, B. P. Lima, M. Scholle, M. Mrksich, W. F. Anderson, B. W. Gibson, B. Schilling, and A. J. Wolfe. Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS One*, 9(4):e94816, 2014.

21. A. I. Lamond, M. Uhlen, S. Horning, A. Makarov, C. V. Robinson, L. Serrano, F. U. Hartl, W. Baumeister, A. K. Werenskiold, J. S. Andersen, O. Vorm, M. Linial, R. Aebersold, and M. Mann. Advancing cell biology through proteomics in space and time (prospects). *Mol Cell Proteomics*, 11(3):O112 017731, 2012.

22. 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.

23. 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.

24. L. McHugh and J. W. Arthur. Computational methods for protein identification from mass spectrometry data. *PLoS Comput Biol*, 4(2):e12, 2008.

25. L. Miao, H. Fang, Y. Li, and H. Chen. Studies of the in vitro nalpha-acetyltransferase activities of e. coli riml protein. *Biochem Biophys Res Commun*, 357(3):641–7, 2007.

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

27. M. V. Nesterchuk, P. V. Sergiev, and O. A. Dontsova. Posttranslational modifications of ribosomal proteins in escherichia coli. *Acta Naturae*, 3(2):22–33, 2011.

28. T. Nystrom. Stationary-phase physiology. *Annu Rev Microbiol*, 58:161–81, 2004.

29. 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.

30. 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.

31. 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.

32. 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.

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

34. 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.

35. V. F. Smith, B. L. Schwartz, L. L. Randall, and R. D. Smith. Electrospray mass spectrometric investigation of the chaperone secb. *Protein Sci*, 5(3):488–94, 1996.

36. 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.

37. B. Soufi, K. Krug, A. Harst, and B. Macek. Characterization of the e. coli proteome and its modifications during growth and ethanol stress. *Frontiers in Microbiology*, 6:1–11, 2015.

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

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

40. 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.

41. B. T. Weinert, V. Iesmantavicius, S. A. Wagner, C. Scholz, B. Gummesson, P. Beli, T. Nystrom, and C. Choudhary. Acetyl-phosphate is a critical determinant of lysine acetylation in e. coli. *Mol Cell*, 51(2):265–72, 2013.

42. J. Wu, S. Chang, X. Gong, D. Liu, and Q. Ma. Identification of n-terminal acetylation of recombinant human prothymosin alpha in escherichia coli. *Biochim Biophys Acta*, 1760(8):1241–7, 2006.

43. C. H. Yi, H. Pan, J. Seebacher, I. H. Jang, S. G. Hyberts, G. J. Heffron, M. G. Vander Heiden, R. Yang, F. Li, J. W. Locasale, H. Sharfi, B. Zhai, R. Rodriguez-Mias, H. Luithardt, L. C. Cantley, G. Q. Daley, J. M. Asara, S. P. Gygi, G. Wagner, C. F. Liu, and J. Yuan. Metabolic regulation of protein n-alpha-acetylation by bcl-xl promotes cell survival. *Cell*, 146(4):607–20, 2011.

44. 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.

45. B. J. Yu, J. A. Kim, J. H. Moon, S. E. Ryu, and J. G. Pan. The diversity of lysine-acetylated proteins in escherichia coli. *J Microbiol Biotechnol*, 18(9):1529–36, 2008.

