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]

May 6, 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, a program that identifies peptide-spectral matches, 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, especially n-term acetylations increase over time and we were able to identify new protein targets that gets acetylated. Third, we show that sulfoxide reductases fix MetSO to Met during stationary phase, protecting *E. coli* from oxidative damage. Fourth, we found that phosphorylations, even though less abundant, increase under these starvation conditions in *E. coli*. Finally, we found some novel post-translational modifications in this data, a frequent one being 39 dalton mass-shift on Isoleucine on exonuclease ABC subunit C.

1 Introduction

Mass-spectrometry based proteomics [24] offers a unique way to characterize different post-translational modifications [PTMs] associated with proteins. PTMs are key in determining protein function, localization and regulation [33]. Hundreds of PTMs [**?**] [**?**] are known *in vivo*, however most of

the current studies focus on identifying only few PTMs from mass-spec data. This is both because of computational limitations and 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 [29], a naive based spectral alignment algorithm to do an unrestricted search of PTMs from mass-spectrometry based proteomics data 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, if required PTMs associated with it. These search algorithms generally 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 [35], Sequest [13], OMSSA [15], X!Tandem [9] 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 [27]. Even though few of these algorithms report an option to do an unrestricted search, there are very few studies that reported an unrestricted search of PTMs using these algorithms. However, recently, software for unrestricted search of PTMs is starting to be available. MODa [29] is one such unrestricted search engine, along with others such as TagRecon [10] and Byonic [3]. We used MODa, a spectral alignment algorithm that uses dynamic programming to identify small mass-tags and then join them reporting the delta masses between these mass tags. Some of the studies that used MODa were able to identify many post-translational modifications, for example, a study on urine proteomics by Liu et. al [25] identified different PTMs without an enrichment of these post-translational modifications. Here we used MODa, first to do an unrestricted search to identify frequently occuring mass-shifts and focussed on the PTMs that seem abundant in this data or being reported to be critical with *E. coli*.

In our current work, we generated mass-spectrometry data for *E. coli* grown under glucose limiting conditions. As shown in many studies [40] [41], *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 caused by limiting nutrients, change in pH value, accumulation of toxic substances in the flasks etc [31]. 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 [40], 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., [41], 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 PTMs, but very briefly. Other than looking at at the pair-wise correlation of the PTMs identified in different stages of growth, there was no further analysis,

such as PTM levels during different phases of growth, protein targets of these PTMs etc. Few other studies of PTMs focussed on multiple PTMs [19], but they looked at only one "snapshot" of interest i.e., a single time point during the growth or used information from a larger database of processed mass-spectrometry data.

Here, apart from doing an unrestricted search of PTMs, we investigated the levels of the PTMs instead of typical protein quantitation. We then identified the most frequent mass-shifts at different phases of the growth, identified artifacts that increase the sensitivity of the search algorithm, when included and finally looked at the protein targets of some of the relevant PTMs. Since, we have the *E. coli* protoemics data at many time-points during the growth, such analysis would provide a holistic view of proteome function.

2 Results

2.1 Running MODa on *E. coli* proteome

Post-translational modifications 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 an overarching goal of mass-spectrometry based proteomics. However, this is limited by both experimental enrichment techniques for PTMs and computational limitations that can identify all the possible PTMs. The *E. coli* dataset used in this work is not enriched for any PTMs.

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 get 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 peptide level information like most of other search engines, (2) it outputs the frequency of the mass-shifts on different amino-acids along with N- and C- term of the peptides. 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

Protein function is shown to have a link to the PTMs that are associated with it. Since, we have the MODa output of modified peptide spectral matches, we first calculated the amount of proteome modified. We did this for all 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. 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. Percentage modified is the percentage of the number of modified peptide spectral matches (PSMs) to the total (modified + unmodified) peptide spectral matches. We used 1 standard error to plot the variation within 3 biological replicates.

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. The number of identifications are a bit lower when compared to our previous analysis using Sequest [21] on the same dataset. However MODa is an algorithm that is more widely used to unravel new PTMs and one of the goals of the current work is to do a large-scale analyses of these PTMs under varying conditions of growth. The above protein identifications are when we considered all the peptide spectral matches i.e., one-hit wonders too. One-hit wonders are the proteins identified with only 1 PSM mapped to it. Later on, when we identify the protein targets of the PTMs, we used a criteria that a protein is considered if there were at least 2 PSMs (either same peptide appearing twice or at least 2 different peptides). This additional criteria resulted in 1861, 1822, 1857, 1839, 1864, 1892, 1866, 1813 and 1758 protein identifications at each of the 9 time points analyzed in this study.

2.3 Mapping mass-shifts to post-translational modifications

The function of the protein, hence PTMs associated with it, depends on the environment. A changing environment results in different set of protein interactions and modified proteins. Since we have the information on each mass-shift as a function of the time, we investigated how these mass-shifts change over time. For this, we post-processed the output from MODa. MODa outputs mass-shifts in integer masses in the peptide hits. A mass-shift indicates either a PTM or a substitution. The mass-shifts are then mapped 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.

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 (Suppl Table 1). As seen from this table, +1 Da is the most frequent mass-shift at all time points. However this +1 Da modification is generally the result of 13C peak-picking. From the amino-acid profile, we clearly see that this happens randomly on all the amino acids confirming that it is indeed mostly 13C peak-picking (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 mass-shift in detail as oxidative stress is widely studied topic. Other than 42Da (acetylation), -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. 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. 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.

We investigated the frequently occuring peptide spectral matches with mass-shifts. Even though +1Da is the most dominant one, since it occurs randomly on all the amino acids possible, +16Da mass-shift in the peptide spectral matches are the dominant ones and occur almost 7 or 8 times in the top 10 most abundant peptide spectral matches. Here, we provide a table of frequently occuring mass-shifts that don't have the +16 Da or other artifacts such as sodium (+22 Da) and potassium (+38 Da) adducts. Table X summarizes these peptide spectral matches at different points of the time growth. It is clear that, apart from oxidation, and -17, -18 Da that are ammonia and water losses, acetylation is the most dominant post-translational modification that seem to occur at all these times points.

Other than the known PTMs, we found a new +62 Dalton on N-terminal Aspargine, which was not shown to be reported previously. The protein mapped to this peptide is excinuclease ABC subnit C. In Uniprot, there was no information on PTM processing at this site. However, this mass-shift was shown to occur only in one biological replicate of 3 hours, indicating a probable artifact caused by either sample preparation or due to the instrument. Unimod has a +62 Dalton usually reported on lysine. The peptide has a K before the cleavage and hence the adduct must have shifted to this residue on N-term. Another new modified peptide spectral match is K.NMITGAAQMDGAI+39LVVAATDGPMPQTR.E. This modified peptide of elongation factor Tu was found at all the time-points of the growth. This amino acid was not shown to be modified earlier, as indicated by UniProt records. Unimod did not show this mass-shift of 39 Dalton on isoleucine.

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 [30]; elongation factor Tu (EFTu) [[2]; and the chaperone SecB [39]. In addition, a number of heterologous eukaryotic proteins are modified with an NtAc when overexpressed in E. coli [4] [6] [47] [28]. 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 [49]; and RimL, which targets the N-terminal Ser of L12 [44]. The E. coli genome also encodes four putative N-acetyltransferases (yjaB, yhhY, yjhQ, and yjgM) that have been suggested as candidate NATs [[44]. The acetyltransferases responsible for Nt-acetylation of non-ribosomal proteins have not been identified, although RimJ [4] [44] and RimL [28] 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 [28], but can be drastically changed by substitutions in downstream AA positions [6]. 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 [23] [46]. 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 [48].

NtAc has a variety of functions in eukaryotic cells, including regulating protein stability, ER trafficking, protein complex formation, and membrane attachment [43], 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 [7], 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 [6] [17]. 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 [50]. 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 [11] 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 [43]. 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 [23]. 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 PTMs, as these were shown to occur frequently in *E. coli*. 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* [26]. 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 [40] probably pointing to a role of phosphorylation in later stages of the growth cycle. When we looked at the protein targets that get phosphorylated, during the exponential phase, only phosphoglucomutase (pgm) was phosphorylated. At later stages, along with pgm, elongation factor Tu, CTP snythetase, glyceraldehyde-3-phosphate dehydrogenase gapA and other transporters such as ferric citrate outer membrane transporter and copper/silver efflux system membrane component seem to be phosphorylated.

