bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

RESEARCH

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

Colin W Brown[1], Viswanadham Sridhara[2], Daniel R Boutz[1,3], Maria D Person[1,4], Edward M Marcotte[1,3,5], Jeffrey E Barrick[1,3,5] and Claus O Wilke[1,2,6*]

*Correspondence:
[1] Institute for Cellular and Molecular Biology, The University of Texas at Austin, Speedway, Austin, USA
[2] Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Speedway, Austin, USA
[6] Department of Integrative Biology, The University of Texas at Austin, Speedway, Austin, USA
Full list of author information is available at the end of the article

Abstract

Background: Post-translational modification (PTM) of proteins is central to many cellular processes across all domains of life, but despite decades of study and a wealth of genomic and proteomic data the biological function of many PTMs remains unknown. This is especially true for prokaryotic PTM systems, many of which have only recently been recognized and studied in depth. It is increasingly apparent that a deep sampling of abundance across a wide range of environmental stresses, growth conditions, and PTM types, rather than simply cataloging targets for a handful of modifications, is critical to understanding the complex pathways that govern PTM deposition and downstream effects.

Results: We utilized a deeply-sampled dataset of MS/MS proteomic analysis covering 9 timepoints spanning the *Escherichia coli* growth cycle and an unbiased PTM search strategy to construct a temporal map of abundance for all PTMs within a 400 Da window of mass shifts. Using this map, we are able to identify novel targets and temporal patterns for N-terminal Nα acetylation, C-terminal glutamylation, and asparagine deamidation. Furthermore, we identify a possible relationship between N-terminal Nα acetylation and regulation of protein degradation in stationary phase, pointing to a previously unrecognized biological function for this poorly-understood PTM.

Conclusions: Unbiased detection of PTM in MS/MS proteomics data facilitates the discovery of novel modification types and previously unobserved dynamic changes in modification across growth timepoints.

Keywords: post-translational modification; proteomics; prokaryote

Background

Post-translational modification of proteins (PTM) is a ubiquitous paradigm for dynamic cellular response and information transfer across all kingdoms of life [1]. Although historically PTM has been studied in the context of discrete and tightly-regulated signal transduction systems such as eukaryotic histone proteins [2], kinase cascades [3, 4], and prokaryotic two-component systems [5], it is only relatively recently, with the development of tandem-mass-spectrometry-based proteomics, that the abundance and complexity of PTM has become apparent [6]. A surprising result from many of these investigations has been that the phylogenetic distribution of many PTMs is much wider than had been assumed. A number of PTM types previously thought to be restricted to eukaryotic and metazoan species, such as lysine acetylation [7], serine/threonine phosphorylation [8], tyrosine phosphorylation

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[9, 10], and ubiquitination-like protein ligation [11], are now known to be relatively common in prokaryotic proteomes as well. This realization, in combination with the recognition that PTM plays a critical role in growth and virulence of important prokaryotic pathogens (e.g. [9, 12, 13, 14, 15, 16]), has highlighted the need for a better understanding of prokaryotic PTM and particularly the need for deeper, proteome-scale analysis of prokaryotic PTMs.

In response to these needs, much progress in recent years has been made in the mapping of important PTMs across a wide range of prokaryotes [7, 17, 1]. However, the vast majority of these studies are limited by only examining a handful of easily-achieved culture conditions and timepoints, and by only examining a single PTM type in isolation. The former limitation is especially important, as batch cultures grown for short time periods in rich media, as is most common for bacterial proteomics experiments, may be a poor reflection of the high-stress, nutrient-starved conditions in which bacteria spend most of their time in the wild [18, 19]. While collecting bacterial samples directly from their natural habitat is generally infeasible for proteomics experiments given the requirements for large cell numbers and pure samples, the starvation conditions commonly encountered in a bacterium's native habitat are thought to be largely recapitulated in long-term batch culture [18, 19]. As an exponentially-growing batch culture exhausts the readily available nutrients in the growth medium, the cells undergo a regulated transition into stasis by activating a stereotypic stress response. This response usually involves a decrease of or complete stop to cell division, steep dropoffs in oxidative metabolism [20] and protein synthesis [21], sequestration of ribosomes [22, 23], activation of oxidative damage response systems [24, 25], and increased protease-mediated protein turnover [26]. Eventually, even this inactive state becomes unsustainable for the majority of cells in the culture, and a large-scale die-off takes place until the culture reaches an equilibrium where the remaining cells are able to survive on the nutrients liberated from their less fortunate culture-mates. This "deep stationary" phase of batch culture is poorly understood, but is characterized by a gradual loss of culturability (the **V**iable **B**ut **N**on **C**ulturable state [27]), likely related to accumulated cell damage, and a dynamic equilibrium of genetic changes as mutations advantageous for stationary phase growth (**G**rowth **A**dvantage in **S**tationary **P**hase, or GASP mutations [19]) are fixed by selection in the population. The low rate of protein synthesis and the potential importance of nonenzymatic protein-damage modifications in stationary phase makes an understanding of PTM chemistry and dynamics during this portion of the growth cycle especially important.

With a few very recent exceptions [28, 29], studies of PTMs at different growth phases in *E. coli* have been restricted to either a single modification or a handful of pre-specified modifications (e.g. [30, 31, 32, 10]). This limitation is largely due to both the relatively low abundance of the PTMs examined, necessitating the enrichment of modified peptides using PTM-specific antibodies [6] or chromatographic separations [33], and data analysis tools that are only useful for examining a small number of pre-specified PTMs. While enrichment is necessary for relatively transient modifications such as phosphorylation, particularly where a broad survey of targets rather than PTM dynamics is the experimental goal, it has a critical shortcoming in that it makes quantitative comparisons among PTM types, and perhaps more

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

importantly between modified and unmodified copies of an individual protein, impossible. Adding to this problem is the fact that many of the most commonly-used software packages for MS/MS spectrum–peptide sequence matching (e.g. Mascot [34], Sequest [35], OMSSA [36], or TANDEM [37]) are limited by the need to create an *in silico* database of theoretical spectra using an existing peptide library; while this approach facilitates rapid searching, it makes searches involving more than a few PTM types computationally unwieldy. The spectrum of PTMs beyond a handful of well-studied examples is therefore largely unexplored.