46. 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
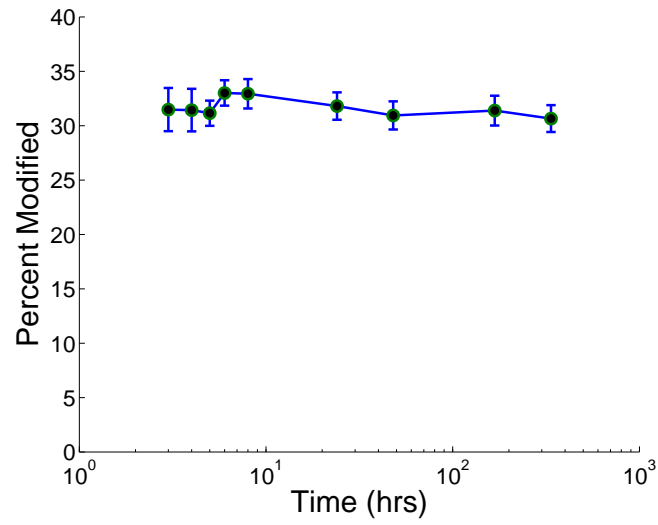
Figure 1. *E. coli* modified proteome. The total number of modified peptide-spectral matches seem to be constant at 30% for all 9 time points. .
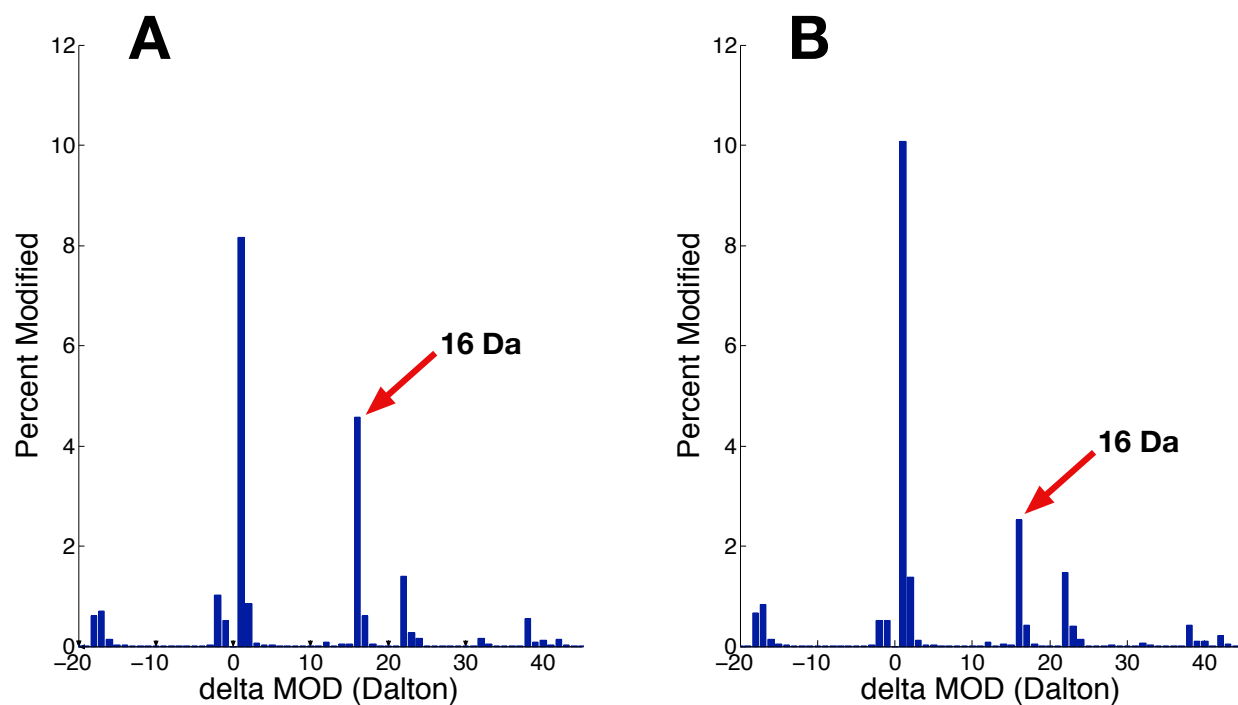
Figure 2. MODa outputs mass-shifts. A naïve based algorithm like MODa can alleviate the requirement of guessing PTMs beforehand. However MODa outputs mass-shifts on the amino acids. We can then use PTM databases like UNIMOD to map the mass-shift to the most probable