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. Carboxylations and nitrosylations could also be caused due to artifacts, but since we did not see a constant level of these PTMs at different phases of growth, there might be some biological PTMs that have a role in cell function.

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 [16]. 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 and more generally through reductases that reduce this oxidized protein

state back to its original non-oxidized state. These reductases where shown to play a criticial role in protection of proteins against oxidative stress in the past [45]. Here, we investigated the levels of oxidation during the different phases of growth, the protein levels of these reductases at different phases of growth and also looked at oxidizied protein targets.

Oxidation frequently occurs on cysteine and methionine and also has been shown to occur relatively less frequently on few other amino acids. However in a typical mass-spectrometry experiment, the cysteines are capped to avoid disulphide bond formation. So, in our analysis, we expect to see lot of methionine oxidation. To validate this, we plotted the frequencies of +!6 Da mass-shift on different amino acids possible. Suppl Figure 6 shows this profile. It is clear from this amino acid profile that methionine oxidation is the frequent one.

Next, we looked at the time-course of these oxidation levels. For this, We looked at both the global oxidation levels (i.e., oxidation occuring on all amino acids) and just the methionine oxidation level. 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 reactive oxygen species (ROS), we expected the oxidation levels 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 are 2 such reductases, MsrA [1] and MsrB [18], in *E. coli*. yeaA is ortholog of MsrB in the REL606 strain. We were able to look at the protein levels of these reductases, as MODa also outputs the peptide/protein level information for each peptide-spectral match. Figure 8 shows these protein abundances. There seems to be a sudden increase in the levels of these proteins during stationary and long-stationary phases, probably trying to fix MetSO back to Met. This might explain the lower levels of methionine oxidation levels at later stages of the growth. We also validated this result using our previous work [21] on protein level analysis on this data using Sequest, a different protein identification 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 course. Previously elongation factor Tu, chaperonin GroEL and other proteins were identified as major protein oxidation targets. In our study, along with these, we identified some other proteins that are oxidized abundantly. In exponential phase, these were Alkyl hydroperoxide reductase subunit (ahpC), S-methyltransferase , cytochrome d ubiquinol oxidase, elongation factor G, phosphoribosylaminoimidazole-succinocarboxamide synthase. Most of the proteins were also seem to be oxidized at later stages of growth i.e., during stationary and late-stationary phases. The supplementary data lists all the proteins along with their frequencies at all the time points of growth. A separate supplementary information also lists that the peptide, along with the positions at which these are oxidized.

Another mass-shift that seem dominant on methionine is -2 Da. In Unimod, this was reported as methionine to glutamine substitution. An alternate description given in unimod is that it could

be misacylation of the tRNA. However, to cope up dealing with the oxidative stress, methionine substitution to glutamine might be an alternate way for *E. coli* to not let protein being oxidized. However, this substitution decreases from exponential to stationary phase (see Suppl Figure), probably suggesting other likely role.

2.7 Other PTMs and artifacts

Some of the frequently occuring mass-shifts (add 22 Da and 38 Da from Suppl Table 1) were annotated as artifacts in UNIMOD database. These were annotated as Na and K adduct formation that were also shown previously [38]. 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 mass-shifts unique to a position on the peptide (See Data folder in github).

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 have a mass greater than 200Da. Either these large PTMs are very low abundant and difficult to identify via mass-spec or other explanation could be *E. coli* is a simple organism and might not have these high level PTMs, that we typically see in eukaryotes.

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), there were

several interesting insights about the post-translational modifications observed on *E. coli* proteins. 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 acetylations goes up from exponential to stationary phases pointing to a likely role of this modification during late stages of the growth. Similar increase was seen for phosphorylation too. However, oxidation levels 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 protocol developed in this study can be used with any mass-spectrometry based proteomics data to identify, characterize and analyze many post-translational modifications in *tandem*.

Recently, the focus has shifted from identifying peptides/proteins to characterizing PTMs as shown in this recent review article [33]. 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 some 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 [34]. Previous work in this area either used the PTMs deposited in the database [19], or limited the analysis to a single PTM [40].

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 [**?**]. In the current work, we used MODa, a naive based spectral alignment algorithm that uses dynamic programing to identify the peptide-spectral matches, leaving the gaps as mass-shifts. 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. Once we have these, we post-processed the resulting files for a better understanding of the PTMs. We calculated the modified peptide spectral matches at each of the time-points. We calculated the frequently occuring mass-shifts during the entire growth phase. To see if these mass-shifts occur on a single amino acid or a set of amino acids, we plotted amino acid frequencies for the mass-shifts. The program outputs mass-shifts which we then used with UNIMOD to map the mass-shifts to the relevant PTMs. 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. In future, we would like to automate the process of identifying PTM types from mass-shifts, however, for the current work, we did this mapping manually once we identified the mass-shift and the amino acids that were observed to be modified with this mass-shift. All the above-mentioned post-processing of MODa output files resulted in analysis of few mass-shifts representing n-term acetylation, oxidation, phosphorylation, Na and K adducts, along with few other mass-shifts that mapped to carboxylation and nitrosylation.

Some studies in the past use multiple programs to validate the peptide hits. For example, MODa results could be validated with other similar programs such as DeNovo, Byonic or TagRecon. One study by [25] used MODa and DeNovo programs to identify the post-translational modifications. One of the main motivations for using multiple programs is to identify very high confident hits, that in turn could be used for biomarkers etc. However, our focus is not to identify very high confident PTMs, but to get a wider coverage. However, a stringent FDR of 1% that we used with MODa would itself identify confident hits.

Colin's discussion N-terminal processing of proteins is a well-known PTM in most species ranging from eukaryotes to bacteria [22]. One such n-terminal processing is n-terminal acetylation. N-terminal acetylation has been shown in the past to provide insights into the nature of the n-terminal of the proteome [20]. In eukaryotes, n-terminal acetyltransferases are well characterized and studied, for example, as shown in these studies [36], [37]. In *E. coli*, NATs are primarily characterized into rim proteins [44], [49]. 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. Here we have shown that n-term acetylation increases from exponential to stationary phase, Lysine acetylation E. coli, acetylation in exponential and stationary phases ***Colin's discussion***.

Protein oxidation by reactive oxygen species (ROS) is linked to various diseases including aging [42]. ROS generally accumulate at later stages of growth, possibly because of 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. In this work, we see that the oxidation levels seem to go down from exponential to stationary phase. So, we looked at the levels of the methionine sulfoxide reductases. *E. coli* has a class of sulfoxide reductases [5] [51] 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 [14]. 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, pointing that there is some oxidation caused *in vivo*.

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 [25].

Current whole-cell models [8] 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. Another advantage of the protocol we developed here is to study the PTM networks in cohert. Most of the PTM networks studied till date either focussed on a single PTM such as phosphorylation [32] or handful of related PTMs, such as glycosylation and phosphorylation. Most of the search engines currently in use can also use the output of MODa to identify the frequently occuring mass-shifts that they can then input to their favorite search engine to increase the sensitivity

of the peptide hits.

4 Conclusions

The amount of modified protein seems to be at a constant 30% during exponential as well as the long stationary phases. Acetylation, in particular n-term acetylation goes up from exponential to stationary phases. Surprisingly oxidation levels went down from exponential to stationary phases, as sulfoxide reductases that fix MetSO back to Met increased in their enzyme levels, playing a key role in protecting *E. coli* from oxidative damage. We found some new protein targets for n-term acetylation. Phosphorylations, even though identified less frequently, increased from exponential to stationary phase, indicating a likely role of this PTM in late-stationary phases of the growth curve. A novel 39 Dalton mass-shift on isoleucine was shown to be dominant on exinuclease ABC subunit C.

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 with this proteomics data [21]. 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 (cultures grown at different times). Detailed information on the mass-spectrometry is provided in the Houser et. al. paper [21] as well. Trypsin was used to digest the proteins and then the sample wass analyzed using liquid chromatography mass spectrometry (LC/MS) on a LTQ-Orbitrap (Thermo Fisher).

5.2 Post-translational modification identification and analysis

Mass-spectrometry raw data is then converted to a mzXML file format to input into MODa [29]. MODa is a naive based spectral alignment algorithm that identifies peptides and their associated PTMs from the input mzXML spectral files. The program needs few other parameters, such as enzyme used, instrument used to capture the mass-spec data, precursor and product ion mass tolerances, fixed modifications, any rules to apply on the digest, such as semi-tryptic or fully-tryptic in our case, number of modifications per peptide and the mass-range to search for the PTMs. 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. 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.