In this work we utilize a recently developed computational tool for unrestricted analysis of PTMs in MS/MS proteomics data, MODa [38], to examine a unique proteomic dataset [39] covering 9 timepoints of the *E. coli* REL606 growth curve in minimal glucose media from early exponential growth (3 hours post-innoculation) to deep stationary phase (336h, or 2 weeks post-innoculation). MODa uses a combination of *de novo* sequence-tag matching and spectral alignment to make assigning PTM-containing spectra across a wide range of mass shifts computationally tractable, and this allows us to construct an unbiased PTM spectrum across all phases of growth for all modifications from −200 Da to +200 Da. The fine temporal resolution of our dataset then allows us to identify novel temporal trends in a number of PTMs, including N-terminal Nα acetylation, C-terminal glutamylation, and asparagine deamidation. In addition, the lack of bias or enrichment for specific PTMs allows us to track behavior of modified and unmodified proteins across the growth cycle, and to identify a potential functional relationship between N-terminal acetylation, protein oxidative damage, and stationary-phase protein degradation.

Results

We took advantage of a previously existing LC-MS/MS proteomics dataset [39] isolated from 3 biological replicate cultures of *E. coli* B REL606 sampled across 9 timepoints, from early exponential phase (3h post-innoculation) to extended late stationary phase (336h, or 2 weeks post-innoculation). The raw spectra from this dataset were used for simultaneous spectrum-sequence matching and PTM identification using the hybrid fragment matching/spectral alignment software MODa [38]. To reduce computation time and limit the occurrence of false positives, we restricted the MODa search to single-peptide mass shifts of +/− 200 Daltons, with one PTM allowed per peptide spectral match (PSM). To further limit the occurrence of false positive matches, we used the MODa "correct match" probability [38] to calculate the false discovery rate (FDR) and construct subsets of the highest-probability PSMs with 5% and 1% FDR (hereafter referred to as FDR5 and FDR1, respectively). The samples in our analysis were treated with IAA to modify cysteines with a 57 Da carbamidomethyl group; during MODa analysis, this was treated as a static modification to cysteine (i.e. all modifications were relative to the molecular weight of Cys + 57Da). However, this results in an incorrect mass shift for any Cys PTMs that prevent carbamidomethylation (e.g. oxidation), so we added 57 Da to all Cys modifications to ensure that mass shifts for these modifications matched those for non-Cys residues.

We identified a total of 2,527,135 PSMs across all 27 samples, corresponding to a total of 32,755 peptides that occur in at least one sample; these peptides represent

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3,544 individual proteins when all timepoints are considered (Table 1). FDR filtering lowers these numbers substantially, yielding 1,980,884 PSMs and 22,776 unique peptides across 2,445 proteins in the 5% FDR set, and 1,473,636 PSMs and 19,265 unique peptides across 2121 proteins in the 1% FDR set (Table 1). These filtered numbers are in agreement with previous proteomic experiments in *E. coli* [39, 40, 31], with the slightly lower number of proteins in our analysis, likely a result of the reduced sensitivity inherent in the larger search space used by MODa.

A large fraction of the *E. coli* proteome undergoes PTM during growth and starvation in glucose

Of the 1,473,636 PSMs identified across all timepoints in the 1% FDR dataset, a remarkably large fraction, 198,277 (13.5%), are predicted by MODa as having a putative PTM. These modified PSMs corresponded to 8,369 out of 19,265 unique peptides (42%) having at least one modification in any sample, and 1,690 out of 2121 proteins (79.7%) having at least one modification on any constituent peptide. Interestingly, the proportion of the proteome predicted to have at least one PTM remains relatively constant across time points and biological replicates. PSMs, unique peptides, and proteins all show very little change in the proportion of overall PTM across all 9 time points (Figure 1).

Composition of the *E. coli* PTM spectrum

A unique feature of our analysis strategy is the ability to conduct an unbiased search for spectra matching post-translationally modified peptides across a wide range of possible mass shifts. We used MODa to search our raw spectral data for 400 potential peptide mass shifts, ranging from −200 Da to +200 Da; counts of PSMs for this range of mass shifts are shown in Figures 2 and S1.

The overall abundance of individual mass shifts varies widely, with the most abundant mass shifts corresponding to small functional group modifications. The most abundant mass shift is a neutral gain of 1 Da (84,357 PSMs, 45% of all modified PSMs). In addition to simple hydrogenation, this mass change can result from a number of more complicated modifications and MS artifacts; see Discussion. Other abundant mass shifts include oxidations (+16 Da, 46,244 PSMs, 24% of all modified PSMs; +32 Da, 1,563, 0.8% of all modified PSMs), metal ion adducts such as sodium (+22 Da, 8,882 PSMs, 4.7% of all modified PSMs) and potassium (+38 Da, 1,490 PSMs 0.79% of all modified PSMs), and neutral losses such as deamidation (−17 Da, 9,780 PSMs, 5.2% of all modified PSMs) and dehydration (−18 Da, 7,169 PSMs, 3.8% of all modified PSMs).

Commonly studied regulatory PTMs are relatively rare in our data, most likely due to their low abundance in the proteome and the fact that our samples did not undergo enrichment for specific modifications prior to analysis. Although a large number of apparent acetylations (+42 Da) were identified, only a handful of these map to known acetylated lysine residues [32, 30, 41]. A small number of phosphorylations (+80 Da) were identified, although the majority of these are modifications to an active-site serine that acts as a phosphoryl group donor during catalysis in the metabolic enzyme phosphoglucomutase (see Table 2). A list of counts for all mass shifts recovered by MODa is included in Additional File 12.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Distribution of target amino-acid residues varies widely among mass shifts

The most commonly modified amino acid across all timepoints is methionine—nearly all of these modifications are a +16 Da shift corresponding to oxidation (see Discussion)—followed by the hydrophobic amino acids Ala, Val, Leu, Ile; amide-containing amino acids Asn and Gln; and their carboxyl counterparts Asp and Glu (Table 3). The observation of a large number of modifications on amino acids with hydrocarbon side-chains, which are generally not expected to undergo PTM, can likely be explained by a combination of incorrect assignment of a mass shift to the amino acid (AA) by MODa, modification of the backbone NH or CO groups, or selection of peaks with isotopically shifted masses during MS2.

We constructed the distribution of targeted amino acids for each mass shift by counting occurrences of each mass shift–AA pair across all nine time points. We observed significant differences among mass shifts in preference for a single type (or, in some cases, groups) of amino acid residues; the +22 Da and +38 Da modifications, for example, show a broad distribution across AA types, while +16 Da and −2 Da show strong (though not exclusive) preference for methionine. To quantify these differences in AA distribution, we ranked mass shifts by the ratio: PSMs for most common AA / mean(PSMs for all other AAs). AA distributions for the the top ranked (most biased towards one AA across multiple biological replicates) mass shifts are shown in Figure 3.