PTM. (A) and (B) are the frequencies of the mass-shifts observed at 3 hours and 2 weeks of the *E. coli* growth.
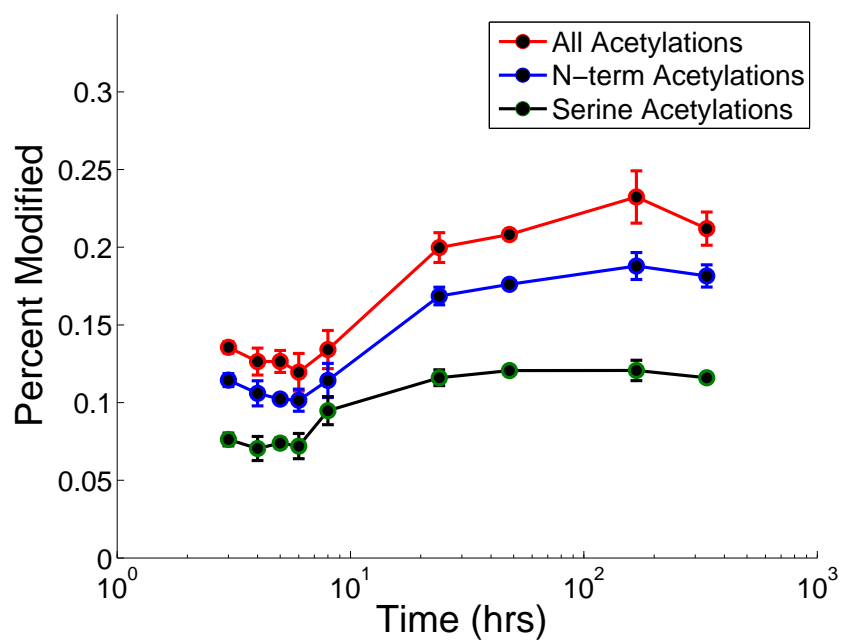
Figure 3. Protein n-term acetylations are dominant. Total number of acetylations as well as the n-term/serine acetylations seem to go up over 2 weeks. *E. coli* grown on glucose generally tend to accumulate acetate, perhaps this resulted in increase in acetylations.
Is this percent of all PSMs, or all peptides? What is the number value?
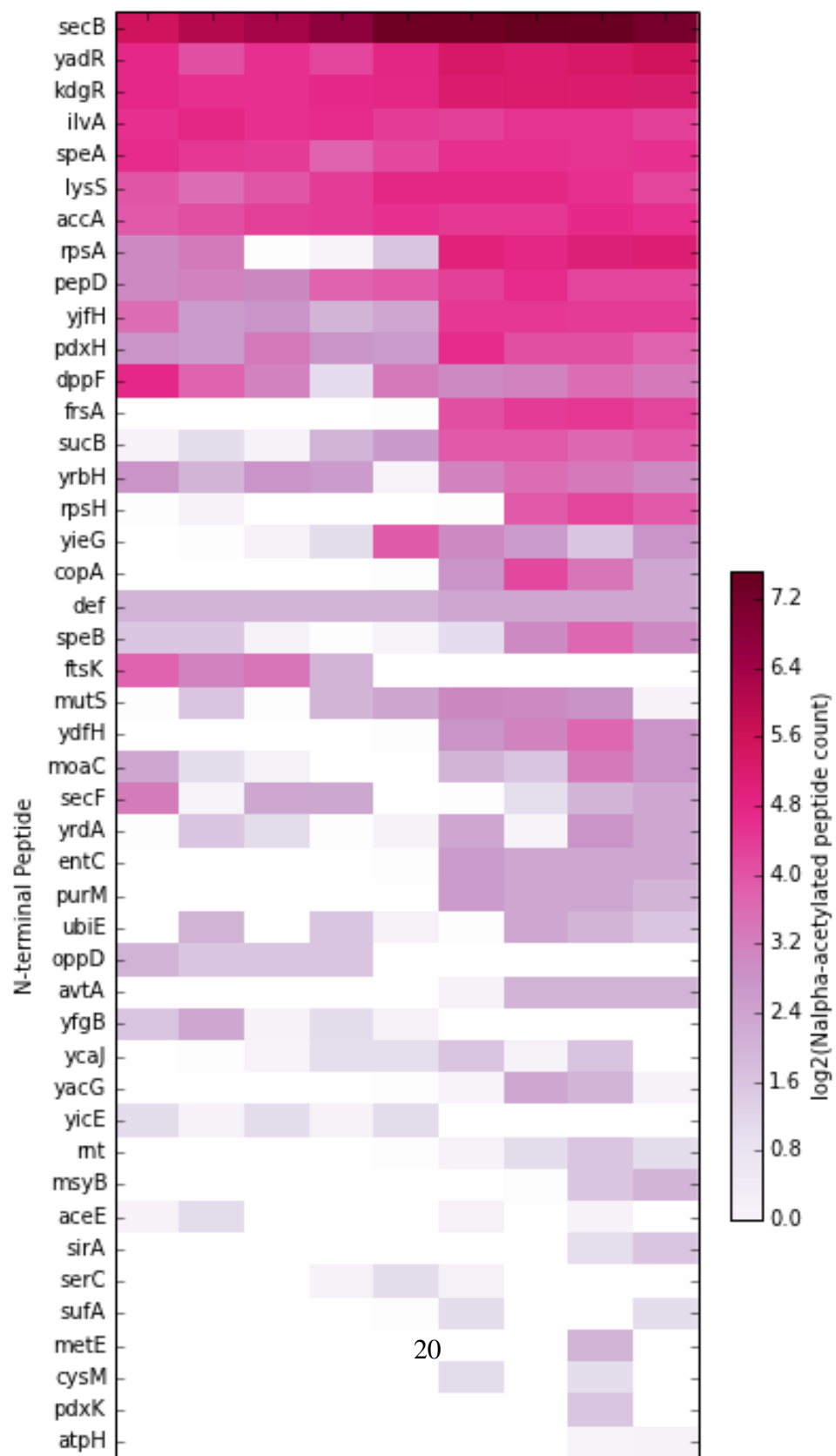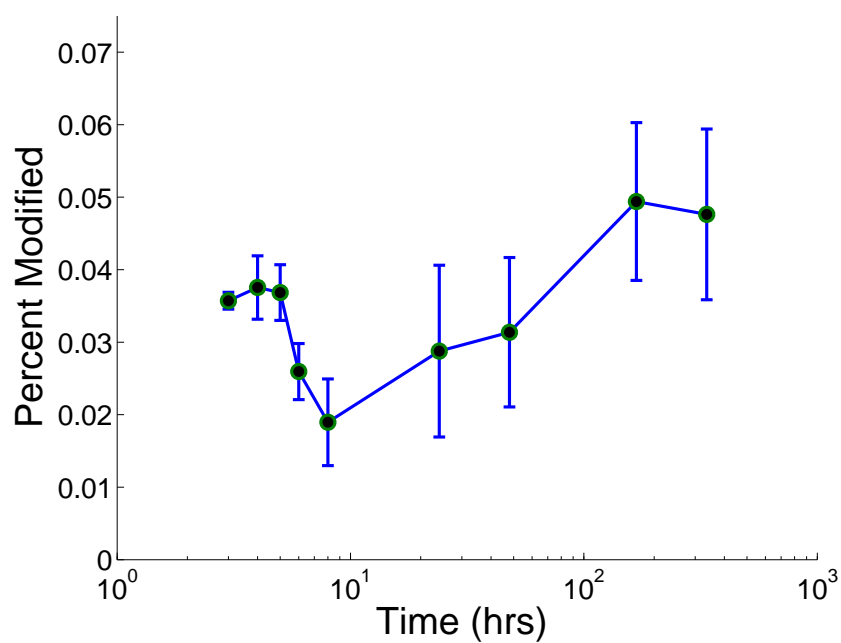