Using this information, first MODa generates sequences tags from experimental MS/MS spectra. In parallel, the program identifies the mass-shifts between the sequence tags. Identifying sequence tags, and then matching only with the peptides in the sequence library that have these tags reduces the search time and false positives significantly. MODa outputs 3 files. In 1 file, it outputs the peptide spectral matches, then it uses the information in this file to output the frequency of mass-shifts on different amino acids along with c-term and n-term across the entire dataset. In a separate 3rd file, it lists out the number of proteins identified at 1%FDR. 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, depending on which amino acid or a set of amino acids, this mass-shift occurs, we identified the possible PTM it matches to, using UNIMOD database. We did not automate this process, and we did this by manual curation. For example, if we see +16Da mass-shift frequently occuring on methionine, from UNIMOD we know that it is 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. 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.

Once we have the peptide hits from MODa output, we used target-decoy approach [12] 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, oxidation, phosphorylation and methionine to glutamine substitution, we did additional analyses at protein level, as we were interested in protein targets of these modifications. We provided a list of all the peptide spectral matches separately for these modifications in our github repository. We have the peptide spectrum frequencies i.e., which peptide occurs frequently along with the frequency of proteins that have these mods. Some proteins seem to appear in exponential phase, while others in stationary and late-stationary phases. To understand the functionality of these proteins, we used DAVID for GO analysis **Confirm with Colin**.

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

This project was funded by ARO Grant W911NF-12-1-0390. We thank John Houser and Kevin Drew for useful discussions. 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. 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.

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

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

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

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. B. Gonzalez-Flecha and B. Demple. Metabolic sources of hydrogen peroxide in aerobically growing escherichia coli. *J Biol Chem*, 270(23):13681–7, 1995.

17. 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.

18. 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.

19. N. Gupta, S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith, and P. A. Pevzner. Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res*, 17(9):1362–77, 2007.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

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

28. 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.

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

30. 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.

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

32. 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.

33. 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.

34. 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.

35. 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.

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

37. 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.

38. K. Schug and H. M. McNair. Adduct formation in electrospray ionization mass spectrometry ii. benzoic acid derivatives. *J Chromatogr A*, 985(1-2):531–9, 2003.

39. 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.

40. 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.

41. 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.

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

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

44. 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.

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

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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

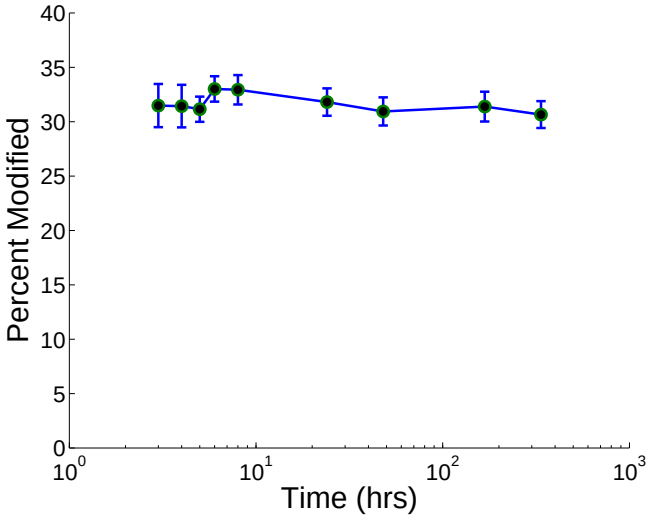


Figure 1. *E. coli* modified proteome. The total number of modified peptide-spectral matches seem to be constant at 30% for all 9 time points.

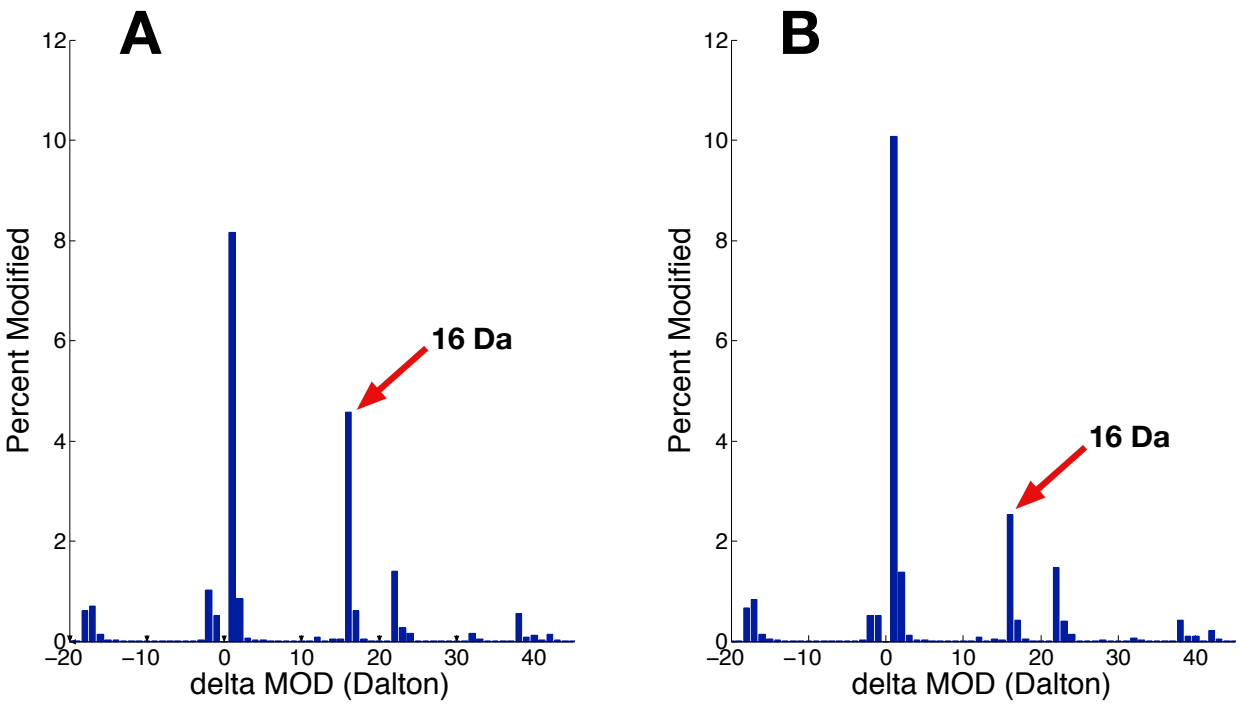


Figure 2. MODa outputs mass-shifts. A naïve based algorithm like MODa can alleviate the requirement of guessing PTMs beforehand. However MODa outputs mass-shifts on the amino acids. We used UNIMOD to map the mass-shift to the most probable PTM, depending on its frequency on the amino acids the mass-shift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