A large number of modifications with a strong preference for cysteine residues were identified in all three biological replicates; most of these are likely artifacts of IAA treatment during sample preparation, and correspond to common modifications co-occuring with carbamidomethylation (+57 Da), e.g. +58 Da (57 + 1 Da), +59 Da (57 + 2 Da), and +40 Da (57 - 17 Da). The +209 Da mass shift, corresponding to a carbamidomethylated dithiothreitol modification of the cysteine thiol group, is a minor artifact of the reduction and alkylation of cysteine during sample preparation. The +48 Da mass shift was almost exclusively found at catalytic cysteine residues in a handful of proteins, and corresponds to the hyperoxidation of the cysteine thiol group (Cys-SH) into cysteic acid (Cys-SO_3H). This modification is likely to be inactivating and irreversible, resulting in the increased accumulation of the modified form throughout the stationary phase. Among modifications targeting non-Cys residues, putative oxidative modifications show the strongest bias towards a single AA, with the +32 Da and +16/+17 Da modifications showing strong preferences for tryptophan and methionine, respectively. The eighth-ranked −48 Da modification is likely also a result of oxidation via dethiomethylation of methionine residues [42]. The strong preference of the acetylation mass shift (+42 Da) for serine is largely due to modifications on protein N-termini (see Section "N-terminal and C-terminal modifications"). A table of counts for each mass shift-amino acid pair is included in Additional File 12.

N-terminal and C-terminal modifications

To search for modifications that preferentially occur at protein N and C termini, we used Fisher's exact test (FET) to compare the ratio of modified : unmodified counts of each mass shift occurring at the N or C terminus of a protein to the same ratio for mass shifts occurring at all other positions. FET p-values for N-terminal and C-terminal enrichment were calculated for all mass shifts within each

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

biological replicate and filtered for consistency by requiring all three replicates to have $p < 0.05$. Nt- and Ct-biased mass shifts are shown in Tables 4 and 5. We also examined the distribution of unique modified positions for these Nt- and Ct-biased mass shifts as a function of normalized protein length, to determine whether the observed positional bias was a general feature of the mass shift or due to a small number of highly abundant modified positions (Figures 4 and 5).

Eight mass shifts were identified as Nt-biased after filtering (Table 4 and Figure 4). The strongest Nt preference is displayed by the +42 Da mass shift, corresponding to N-terminal acetylation, with modified N termini representing 35–42 % of total observed counts for positions with at least one +42 Da count. The remaining Nt-biased mass shifts fall into two broad categories. The first are rare modifications that occur at a small number of positions at high frequency, such as the +28 Da mass shift (possible retention of formylation on an Nt-terminal fMet, 12–32%), the +64 Da mass shift (possible modification by acetate, 8–22%) and the +172 Da mass shift (100% in all replicates). The second category is comprised of common modifications that occur at low frequency across a larger number of positions; this includes oxidation (+16 Da, 10–14%), most commonly of a retained Nt methionine, and hydrogenation (+1 Da, 2–3%). A beneficial feature of our analysis is the ability of MODa to identify modified N-terminal residues even in the presence of un-annotated N-terminal methionine cleavages. For the protein SecB, for example, we recovered abundant N-terminal peptides which had both undergone N-terminal Met cleavage and putative acetylation at the penultimate N-terminal Ser residue (see Additional File 12), despite the fact that this protein had not been annotated as having its N-terminal Met cleaved in the UniProt database.

Only three mass shifts were identified as Ct-biased after consistency filtering (Table 5 and Figure 5). Two of these, +129 Da (16–24% of counts at C-terminal positions across the three replicates modified, compared to $< 1\%$ of counts at all other positions) and +130 Da (6.5–14% of counts at C-terminal positions across the three replicates modified, compared to $< 1\%$ of counts at all other positions), most likely correspond to the same modification, C-terminal addition of a glutamate residue. Interestingly, the third C-terminal mass shift is oxidation (+16 Da), which is observed to occur at high frequency (17–20% modified counts across replicates at C-terminal residues with at least one +16 Da modification, compared to 2.6–5% at all other modified positions) on C-terminal residues as well as N-terminal residues, although the C-terminal modification is observed for a smaller set of proteins.

The C-terminal glutamylation modification is especially interesting. The most frequent target for this modification is the C terminus of the 30S ribosomal protein S6 (RpsF), which is known to undergo post-translational modification with 1–4 glutamate residues (mass = 129 Da) [43]. The enzymatic addition of these Glu residues to S6 proceeds in a stepwise fashion, and any modification of two or more Glu residues would fall outside the range of mass shifts that were considered in our analysis, so it is likely that the mono-glutamylated S6 we observed only represents a subset of the total modified S6 present in our samples.

We also identified a previously unreported C-terminal +129 Da modification of the stationary phase ribosomal stability factor RaiA / YfiA [22]. YfiA binds within the mRNA tunnel of the 30S subunit [44, 45], where it inhibits translation [44, 46]

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and prevents subunit dissociation and 100S dimer formation for a subset of ribosomes in stationary phase [47]. YfiA and S6 lie near one another within the 30S subunit, and both proteins' C termini extend towards the same region of the 16S rRNA on the subunit surface (Figure S2), although the modified C-terminal tails themselves are not resolved in the crystal structure. The temporal modification patterns of S6 and YfiA differ dramatically (Figure S3). S6 levels of both total PSM counts and Ct +129-Da modified counts peak in mid-exponential phase, followed by a steep drop to a lower number of counts that is maintained through late stationary phase; the relative proportion of +129 Da modified counts remains nearly unchanged across all time points. In contrast, YfiA shows low or no counts of either modified or unmodified PSMs until the onset of stationary phase, when overall counts increase dramatically, accompanied by a low but constant level of C-terminal +129 Da modification through late stationary phase. The exponential phase enrichment we observed for the +129 Da mass shift is therefore due largely to changes in overall expression of its target proteins rather than differential modification.