Figure 5. Phosphorylations are rare. Phosphorylations seem to be low and tend to increase during last week of growth. 2 frequently phosphorylated proteins in MODa search output are phosphoglucomutase and elongation factor Tu.
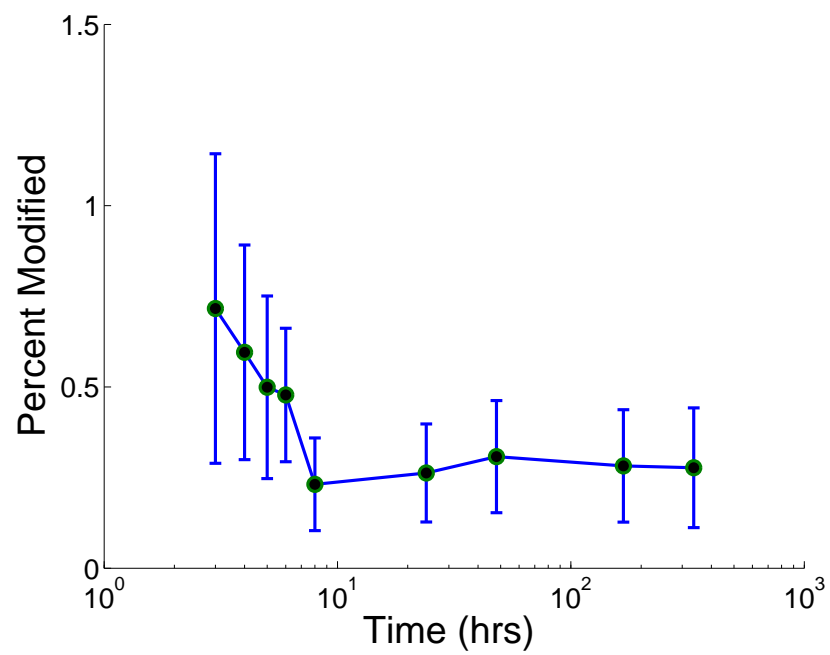
Figure 6. Methionine to Glutamine subsitution. .
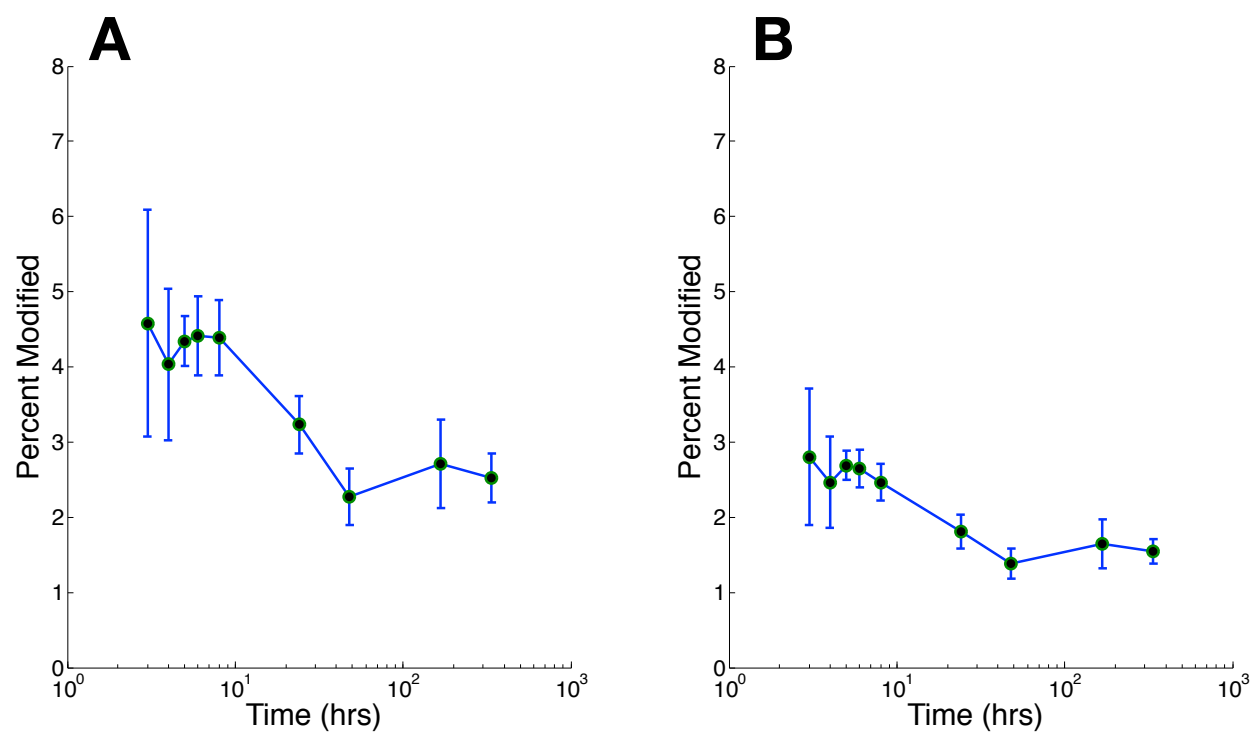
Figure 7. Oxidations go down over 2 weeks. (A) Total number of oxidations seem to go down from exponential to stationary phases. (B) The same trend follows for methionine oxidations too.