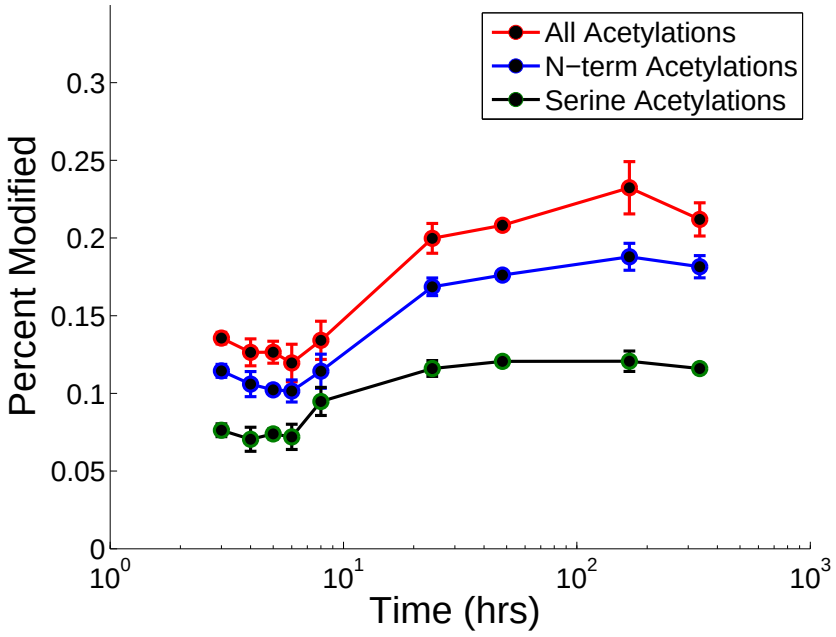
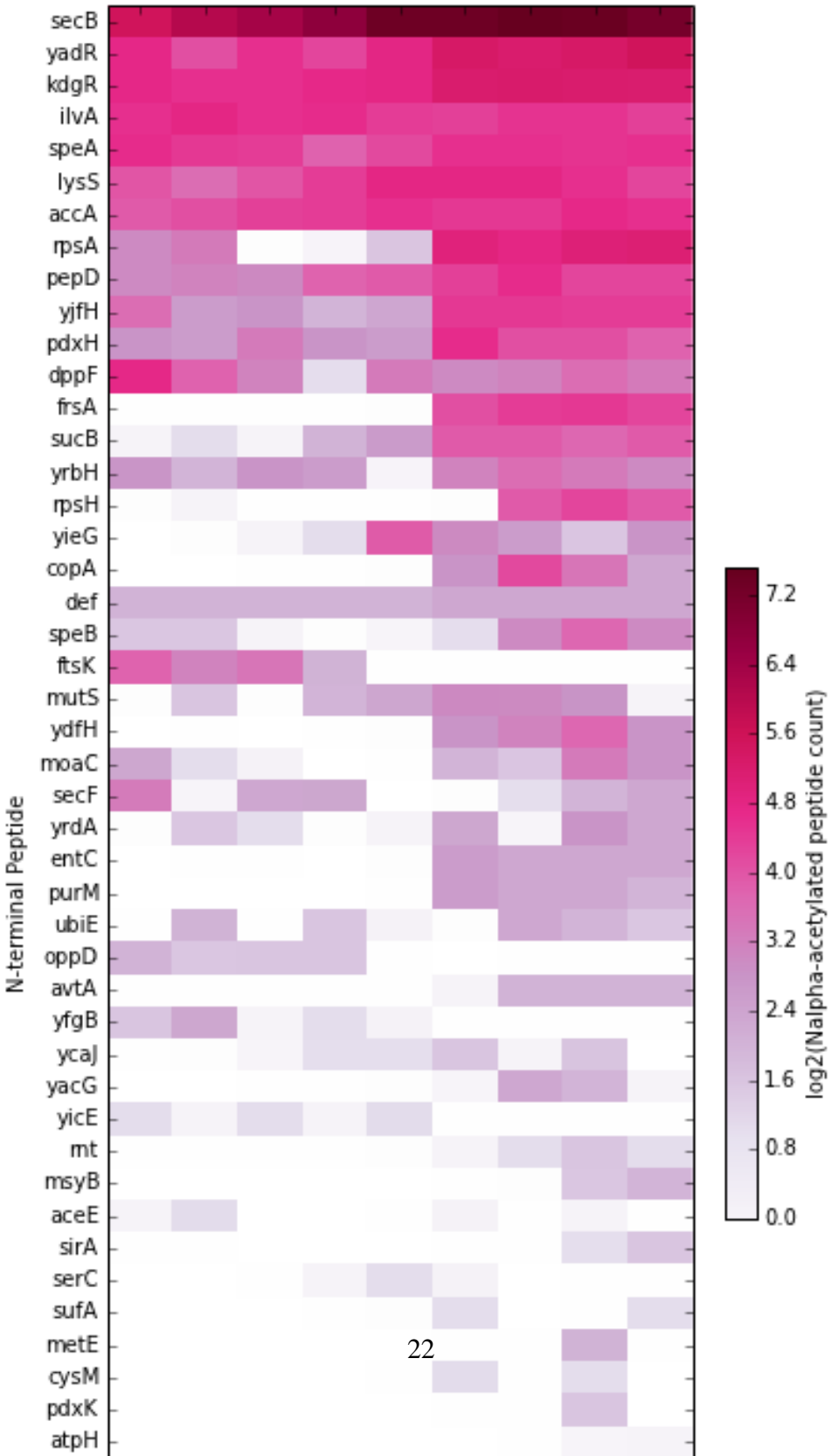


Figure 3. Protein n-term acetylations are dominant. Total number of acetylations as well as the n-term/serine acetylations seem to go up over 2 weeks. *E. coli* grown on glucose generally tend to accumulate acetate, perhaps this resulted in increase in acetylations.
y-axis is the percentage of modified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


secB
yadR
kdgR
ilvA
speA
lysS
accA
rpsA
pepD
yjfH
pdxH
dppF
frsA
sucB
yrbH
rpsH
yieG
copA
def
speB
ftsK
mutS
ydfH
moaC
secF
yrdA
entC
purM
ubiE
oppD
avtA
yfgB
ycaJ
yacG
yicE
rnt
msyB
aceE
sirA
serC
sufA
metE
cysM
pdxK
atpH
N-terminal Peptide
7.2
6.4
5.6
4.8
4.0
3.2
2.4
1.6
0.8
0.0
log2(Nalpha-acetylated peptide count)

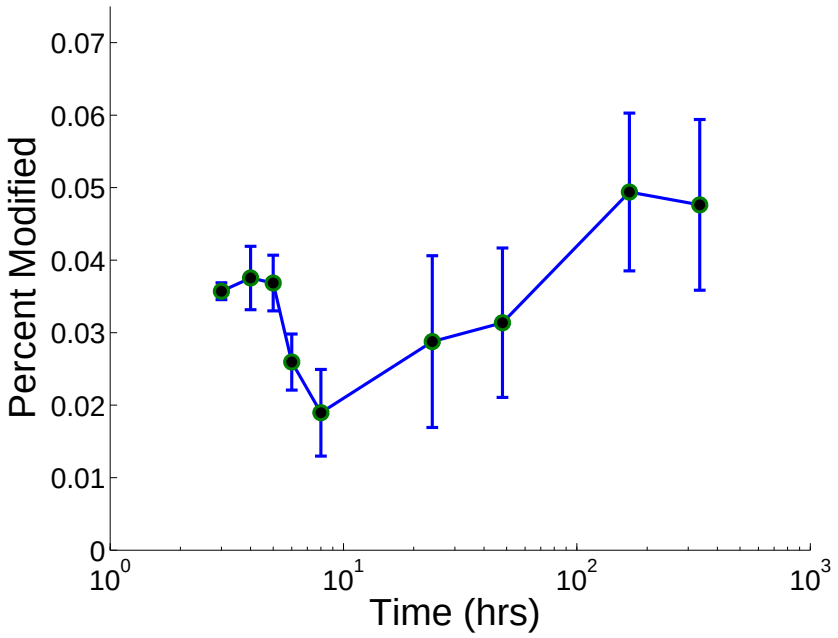


Figure 5. Phosphorylations are rare. Phosphorylations seem to be less abundant in this data set and tend to increase during last week of growth. Phosphoglucomutase (pgl) is the only phosphorylated protein identified during the exponential phase, while there were few proteins that were phosphorylated during stationary phase and long-stationary phase. Some of these proteins were elongation factor Tu, CTP synthetase along with pgl.

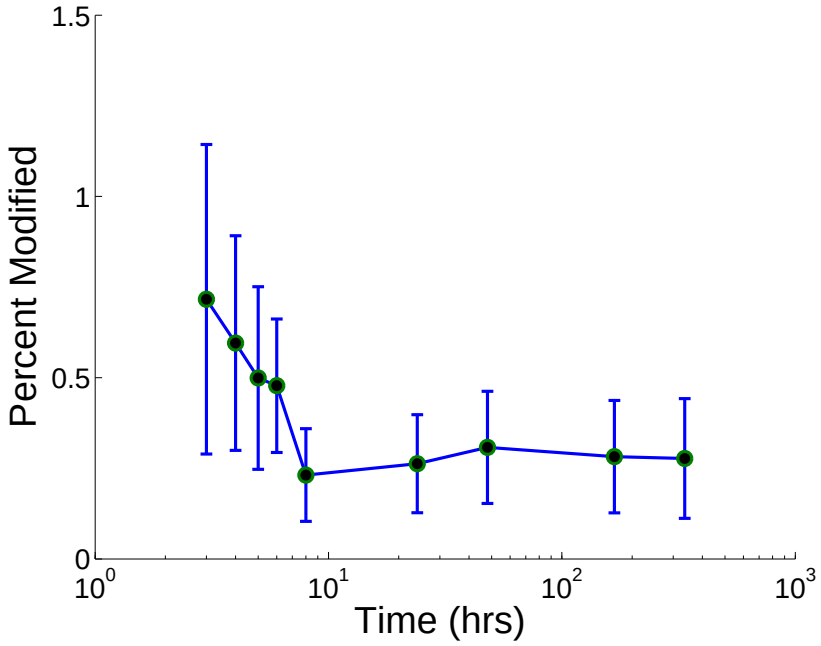


Figure 6. Methionine to Glutamine subsitution. One mass-shift observed on methionine is -2Da. This mass-shift was annotated as methionine to glutamine subsitution in UNIMOD.