Temporal patterns

The glucose starvation dataset used in our analysis is unique in the wide range of timepoints (3h–336h) that were sampled. Changes in abundance during different phases of the growth cycle in liquid culture have been observed for individual PTMs, but an unbiased examination of temporal variation in the global PTM profile has not been performed in *E. coli*. To identify mass shifts with significant frequency changes over the growth cycle, we first pooled four of our nine time-point samples into exponential-phase samples (3h, 4h, 5h, and 6h, EXP) and four into stationary-phase samples (24h, 48h, 168h, 336h, STA). (We did not include the 8h sample in this analysis.) We then grouped counts across modified amino-acid positions by mass shift–AA pairs and compared the ratio of modified:unmodified counts at all modified positions in the EXP and STA pools using Fisher's exact test (FET) [48]. Mass shift–AA pairs were called as significant if their FET p-values passed a false-discovery rate filter ($< 5\%$ FDR by the Benjamini-Hochberg step-down procedure) in all three biological replicates. Because we used a two-tailed test that was unable to determine the direction of enrichment (i.e., EXP $>$ STA or EXP $<$ STA), we subsequently divided significant mass shift-AA pairs into EXP $>$ STA or EXP $<$ STA groups using the FET log-odds score.

We identified only a single mass shift that consistently shows significantly higher levels of modification in exponential phase across all three biological replicates, a +16 Da modification of tryptophan (3.78–4.33% of total counts at modified positions across the three biological replicates have the mass shift in exponential phase, 1.60–1.69% in stationary phase, Table 6). The behavior of this mass shift differs slightly across the three biological replicates: in biological replicates 1 and 2, the +16 Da Trp modification shows a spike in abundance near the Exponential-Stationary phase transition (8h), followed by a drop to near zero by mid-stationary phase (48h), while replicate 3 shows a spike of enrichment earlier in exponential phase (4h) followed by a steep drop off at the 5h timepoint (Figure 6).

We identified five mass shifts that consistently show significantly higher levels of modification in stationary phase across all three biological replicates: a +1 Da

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

modification of asparagine (1.90–3.06% of total counts at modified positions have the mass shift in exponential phase, 2.95–4.60% in stationary phase); +42 Da modifications of serine, alanine, and threonine (29.78–31.71% EXP, 46.30–60.07% STA; 18.33–22.45% EXP, 34.81–46.46% STA; and 0.0–3.37% EXP, 9.46–15.70%, respectively), and a +48 Da modification of cysteine (0.94–1.09% EXP, 3.11–4.19% STA) (Table 7). As with the exponential-phase-biased mass shifts, we observed different temporal patterns when timepoints are considered individually (Figure 7). For example, the +1 Da asparagine modification and the +48 Da cysteine modification show steady increases across stationary phase, reaching their highest value at the latest stationary phase timepoint (336h), while the +42 Da modification to serine shows a more step-like increase in abundance near the onset of stationary phase, with abundance remaining fairly constant through the latest timepoints.

Preferential persistence of N-terminally acetylated proteins in stationary phase

The N-terminal bias and preference for serine, alanine, and threonine residues observed for the +42 Da mass shift strongly suggests that this modification corresponds to N-terminal Nα-acetylation. Although cotranslational N-terminal Nα 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 (encoded by the rpsE gene), S18 (encoded by the rpsR gene), and L12/7 (encoded by the rplL gene)[49]; elongation factor Tu (EFTu, encoded by the tufB gene) [50]; and the chaperone SecB [51]. In addition, a number of heterologous eukaryotic proteins are modified with an NtAc when overexpressed in *E. coli* [52, 53, 54, 55].

We identified more than 40 Nt-acetylated proteins, and were able to recover modified peptides from known Nt-acetylation target SecB (Figure S4) and a small number of peptides matching Nt-acetylated ribosomal protein S5 (Figure S6) in our initial MODa dataset. The low peptide counts for S5, as well as the absence of modified PSMs for the other known (and highly abundant) targets ribosomal proteins S18 and L7/12, as well as EFTu, are likely due to the presence of tryptic cleavage sites within a few residues of the N-terminus in all three of these proteins (Nt-AHIE**K**QAGE for S5, Nt-A**R**YF**RRRK**F for S18, Nt-SIT**K**DQIEE for L7/12, and Nt-S**K**E**K**FERT**K** for EFTu). This means that most copies of the protein present in our samples will produce N-terminal peptides too short to recover during subsequent liquid chromatography and MS/MS steps. Consistent with this interpretation, we were able to recover abundant peptides from non-N-terminal regions of all four of these proteins, and the small number of S5 N-terminal peptides that were recovered were all the result of missed cleavage events at the N-terminal-most cleavage site. Among the NtAC peptides that were recovered in our modA dataset, the Nt fragment from SecB is by far the most frequently observed, representing 15–41% of the total Nt-Acetylated peptides across the nine time points. In addition, six other proteins from our dataset were previously identified as Nt-acetylation targets in an enrichment-based analysis of N-terminal modifications in *Pseudomonas aeruginosa*[13] (See Table 8).

We observed that NtAc modified proteins are proportionally more heavily modified in stationary phase (Figure 7). This pattern could be explained by (i) an

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

increase in acetylation activity in stationary phase and/or (ii) a proportionally larger decrease in non-acetylated copies of a protein relative to acetylated copies in stationary phase. To differentiate between these scenarios, we plotted total PSM counts and NtAc-modified PSM counts for pooled NtAc-targeted proteins across all nine time points (Figure 8). When all NtAc-targeted proteins are considered (Figure 8, top left panel), the total number of PSMs stays appoximately constant, while the number of NtAc-modified PSMs increases by nearly twofold in early stationary phase, consistent with scenario (i). However, NtAc-targeted proteins pooled by penultimate amino acid (Figure 8) or individual NtAc-targeted proteins (Figures S4, S5, and S6) show a mixture of both scenarios. NtAc-targeted proteins with a penultimate serine or threonine residue, for example, exhibit a pattern consistent with scenario (i), similar to the pattern for all targets (Figure 8, top right and bottom left panels). Proteins with a penultimate alanine, however, show a slight increase in modified peptides at the onset of stationary phase, accompanied by a large drop in unmodified peptides (Figure 8, top right panel). Many of the the most heavily NtAc-modified proteins also show this pattern, such as LysS, SpeA, pdxH, and SecB (Figure S4), and IlvA and KdgR (Figure S4). This preferential retention of NtAc-modified peptides in stationary phase suggests that NtAc may play role in protein stability by acting as an anti-degradation signal (see Discussion). A table of all Nt-acetylation sites recovered by MODa is included in Additional File 12.

Asparagine deamidation is strongly enriched in very late stationary phase

An interesting temporal pattern was also identified for the +1 Da modification of Asparagine residues, which increases in frequency throughout stationary phase and peaks at the last timepoint (336h) (Figures 7 and S7). +1 Da modifications were the most frequently observed modifications in our dataset, and likely result from a variety of different modification chemistries. A +1 Da modification occurring on an asparagine residue, however, is known to be a signature of nonenzymatic asparagine deamidation, in which a backbone nitrogen initiates a nucleophilic attack on the amide carbon of the asparagine side chain (or the asparagine amide nitrogen on the backbone carbonyl carbon) to form a cyclic succinimide intermediate [56, 57, 58]. This intermediate can then resolve by hydrolysis to either convert the original asparagine to an aspartate residue, or rearrange to form an isopeptide linkage through isoaspartate; both of these events result in a +1 Da mass shift.