Supplementary Figures
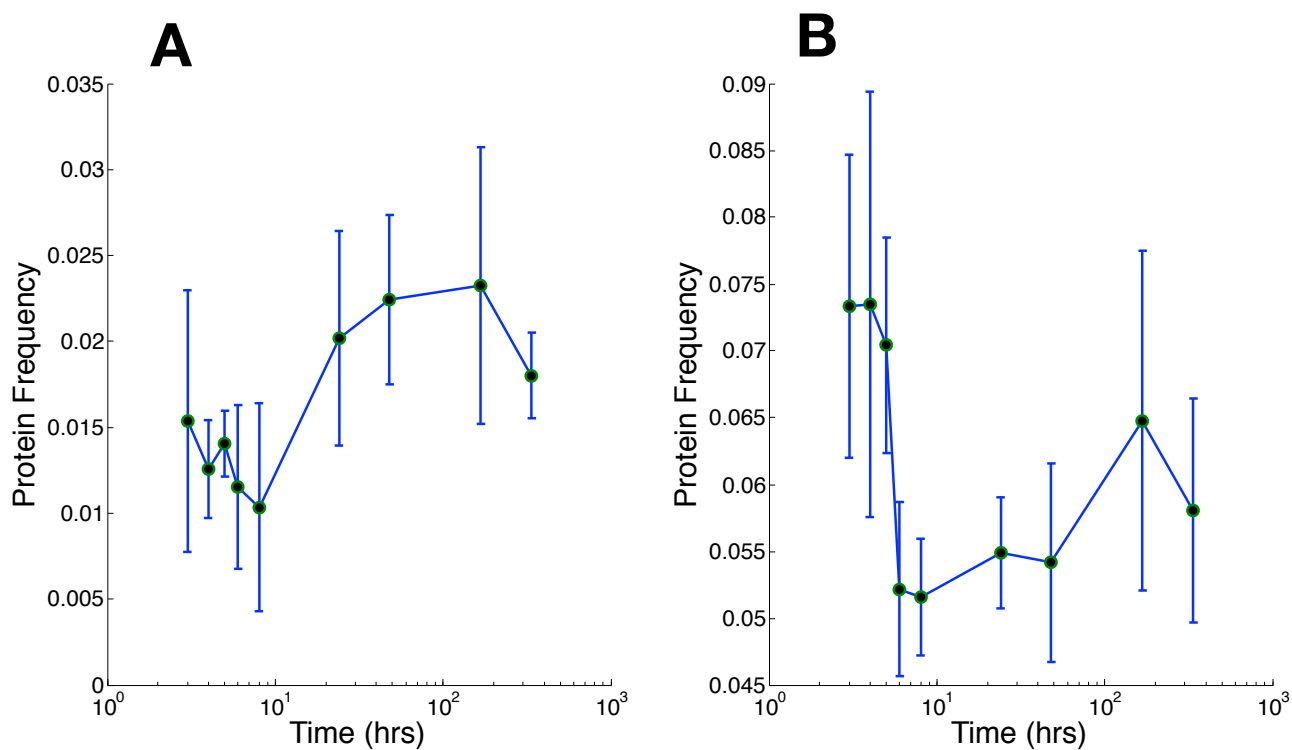
Figure 8. Relative protein abundances of methionine sulfoxide reductases MsrA and MsrB. (A) Protein abundances of MsrA seems to increase going from exponential to long-stationary phase. One of the sites oxidized on MsrA is FQAA[M+16]LAADDDR. (B) MsrB also seems to increase going into the long-stationary phase, protecting *E. coli* from oxidative damage.
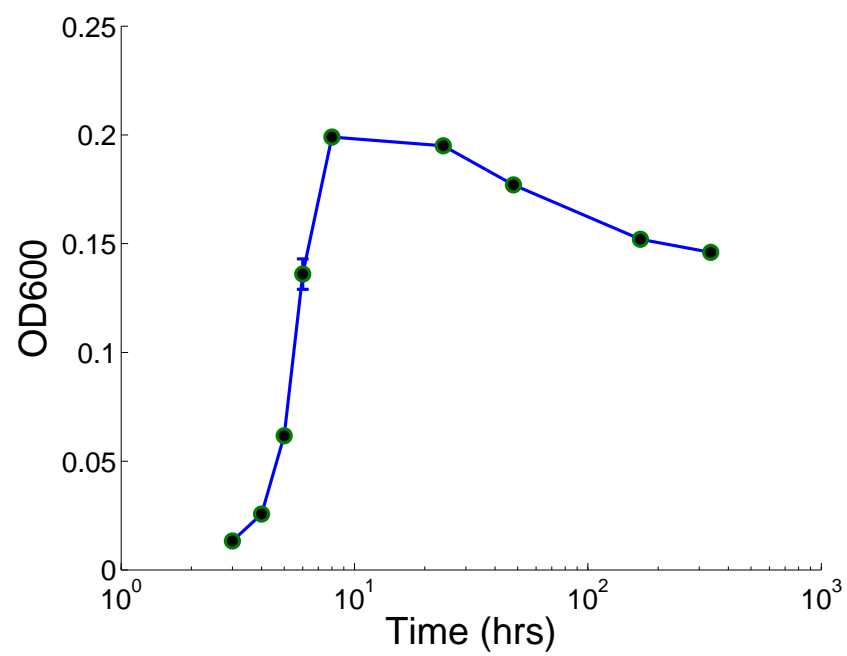
Figure S1: OD600 curve . Growth curve (OD600) of REL606 under glucose starvation conditions.
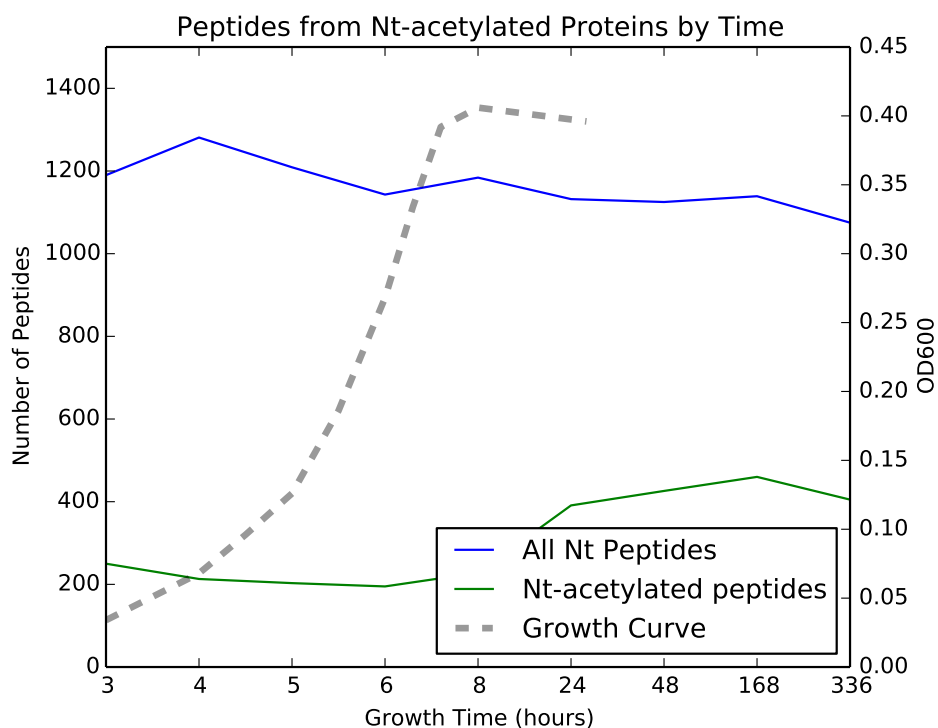
Figure S2: N-terminal Nalpha Acetylated Peptide Occurance across Growth Timepoints.
Blue line represents total count of N-terminal peptides with any (or no) modification originating