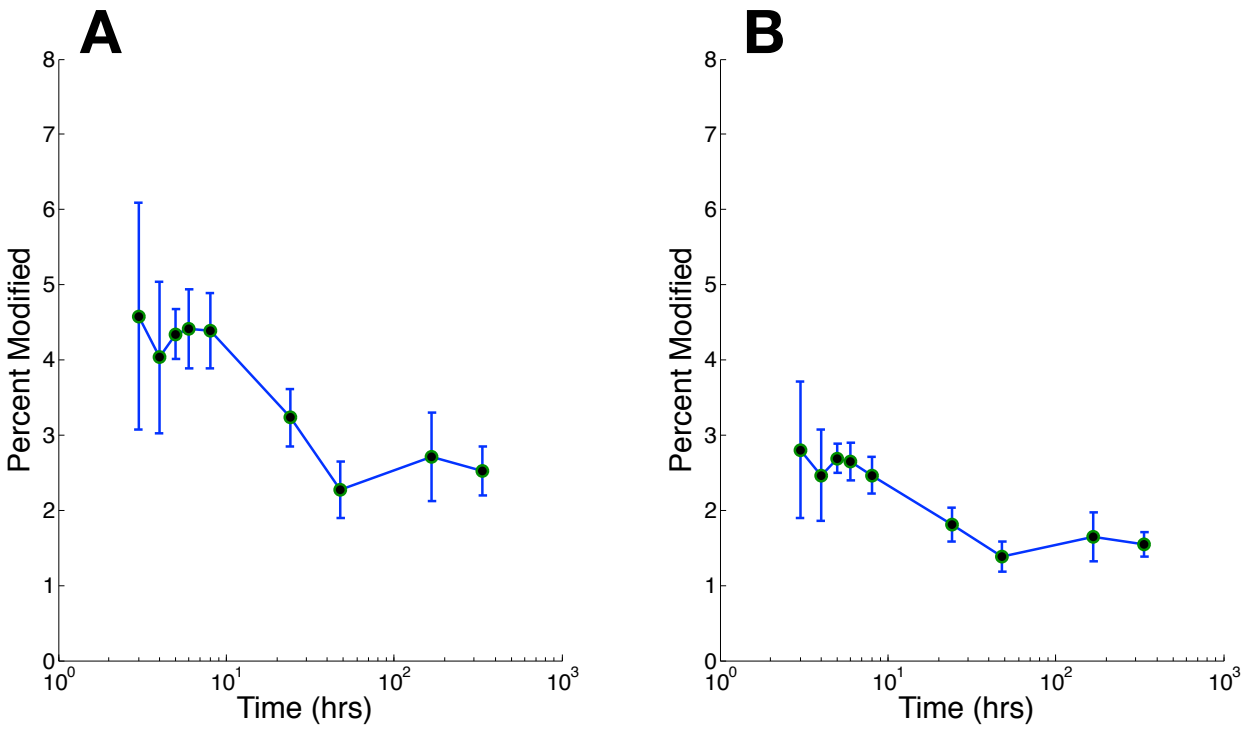


Figure 7. Oxidations go down over 2 weeks. (A) Total number of oxidations seem to go down from exponential to stationary phases. (B) The same trend follows for methionine oxidations too.

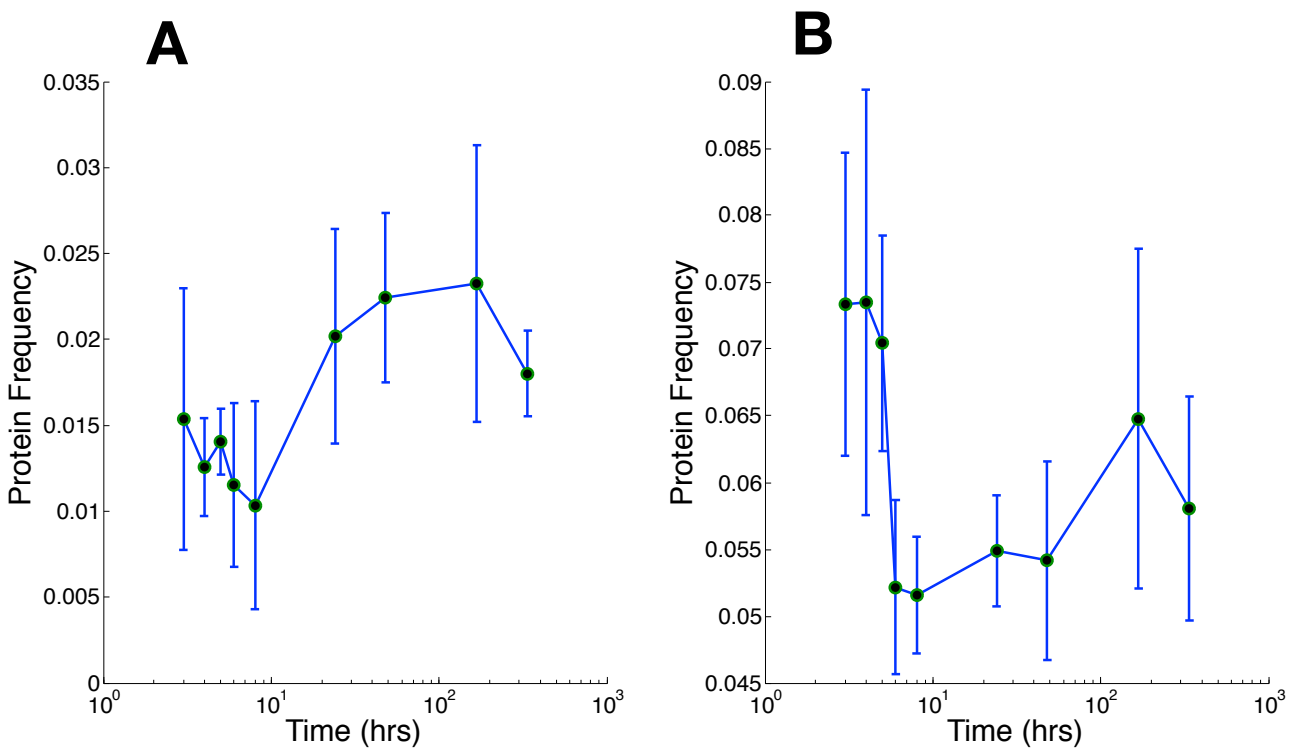


Figure 8. Relative protein abundances of methionine sulfoxide reductases MsrA and MsrB. (A) Protein abundances of MsrA seems to increase going from exponential to long-stationary phase. One of the sites oxidized on MsrA is FQAA[M+16]LAADDDR. (B) MsrB also seems to increase going into the long-stationary phase, protecting *E. coli* from oxidative damage.