Although asparagine deamidation can occur spontaneously as an experimental artifact during preparation of proteomic samples [59], a number of lines of reasoning suggest that at least a subset of the modifications we observe were present in the samples prior to processing. First, we observe a nearly identical pattern of increasing Asp +1 Da modification across all three of our biological replicates (Figure 7). All timepoints were collected from a single set of cultures started on the same day, each biological replicate was grown independently (on a different day) from the others, and all timepoints from a single replicate were processed for proteomic analysis in parallel. The bulk of nonenzymatic deamidation during proteomic sample prep has been shown to occur during tryptic digest [59], with both longer incubation time and basic pH increasing the occurrence of deamidated peptides. Although no special steps were taken during preparation of the samples included in our analysis, both

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of these variables should be identical across all timepoints (because samples were processed together), and any nonenzymatic deamidation should therefore also be constant across timepoints. The pattern does not appear to be explained simply by increased expression of modified proteins during stationary phase, as the pattern is observed even for individual modifications that have high abundance in both exponential and stationary phases (Figure S7). A table of all putative asparagine deamidation sites recovered by MODa is included in Additional File 12.

Oxidative modifications of methionine and tryptophan are variable across biological replicates

Oxidation (+16 Da) modifications, particularly of methionine, are very common in our data, but with the exception of +16 Da modification of tryptophan residues (Figure 6), oxidations in general are not identified as having a significant bias for either exponential or stationary phase. Both Met +16 Da and Trp +16 Da show significant variability among the three biological replicates, with replicates one and two showing a similar pattern of relative modification enrichment over time, while replicate three has a different pattern (Figures S10 and S8). In addition, for both modifications replicates one and two show a peak of modified peptide counts centered at or near the 8 h time point (the exponential-stationary phase transition; this timepoint was excluded from our initial comparisons of stationary vs. exponential enrichment), with the proportion of modified PSMs then decreasing to early-exponential-phase levels or below by 24 h.

The reason for the discrepancy between the third replicate and the two others is unclear; this observation in combination with the common occurrence of oxidative modifications as experimental artifacts [60] makes it difficult to draw any biological conclusions from the temporal patterns of oxidative modifications. We observe the discrepancy among replicates only for oxidative modifications and not for other modified peptide counts or overall peptide levels, so one possibility is that a difference in redox conditions in sample processing influenced the number of oxidized peptides that were recovered. Differential modification in the third replicate is apparent in the temporal modification patterns of individual target sites (Figures S9 and S11), but does not display a consistent pattern across sites. Tables of all methionine and tryptophan oxidation sites are included in Additional File 12.

Discussion

We have leveraged a large proteomics dataset [39] and the fast multi-blind spectral alignment algorithm MODa [38] to construct a comprehensive, unbiased map of all protein post-translational modifications between −200 and +200 Da at 9 timepoints, spanning early exponential phase (3h post-innoculation) through late stationary / starvation phase (336h, or 2 weeks post-innoculation). From this map, we have identified post-translational mass shifts with statistically significant differences in modification stoichiometry between N- and C-terminal ends of proteins and between exponential and stationary phases. This analysis has enabled us to identify previously unobserved temporal patterns and novel target proteins for known modifications, and to identify possible novel modifications. Finally, by comparing temporal patterns of modified and unmodified PSM counts for individual AA positions,

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

we have been able to identify a possible relationship between post-translational modification and protein degradation rate in stationary phase.

Although decades of work have been dedicated to studying the biochemical and physiological function of post-translational modifications, much of this work has focused on a handful of modification chemistries such as Ser/Thr phosphorylation and Lys acetylation. Technical limitations in instrument sensitivity, sample preparation, and data analysis have meant that even these well-studied PTMs are often studied in isolation, and their place in the overall context of the cell, in terms of the overall set of pathways and proteins that utilize them, their interaction with other modifications, and their abundance relative to other modifications, is lost. An intriguing feature of our dataset is the relative scarcity of the most commonly studied regulatory modifications, such as phosphorylation and acetylation; in the few cases where such modifications are identified, they tend to occur at low frequency, even on very abundant proteins (Table 2).

By contrast, we have identified examples of abundant modification for N-terminal acetylation and C-terminal glutamylation. While further work is necessary to establish that the orders-of-magnitude differences in abundance between these modifications and more well-known regulatory modifications reflects their actual abundance in the cell, our findings do suggest that these modifications may play a more important physiological role than previously thought. Both of these modifications are known to be installed in a regulated and specific pattern on ribosomal proteins, but their function either in the ribosomal context or on other targets is largely unknown. In eukaryotic cells, N-terminal (Nt) acetylation has a variety of functions, including regulating protein stability, ER trafficking, protein complex formation, and membrane attachment [61], 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 [62], but no function for prokaryotic Nt acetylation outside of the ribosome has been proposed. While Nt acetylation of eukaryotic proteins can either inhibit [63] or enhance [64] degradation rates, our evidence suggests that Nt-acetylated proteins in *E. coli* are subject to lower levels of degradation than their unmodified counterparts. The viability of mutants in the three known *E. coli* Nt-acetyltransferase enzymes, RimI, RimJ, and RimL [49], should make experimental investigation of this hypothesis a tractable and interesting avenue for future research.

Similarly, the physiological role of C-terminal (Ct) glutamylation has only recently begun to be uncovered. Early investigations identified a Ct glutamyltransferase enzyme, RimK, that installs poly-E tails on ribosomal protein S6 *in vivo* and *in vitro* [65], but the only phenotypic effect observed in *E. coli rimK* mutant strains (other than loss of S6 glutamylation) is increased resistance to the aminoglycoside antibiotics streptomycin, neomycin, and kanamycin [66, 67]. Nonetheless, RimK, and presumably S6 Ct glutamylation, are conserved across a wide range of bacterial species [68], and recent work in *Pseudomonas* found profound changes in proteome composition and compromised colonization and virulence phenotypes in $\Delta rimK$ strains [69]. Our novel finding of an additional target of C-terminal glutamylation, the ribosomal hibernation factor YfiA, offers an additional experimental handle with which to examine the biological and molecular functions of this modification.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The association of both Ct-glutamylation target proteins with the ribosome is especially interesting, because some evidence suggests that RimK modifies S6 C-termini specifically on intact ribosomes [68, 70], and RimK is known to catalyze poly-L-glutamine formation in the absence of S6 [71]. The C-terminal amino-acid residues of YfiA resemble those of S6 only in the presence of two glutamate residues in the last two positions (DDAEAGDSEE for S6 and ANFVEEVEEE for YfiA), indicating that targeting may largely be a function of YfiA's structural association with the ribosome rather than due to a specific sequence signal.