from parent proteins with at least one Nt-acetylated peptide in the dataset. Green line represents
total count of Nt-acetylated peptides from these proteins. Both counts are the sum of three biolog-
ical replicates for each timepoint. Dashed line (right axis) shows OD600 values for an example
culture grown under identical glucose-starvation conditions.
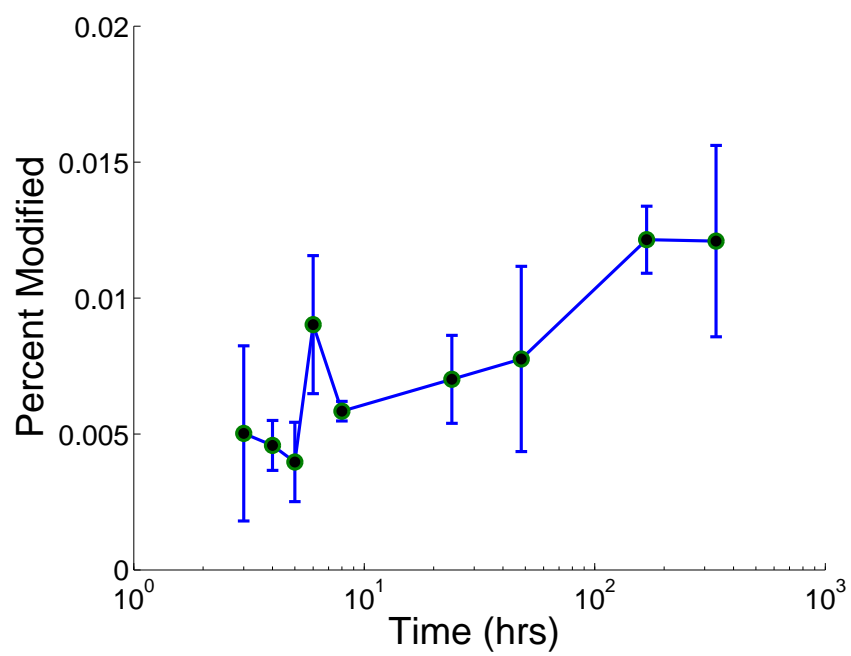
Figure S3: E. coli Carboxylations . Carboxylations seem to go up.
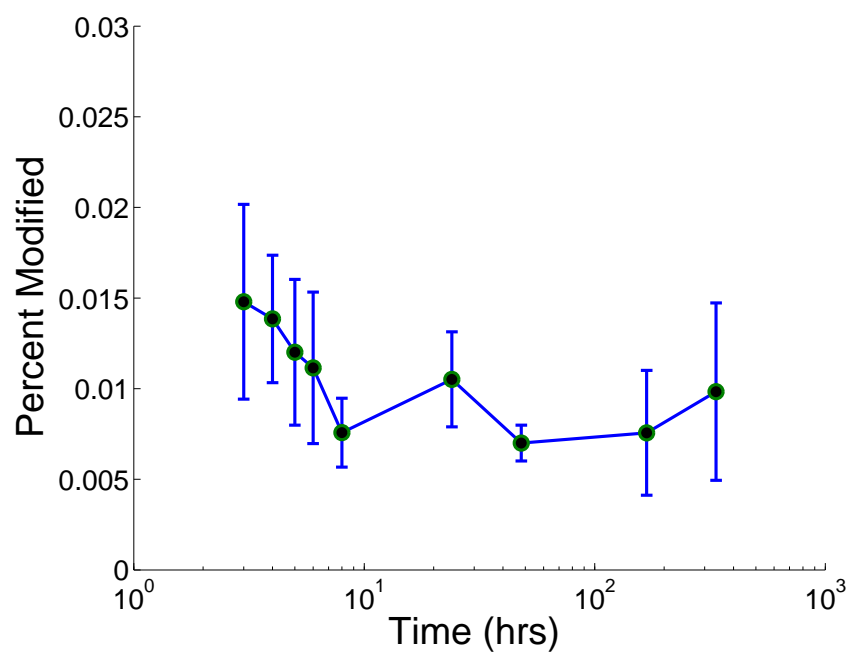
Figure S4: E. coli Nitrosylations. Nitrosylations seem not to change much during the entire growth period.
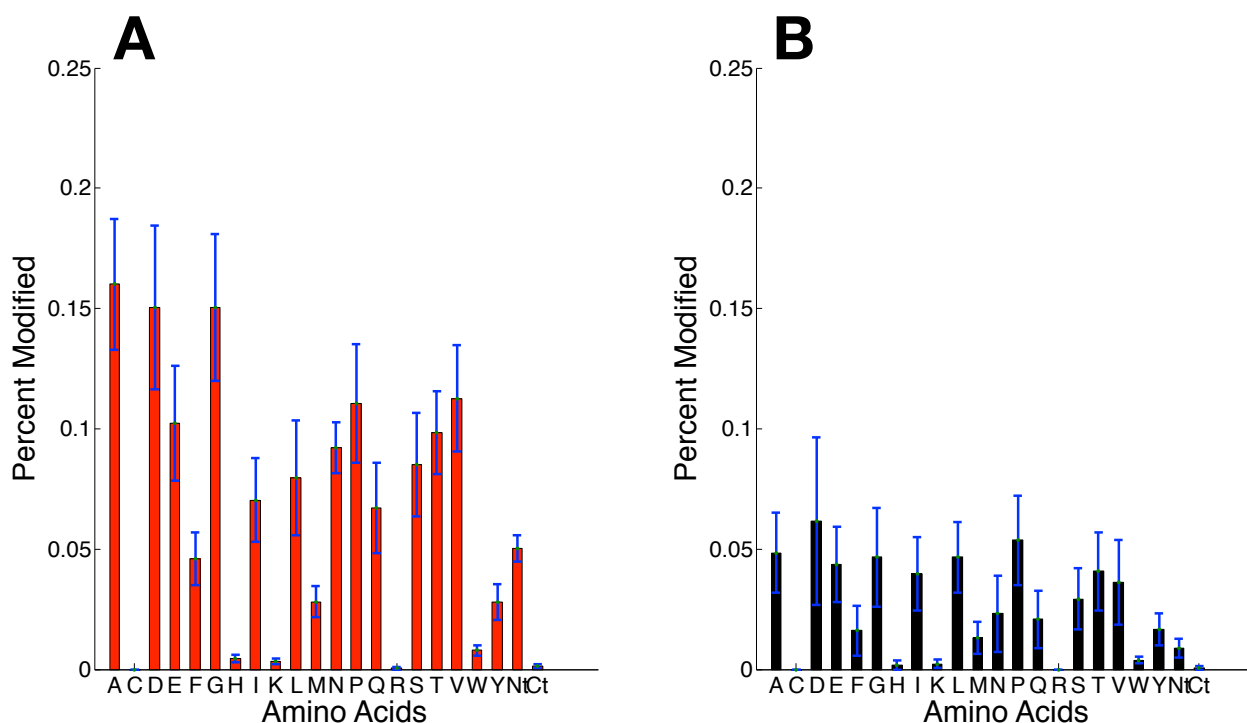