Tables

Supplementary Figures

!ht

Table 1. This table lists the top 3 abundant hits at each one of the biological replicates and all the time points of the growth

Frequency	Peptide	Biological Replicate
28	M.S+42EQNNTEMTFQIQR.I	10
25	R.I+2TTSHYDVGGAAIDNHGQPLPPATVEGCEQADAVLFGSVGGPK.W	10
13	K.TPGHPEVGYTAGVETTTGP+2LGQGIANAVGMAIAEK.T	10
71	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	11
39	M.S+42EQNNTEMTFQIQR.I	11
28	R.KMDAFIQYGIVAGVQ+2AMQDSGLEITEENATR.I	11
57	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	12
34	M.S+42EQNNTEMTFQIQR.I	12
23	K.EGVITVEDGTGLQDE+23LDVVEGMQFDR.G	12
67	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	13
32	M.S+42EQNNTEMTFQIQR.I	13
25	K.EGVITVEDGT+23GLQDELDVVEGMQFDR.G	13
37	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	14
33	M.S+42EQNNTEMTFQIQR.I	14
28	M.S+43EQNNTEMTFQIQR.I	14
55	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	15
41	K.AEIFSLFAENN+23MELTDSTIVTPGLR.F	15
29	M.S+42EQNNTEMTFQIQR.I	15
15	M.S+42EQNNTEMTFQIQR.I	16
13	K.LVSW+32YDNETGYSNK.V	16
8	M.S+42LNFLDFEQPIAELEAK.I	16
22	M.S+42EQNNTEMTFQIQR.I	17
12	K.EGVITVEDGT+23GLQDELDVVEGMQFDR.G	17
11	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	17
26	M.S+42EQNNTEMTFQIQR.I	18
10	R.I+2TTSHYDVGGAAIDNHGQPLPPATVEGCEQADAVLFGSVGGPK.W	18
10	M.A+42DSQPLSGTPEGAEYLR.A	18
26	M.S+42EQNNTEMTFQIQR.I	19
15	K.NMITGAAQMDGAI+39LVVAATDGPMPQTR.E	19
14	K.DGAVEAEDRVTI+2DFTGSVDGEEFEGGK.A	19
53	M.S+42EQNNTEMTFQIQR.I	20
32	K.NMITGAAQMDGAI+39LVVAATDGPMPQTR.E	20
20	K.EISPYSIVGLSATW+32DVTK.N	20
48	M.S+42EQNNTEMTFQIQR.I	21
29	K.EGVITVEDGTGLQDE+23LDVVEGMQFDR.G	21
17	K.EGVITVEDGTG+24LQDELDVVEGMQFDR.G	21
47	M.S+42EQNNTEMTFQIQR.I	22
17	K.DGAVEAEDRVTI+2DFTGSVDGEEFEGGK.A	22
14	K.M+32MTMLDPANAER.Y	22
30	K.M+32MTMLDPANAER.Y	23
25	M.S+42EQNNTEMTFQIQR.I	23
20	K.EGVITVEDGTGLQDE+23LDVVEGMQFDR.G	23
27	K.EGVITVEDGTGLQDE+23LDVVEGMQFDR.G	24
25	M.S+42EQNNTEMTFQIQR.I	24
23	K.EGVITVEDGTG+23LQDELDVVEGMQFDR.G	24

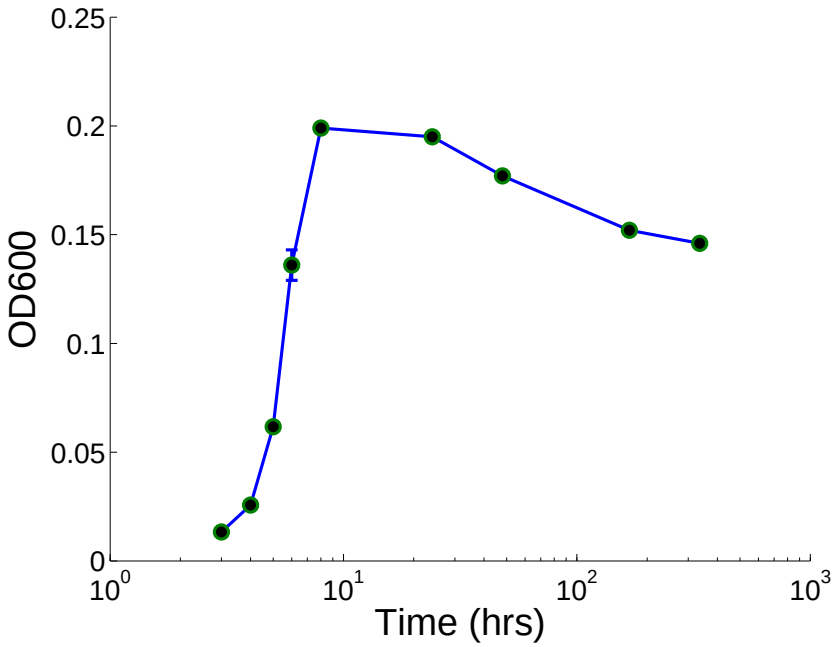


Figure S1: OD600 curve . Growth curve (OD600) of REL606 under glucose starvation conditions.

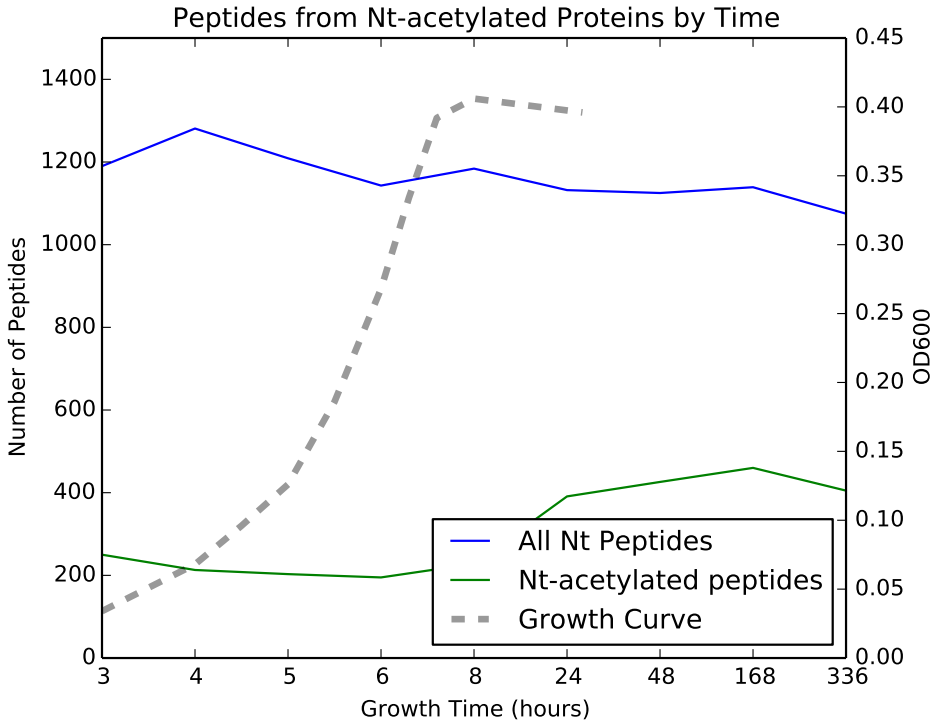


Figure S2: N-terminal Nalpha Acetylated Peptide Occurance across Growth Timepoints. Blue line represents total count of N-terminal peptides with any (or no) modification originating from parent proteins with at least one Nt-acetylated peptide in the dataset. Green line represents total count of Nt-acetylated peptides from these proteins. Both counts are the sum of three biological replicates for each timepoint. Dashed line (right axis) shows OD600 values for an example culture grown under identical glucose-starvation conditions.

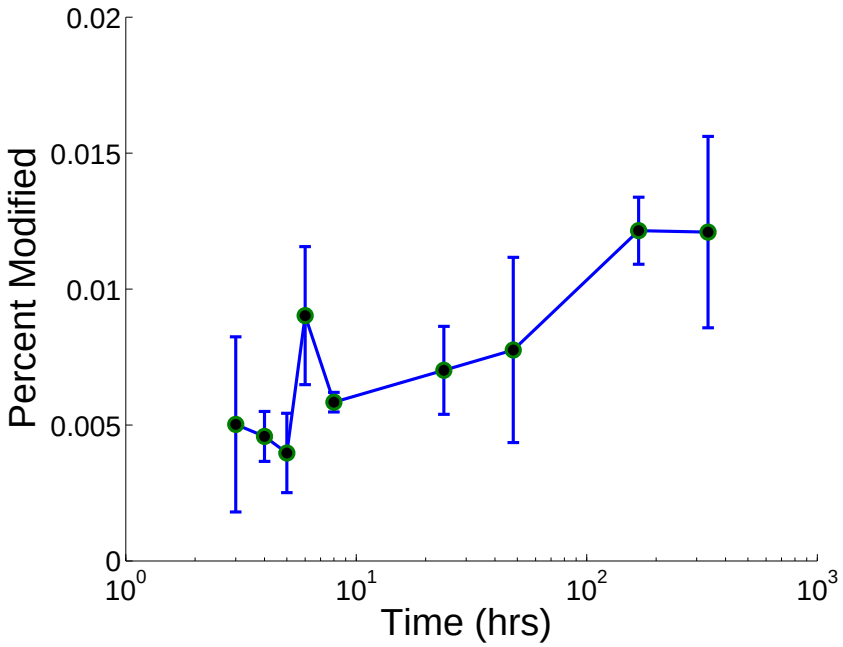


Figure S3: E. coli Carboxylations . Carboxylations seem to go up.