The presence of a gradual increase in asparagine deamidation throughout our stationary phase samples is an intriguing observation. Asparagine deamidation/isomerization events occur spontaneously at a low frequency at specific protein residues with favorable local structure and sequence context [57, 72], and they are often observed in proteins that undergo a low frequency of turnover such as muscle fiber proteins [73] and lens crystallins [74]. This clock-like behavior of Asp deamidation is consistent with our observation of a steady accumulation of the Asp +1 Da mark through very late stationary phase (336h), and it suggests that proteins having this modification have been retained with little or no turnover throughout stationary phase. Remarkably, many of the most heavily modified target proteins are part of large supra-molecular complexes, including six on ribosomal proteins (N113 and N544 of ribosomal protein S1, encoded by the *rpsA* gene, N77 and N146 of ribosomal protein S5, encoded by the *rpsE* gene, and N89 of ribosomal protein L14, encoded by the *rplN* gene), N64 of EFTu, N77 of the genomic DNA structural protein H-NS [75], and two positions (N110 and N111) on SucB, the E2 subunit of the 2-oxoglutarate dehydrogenase multienzyme complex (OGDHC) [76]. Although retention of intact ribosomes through stationary phase is a well-documented phenomenon [22, 23], and H-NS has been shown to be involved in late-stationary-phase survival [77], retention of the OGDHC complexes has not been previously observed.

Our work has several limitations. First, although the consistent temporal signal across multiple replicates strongly indicates that the major modifications discussed above are of biological origin, we cannot rule out the possibility that a subset of these modifications are experimental artifacts; the oxidative modifications and asparagine deamidation in particular are known to occur as artifacts of downstream sample processing in MS/MS [78, 60], so further experimental verification will be needed to confirm their biological origin. Our study is also limited by the need to examine a relatively small window of mass shifts (−200 Da to +200 Da); many known modifications fall outside of this window, such as glycosylation, longer chain acylations, and lipidations [79]. Finally, while the lack of bias for particular modifications offers a number of advantages in our analysis strategy, it also means that our results are more limited by the inherent sensitivity of both shotgun MS/MS and computational identification of PTMs. Consequently, our data are biased towards highly abundant proteins and mass shifts, a factor that likely explains the scarcity of well-known PTMs such as lysine acetylation and phosphorylation in our data. Equipment improvements and/or novel experimental procedures (e.g. [80]) will likely be needed to enable detection of low-abundance or short-lived PTM and other rare effects such as translational mutations [81].

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Conclusions

In summary, the work presented here highlights the holistic perspective and novel biological insights that can be generated by combining unbiased PTM detection and deep temporal sampling of bacterial growth. Stationary phase biology and post-translational modification in prokaryotic systems are both still areas of active research with many open questions, and we hope that the analysis paradigm presented here can be applied to additional organisms and growth conditions to gain broader insight into prokaryotic physiology and evolution.

Methods

Origin of the analyzed data

All data were taken from a previously published *E. coli* time course [39]. In that study, *E. coli* was grown in glucose minimal media and samples were collected at 8 different time points: 3, 4, 5, 8, 24, 48, 168 and 336 hours past inoculation. The entire experiment was carried out in triplicate, with cultures in each time course grown at different times. Mass-spectrometry on these samples was carried out as follows [39]: Protein samples were prepared by trypsin digest and each sample was then analyzed using liquid chromatography mass spectrometry (LC/MS) on a LTQ-Orbitrap (Thermo Fisher). The resulting data are available from the ProteomeXchange Consortium (accession PXD002140) [82].

Post-translational modification identification and analysis

We analyzed the raw mass-spectrometry data via MODa [38]. MODa is a naive based spectral alignment algorithm that identifies peptides and their associated PTMs from the input mzXML spectral files. The program needs a few additional 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, number of modifications per peptide, and the mass-range to search for PTMs. We ran separate MODa searches for each of the 9 time points. Since there were 3 biological replicates, this resulted in a total of 27 MODa searches. We set the enzyme used in the searches to trypsin, with fully-tryptic and no-proline rules. We allowed for 2 missed cleavages. We used a mass-tolerance for the precursor ion of 10 ppm, and the mass-tolerance used for the product ion was set to 0.5 Da. Finally, we set carbamidomethylation (+57 Da) of cysteine as a static or fixed modification. As mentioned earlier, MODa requires a mass range to search for variable modifications, so we ran MODa searches for the mass range between -200 and $+200$ Da. We used the *E. coli* B REL606 genome sequence (GenBank:NC_012967.1 [83]) to create the reference proteome. For the independent analysis of N-terminal acetylation, a modified *E. coli* B REL606 database was constructed in which N-terminal methionines were removed for 359 proteins annotated in UniProt for cleavage of "initiator methionine". Raw spectral data were searched against this database by SEQUEST-HT (Proteome Discoverer 1.4; Thermo Scientific). Only fully-tryptic peptides were considered, allowing up to two missed cleavages. A precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.5 Da were used. Oxidized methionine and N-terminal acetylation were considered as dynamic modifications; carbamidomethylation of cysteine was selected as a

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

static modification. High confidence PSMs were filtered at a false discovery rate of ¡ 1% as calculated by Percolator (q-value ¡ 0.01, Proteome Discoverer 1.4; Thermo Scientific).

FDR calculations using MODa probabilities

For each PSM assigned to a spectra by MODa, the algorithm calculates a probability P_{MODa} using a logistic regression model that uses a variety of spectral features as parameters, trained on a standard set of correct and incorrect spectral matches [38]. To restrict our dataset to only high-quality PSMs, we used this probability to estimate the False-Discovery Rate (FDR) of incorrect matches in our dataset by (i) ranking all PSMs by their P_{MODa} values, and then (ii) iteratively adding PSMs, starting from the highest-probability matches, and calculating the FDR as

$$\text{FDR} = \frac{1}{k}\sum_{i=1}^{k}(1 - P_i),$$

where k is the rank index of the last added PSM and P_i is the P_{MODa} of the ith ranked PSM, until adding any additional PSMs would result in an FDR above the chosen cutoff value.