Figure S5: Na and K adducts. (A) Na and (B) K adducts seem to happen on all amino acids except histidine, lysine and arginine. H, K and R are amino acids that are basic and carry some (+) charge at physiological pH. Here Nt and Ct represent n-termins and c-terminus of the peptide.
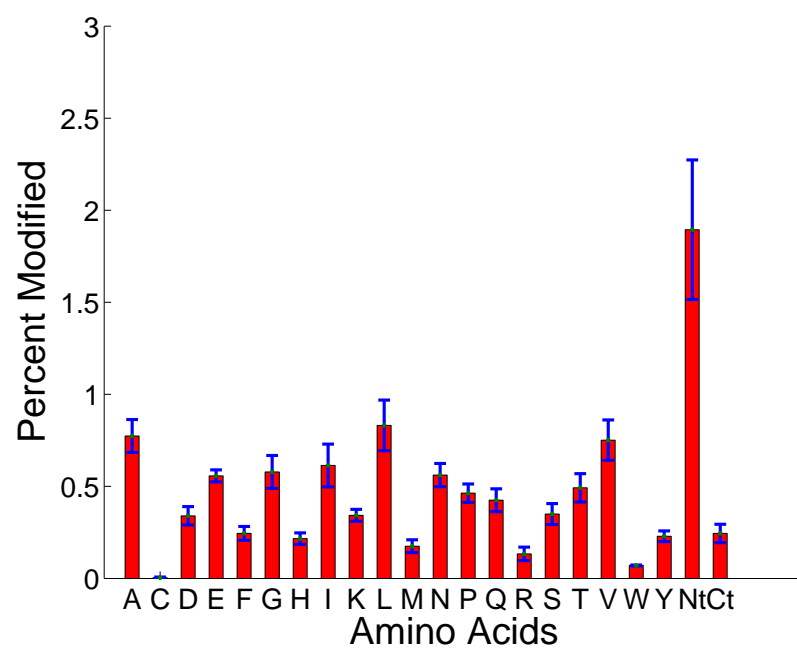
Figure S6: +1Da shift occurs randomly. We cannot infer that this +1 Dalton mass-shift is deamidation as it occurs randomly on all amino acids, inferring it is carbon-13 (C13) peak picking as previously shown in many studies.
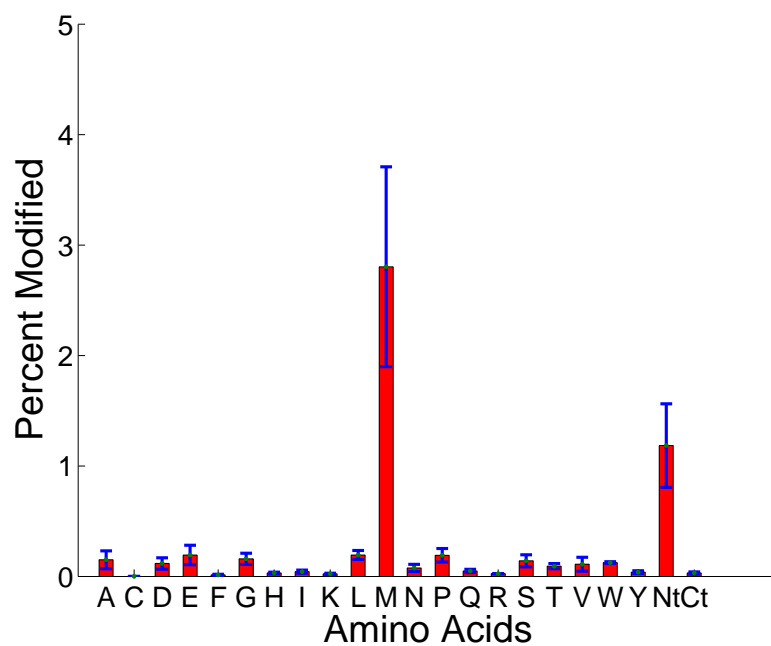
Figure S7: Oxidation is dominant on methionine. Even though many amino acids could be oxidized, in this data set, oxidation seems to occur primarily on methionine, as expected.
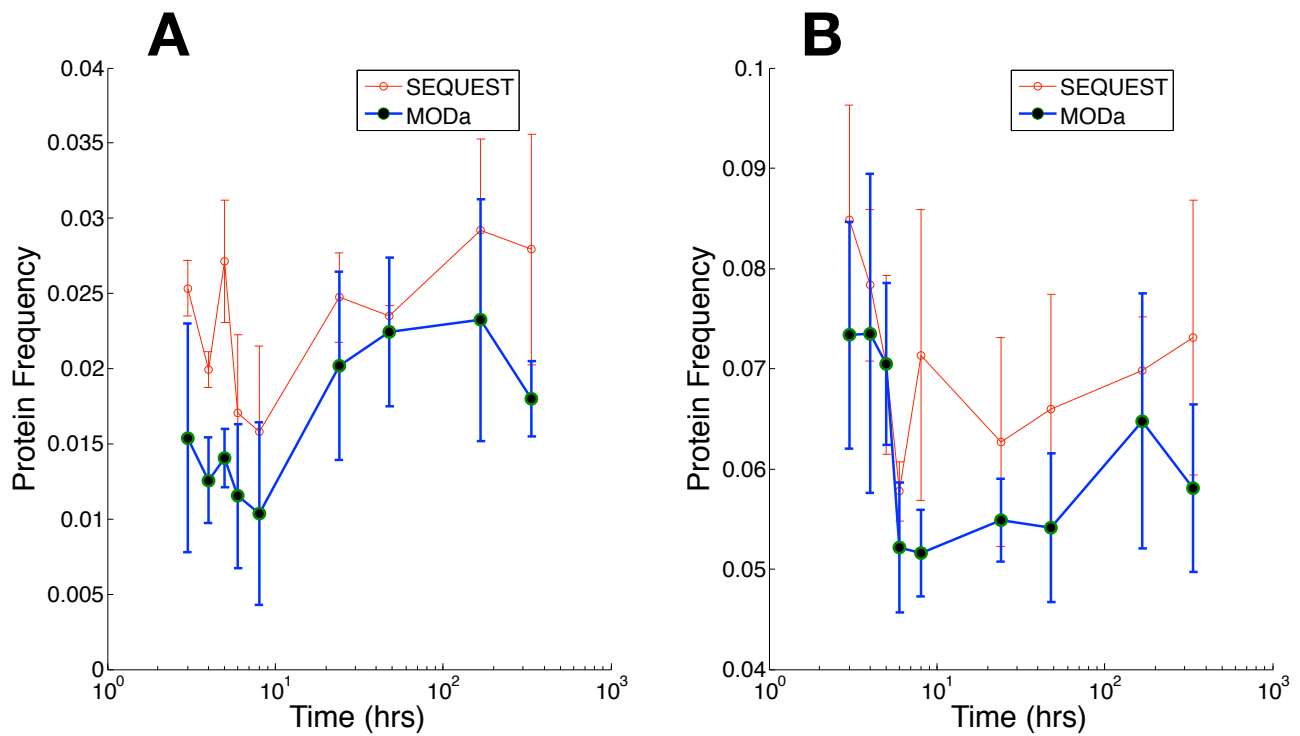