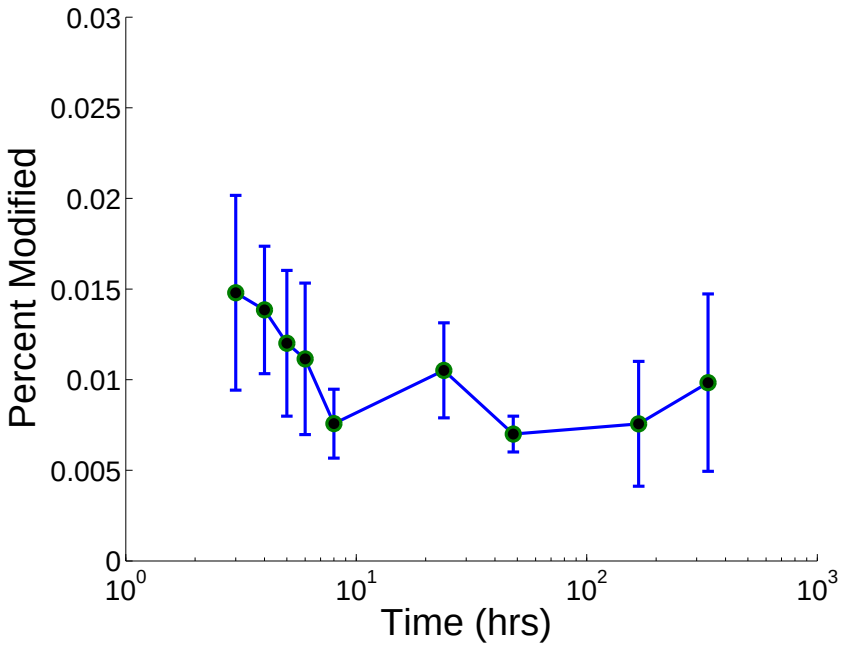


Figure S4: E. coli Nitrosylations. Nitrosylations seem not to change much during the entire growth period.

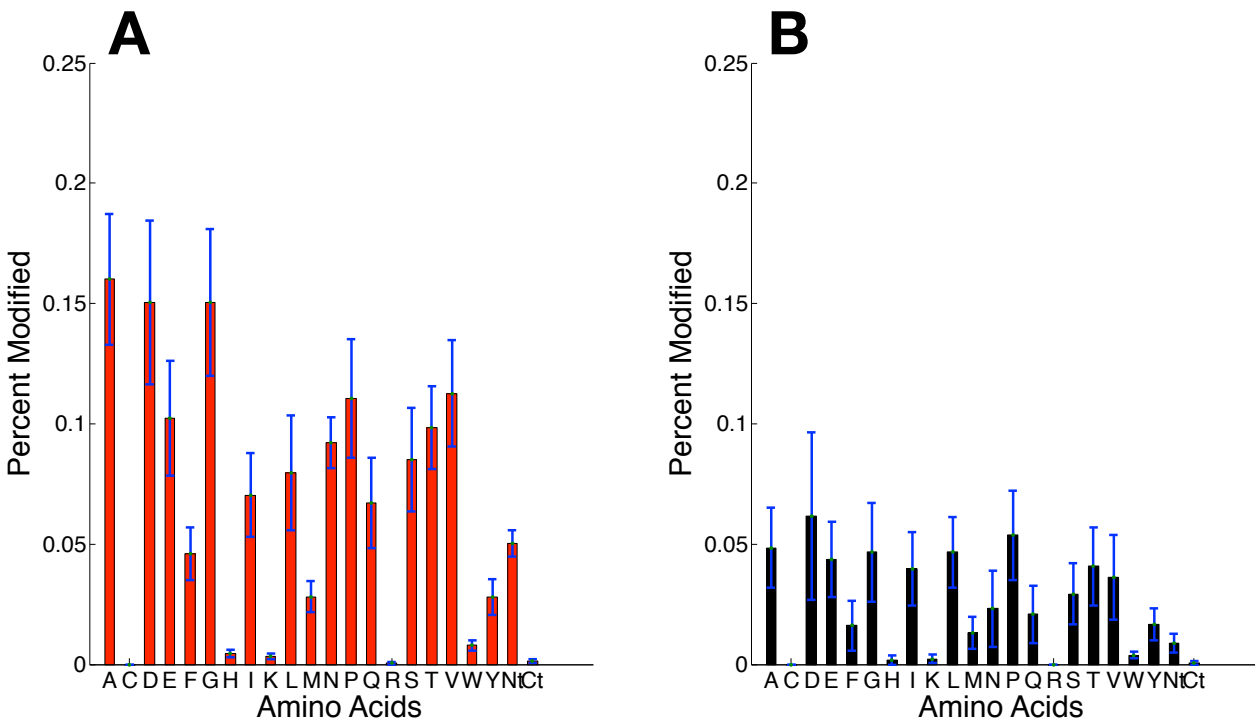


Figure S5: Na and K adducts. (A) Na and (B) K adducts seem to happen on all amino acids except histidine, lysine and arginine. H, K and R are amino acids that are basic and carry some (+) charge at physiological pH. Here Nt and Ct represent n-termins and c-terminus of the peptide.

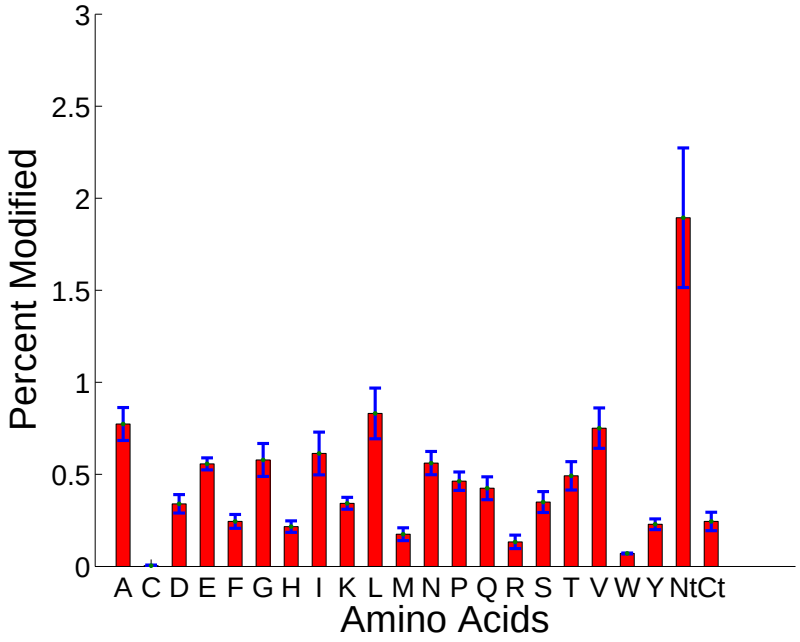


Figure S6: +1Da shift occurs randomly. We cannot infer that this +1 Dalton mass-shift is deamidation as it occurs randomly on all amino acids, inferring it is carbon-13 (C13) peak picking as previously shown in many studies.