Metrics and statistical tests for single amino-acid bias, N-terminal/C-terminal bias, and growth-phase bias

To test the preference of each mass shift for modification of a single type of amino acid, we calculated a single-AA bias score $B_s(A)$ for mass shift s and length 20 vector A of counts of unique positions bearing at least one modification matching s for each amino acid type:

$$B_s(A) = \frac{U_m(A) + 1}{\bar{U}_{nm}(A) + 1}$$

where $U_m(A) = \max(U_{a \in A})$ and $\bar{U}_{nm}(A) = \frac{1}{19}\sum_{a \neq m} U_a$. Note that "unique position" means that a given position in a protein is counted at most once regardless of total PSM counts at that position; this choice was intended to reduce bias from modifications with high abundance at a small number of positions.

To simplify our analysis, we constructed an intermediate dataset of PSM counts calculated by amino-acid position across all proteins in the REL606 annotated proteome. Unmodified counts $n_{p,\text{unmod}}$ for each position p (having at least one modified or unmodified PSM) were calculated by summing PSM counts for any peptides that overlap p but do not have a modification (of any mass shift) at p. Modified counts $n_{p,s}$ were calculated by summing PSM counts for any peptides with a modification of mass shift s at protein position p.

To test for higher fractional modification by specific mass shifts at the protein termini, we constructed 2×2 contingency tables of the form

	$N_{\text{mod}}(s)$	$N_{\text{unmod}}(s)$
$N_{\text{Xt}}(s)$	$N_{\text{Xt, mod}}(s)$	$N_{\text{Xt, unmod}}(s)$
$N_{\text{non-Xt}}(s)$	$N_{\text{non-Xt, mod}}(s)$	$N_{\text{non-Xt, unmod}}(s)$

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

for each mass shift s in each of the three biological replicates, where $N_{\mathrm{Xt}}(s)$ is the sum $\sum_{p=\mathrm{Xt}}(n_{p,s}+n_{p,\mathrm{unmod}})$ for positions having at least one PSM with mass shift s occuring at the terminus Xt (either C- or N-terminus of a protein); $N_{\text{non-Xt}}(s)$ is the sum $\sum_{p\neq\mathrm{Xt}}(n_{p,s}+n_{p,\mathrm{unmod}})$ for positions p having at least one PSM with mass shift s, occuring at all other positions (including the opposite terminus); $N_{\mathrm{mod}}(s)$ is the sum $\sum n_{p,s}$ for positions having at least one PSM with mass shift s; and $N_{\mathrm{unmod}}(s)$ is the sum $\sum n_{p,\mathrm{unmod}}$ for positions having at least one PSM with mass shift s. We used these tables to perform Fisher's exact tests using a two-sided alternative hypothesis, implemented in Python using the statistics module of NumPy [84].

Similarly, to test for higher fractional modification by specific mass shift $\times$ amino acid pairs in either exponential or stationary phases of growth, we constructed 2×2 contingency tables of the form

	$N_{\mathrm{mod}}(s,a)$	$N_{\mathrm{unmod}}(s,a)$
$N_{\mathrm{STA}}(s,a)$	$N_{\mathrm{Sta,\ mod}}(s,a)$	$N_{\mathrm{STA,\ unmod}}(s,a)$
$N_{\mathrm{EXP}}(s,a)$	$N_{\mathrm{Exp,\ mod}}(s,a)$	$N_{\mathrm{EXP,\ unmod}}(s,a)$

for each mass-shift–amino-acid pair in the three biological replicates, where $N_{\mathrm{mod}}(s,a)$ and $N_{\mathrm{unmod}}(s,a)$ are as above and $N_{\mathrm{EXP}}(s,a)$ is the sum $\sum_{t=3,4,5,6}(n_{p,s,t}+n_{p,\mathrm{unmod},t})$, where $n_{p,s,t}$ is the count of PSMs modified by mass shift s at position p in timepoint t for positions p of amino acid type a having at least one PSM with mass shift s; and $N_{\mathrm{STA}}(s,a)$ is the sum $\sum_{t=24,48,168,336}(n_{p,s,t}+n_{p,\mathrm{unmod},t})$ for positions p of amino acid type a having at least one PSM with mass shift s. We used these tables to perform Fisher's exact tests using a two-sided alternative hypothesis, implemented in Python using the statistics module of NumPy [84].

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`. The analysis was performed in iPython [85] notebooks using the NumPy and SciPy libraries [84] for numerical calculations, the Pandas library [86] for data processing, and the MatPlotLib library [87] for plotting. Macromolecular structures in Figure S2 were assembled in MacPyMOL (version v1.7.4.4; Schrödinger, LLC).

Competing interests

The authors declare that they have no competing interests.

Author's contributions

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

Acknowledgements

This project was funded by Army Research Office Grant W911NF-12-1-0390, National Institutes of Health Grant R01 GM088344, Welch Foundation Grant F-1780, and CPRIT Grant RP110782. EMM acknowledges additional funding from the National Institutes of Health (DP1 OD009572), National Science Foundation (IOS1237975), and the Welch Foundation (F-1515).

We thank John Houser and Kevin Drew for helpful discussions. The Texas Advanced Computing Center (TACC) provided high-performance computing support.