Figure S8: Sequest and MODa results agree. We compared MODa protein abundance results with Sequest to double-check the protein levels of the sulfoxide reductases that protect *E. coli* from oxidative stress by fixing MetSO to Met. As seen, MODa and Sequest results seem to agree well for (A) MsrA and (B) MsrB.
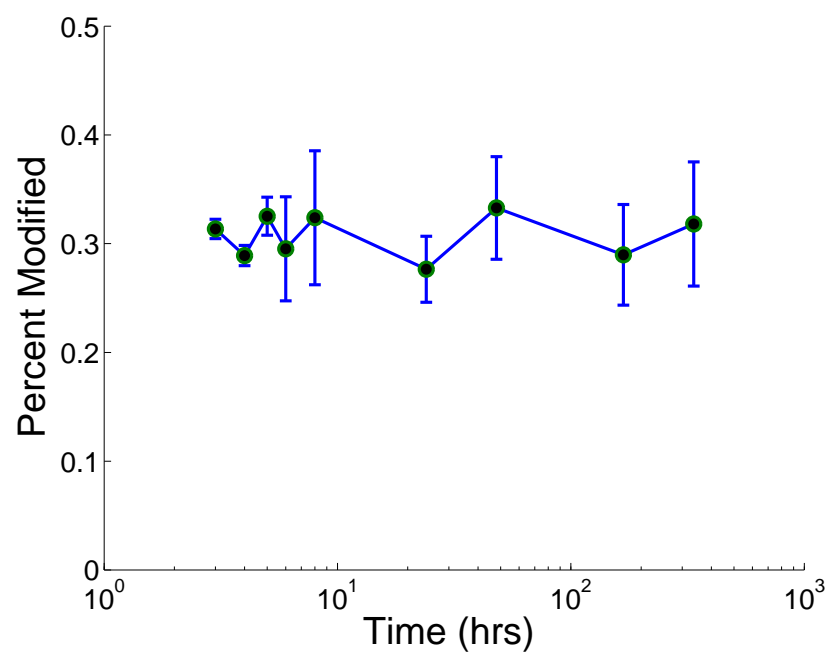
Figure S9: Glutamine to pyroglutamate conversion. Glutamine to pyroglutamate happens to stabilize the protein. This conversion seems to be consistent across both the exponential and stationary phases. *Peptide pQ is artifact, protein N-terminal pQ is stabilizing. Those cases need to be separated.*
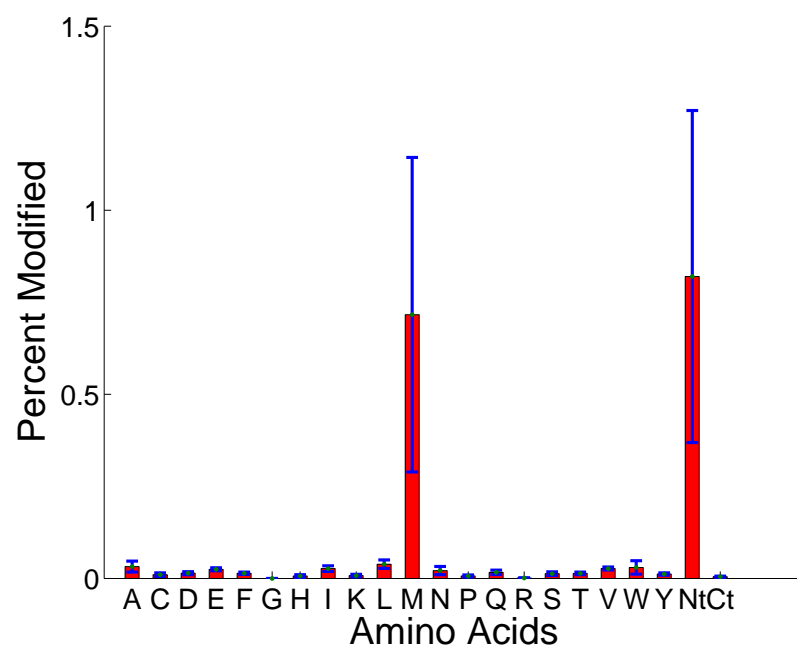
Figure S10: Methionine to Glutamine subsitution. Substitution ...

Supplementary Tables

Table S1: . These .
File: `Figures/mS.csv`