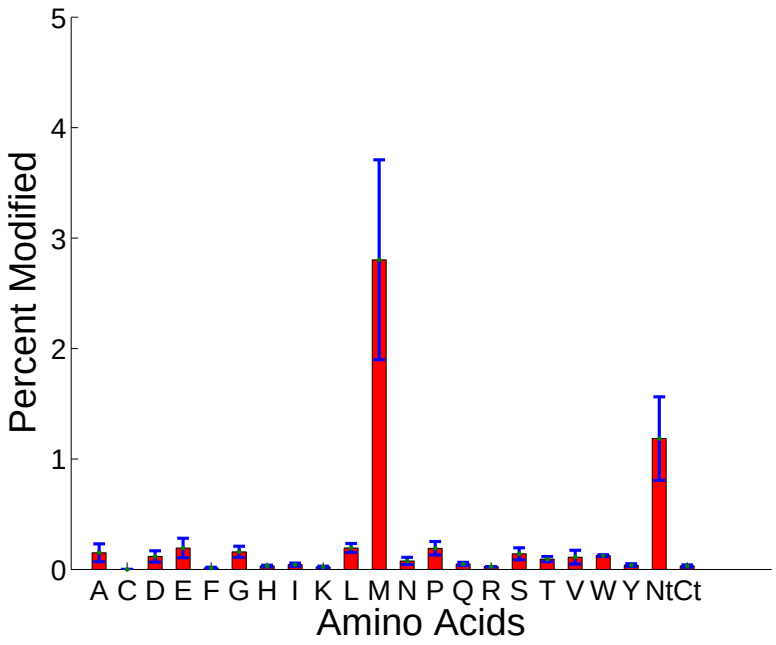


Figure S7: Oxidation is dominant on methionine. Even though many amino acids could be oxidized, in this data set, oxidation seems to occur primarily on methionine, as expected.

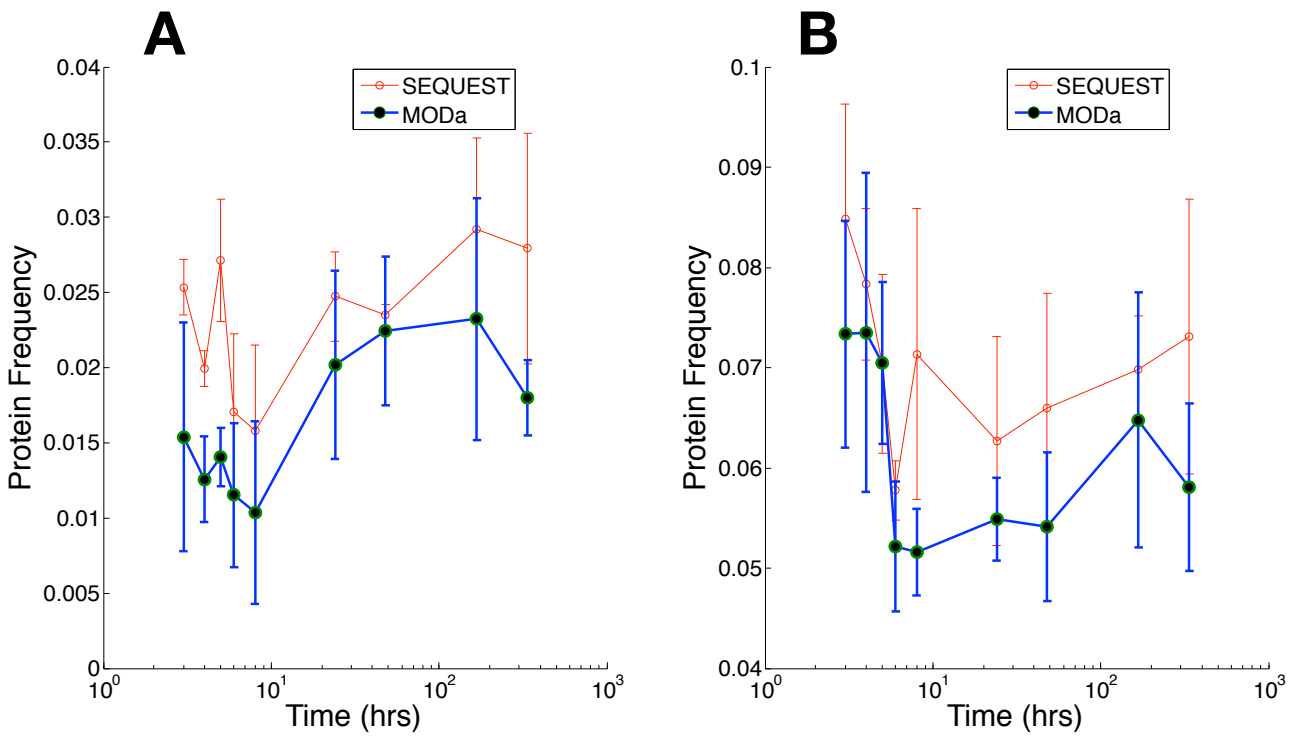


Figure S8: Sequest and MODa results agree. We compared MODa protein abundance results with Sequest to double-check the protein levels of the sulfoxide reductases that protect *E. coli* from oxidative stress by fixing MetSO to Met. As seen, MODa and Sequest results seem to agree well for (A) MsrA and (B) MsrB.

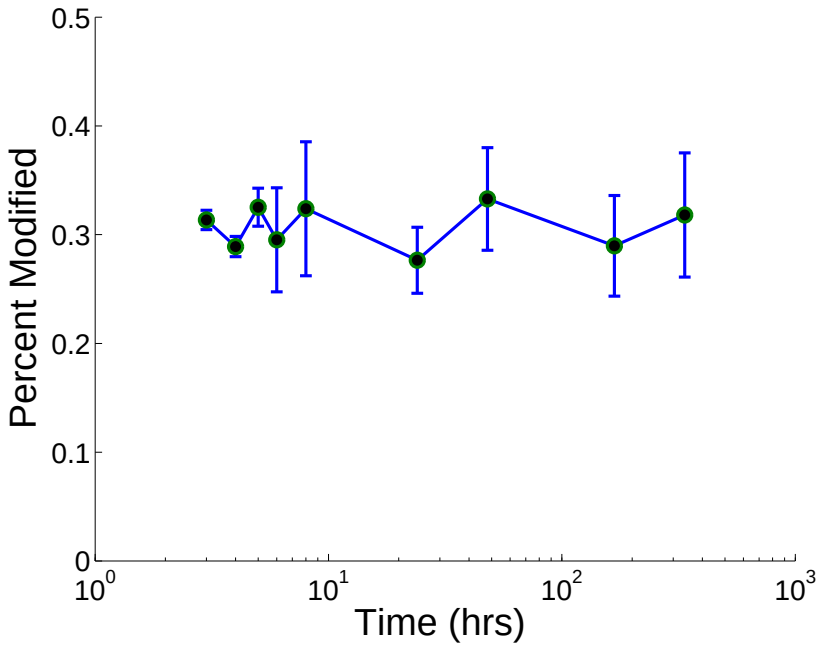


Figure S9: Glutamine to pyroglutamate conversion. Glutamine to pyroglutamate happens to stabilize the protein. This conversion seems to be consistent across both the exponential and stationary phases. ***Peptide pQ is artifact, protein N-terminal pQ is stabilizing. Those cases need to be separated.***

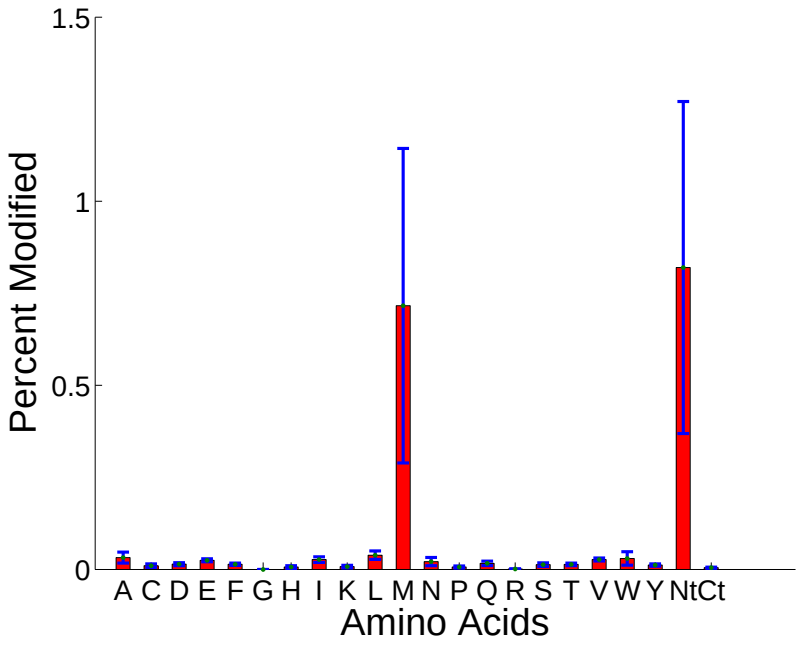


Figure S10: Methionine to Glutamine subsitution. Substitution ...

Supplementary Tables

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