Author details

[1] Institute for Cellular and Molecular Biology, The University of Texas at Austin, Speedway, Austin, USA. [2] Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Speedway, Austin, USA. [3] Center for Systems and Synthetic Biology, The University of Texas at Austin, Speedway, Austin, USA. [4] College of Pharmacy, The University of Texas at Austin, Speedway, Austin, USA. [5] Department of Molecular Biosciences, The University of Texas at Austin, Speedway, Austin, USA. [6] Department of Integrative Biology, The University of Texas at Austin, Speedway, Austin, USA.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. Khoury, G.A., Baliban, R.C., Floudas, C.A.: Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep **1** (2011). doi:10.1038/srep00090
2. Jenuwein, T., Allis, C.D.: Translating the histone code. Science **293**(5532), 1074–80 (2001). doi:10.1126/science.1063127
3. Pawson, T.: Protein modules and signalling networks. Nature **373**(6515), 573–80 (1995). doi:10.1038/373573a0
4. Lim, W.A.: The modular logic of signaling proteins: building allosteric switches from simple binding domains. Curr Opin Struct Biol **12**(1), 61–8 (2002)
5. Laub, M.T., Goulian, M.: Specificity in two-component signal transduction pathways. Annu Rev Genet **41**, 121–145 (2007). doi:10.1146/annurev.genet.41.042007.170548. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. laubmit.edu
6. Choudhary, C., Mann, M.: Decoding signalling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol **11**(6), 427–439 (2010). doi:10.1038/nrm2900
7. Jones, J.D., O'Connor, C.D.: Protein acetylation in prokaryotes. Proteomics **11**(15), 3012–22 (2011). doi:10.1002/pmic.201000812
8. Dworkin, J.: Ser/thr phosphorylation as a regulatory mechanism in bacteria. Curr Opin Microbiol **24**, 47–52 (2015). doi:10.1016/j.mib.2015.01.005
9. Kusebauch, U., Ortega, C., Ollodart, A., Rogers, R.S., Sherman, D.R., Moritz, R.L., Grundner, C.: Mycobacterium tuberculosis supports protein tyrosine phosphorylation. Proc Natl Acad Sci U S A **111**(25), 9265–70 (2014). doi:10.1073/pnas.1323894111
10. Hansen, A.-M.M., Chaerkady, R., Sharma, J., Díaz-Mejía, J.J., Tyagi, N., Renuse, S., Jacob, H.K.C., Pinto, S.M., Sahasrabuddhe, N.A., Kim, M.-S.S., Delanghe, B., Srinivasan, N., Emili, A., Kaper, J.B., Pandey, A.: The escherichia coli phosphotyrosine proteome relates to core pathways and virulence. PLoS Pathog **9**(6), 1003403 (2013). doi:10.1371/journal.ppat.1003403
11. Striebel, F., Imkamp, F., Özcelik, D., Weber-Ban, E.: Pupylation as a signal for proteasomal degradation in bacteria. Biochim Biophys Acta **1843**(1), 103–113 (2014). doi:10.1016/j.bbamcr.2013.03.022
12. Alber, T.: Signaling mechanisms of the mycobacterium tuberculosis receptor ser/thr protein kinases. Curr Opin Struct Biol **19**(6), 650–7 (2009). doi:10.1016/j.sbi.2009.10.017
13. Ouidir, T., Jarnier, F., Cosette, P., Jouenne, T., Hardouin, J.: Characterization of n-terminal protein modifications in pseudomonas aeruginosa pa14. J Proteomics (2014). doi:10.1016/j.jprot.2014.11.006
14. Hentchel, K.L., Escalante-Semerena, J.C.: Acylation of biomolecules in prokaryotes: a widespread strategy for the control of biological function and metabolic stress. Microbiol Mol Biol Rev **79**(3), 321–46 (2015). doi:10.1128/MMBR.00020-15
15. Starai, V.J., Escalante-Semerena, J.C.: Identification of the protein acetyltransferase (pat) enzyme that acetylates acetyl-coa synthetase in salmonella enterica. J Mol Biol **340**(5), 1005–1012 (2004). doi:10.1016/j.jmb.2004.05.010
16. Salomon, D., Orth, K.: What pathogens have taught us about posttranslational modifications. Cell Host Microbe **14**(3), 269–79 (2013). doi:10.1016/j.chom.2013.07.008
17. Gnad, F., Forner, F., Zielinska, D.F., Birney, E., Gunawardena, J., Mann, M.: Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes, and mitochondria. Mol Cell Proteomics **9**(12), 2642–2653 (2010). doi:10.1074/mcp.M110.001594
18. Navarro Llorens, J.M., Tormo, A., Martínez-García, E.: Stationary phase in gram-negative bacteria. FEMS Microbiol Rev **34**(4), 476–95 (2010). doi:10.1111/j.1574-6976.2010.00213.x
19. Finkel, S.E.: Long-term survival during stationary phase: evolution and the gasp phenotype. Nat Rev Microbiol **4**(2), 113–20 (2006). doi:10.1038/nrmicro1340
20. Nyström, T.: The glucose-starvation stimulon of escherichia coli: induced and repressed synthesis of enzymes of central metabolic pathways and role of acetyl phosphate in gene expression and starvation survival. Mol Microbiol **12**(5), 833–43 (1994)
21. Reeve, C.A., Amy, P.S., Matin, A.: Role of protein synthesis in the survival of carbon-starved escherichia coli k-12. J Bacteriol **160**(3), 1041–6 (1984)
22. Maki, Y., Yoshida, H., Wada, A.: Two proteins, yfia and yhbh, associated with resting ribosomes in stationary phase escherichia coli. Genes Cells **5**(12), 965–74 (2000)
23. Wada, A.: Growth phase coupled modulation of escherichia coli ribosomes. Genes Cells **3**(4), 203–8 (1998)
24. Nyström, T.: Translational fidelity, protein oxidation, and senescence: lessons from bacteria. Ageing Res Rev **1**(4), 1–11 (2002)
25. Petropoulos, I., Friguet, B.: Protein maintenance in aging and replicative senescence: a role for the peptide methionine sulfoxide reductases. Biochim Biophys Acta **1703**(2), 261–6 (2005). doi:10.1016/j.bbapap.2004.08.018
26. Weichart, D., Querfurth, N., Dreger, M., Hengge-Aronis, R.: Global role for clpp-containing proteases in stationary-phase adaptation of escherichia coli. J Bacteriol **185**(1), 115–25 (2003)
27. Na, S.H., Miyanaga, K., Unno, H., Tanji, Y.: The survival response of escherichia coli k12 in a natural environment. Appl Microbiol Biotechnol **72**(2), 386–92 (2006). doi:10.1007/s00253-005-0268-3
28. Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., Heinemann, M.: The quantitative and condition-dependent escherichia coli proteome. Nat Biotechnol **34**(1), 104–10 (2016). doi:10.1038/nbt.3418
29. Soufi, B., Krug, K., Harst, A., Macek, B.: Characterization of the e. coli proteome and its modifications during growth and ethanol stress. Front Microbiol **6**, 103 (2015). doi:10.3389/fmicb.2015.00103
30. Weinert, B.T., Iesmantavicius, V., Wagner, S.A., Schölz, C., Gummesson, B., Beli, P., Nyström, T., Choudhary, C.: Acetyl-phosphate is a critical determinant of lysine acetylation in e. coli. Mol Cell **51**(2), 265–272 (2013). doi:10.1016/j.molcel.2013.06.003. The NNF Center for Protein Research, Faculty of Health and Medical

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark.

31. Soares, N.C., Spät, P., Krug, K., Macek, B.: Global dynamics of the escherichia coli proteome and phosphoproteome during growth in minimal medium. J Proteome Res **12**(6), 2611–21 (2013). doi:10.1021/pr3011843
32. Kuhn, M.L., Zemaitaitis, B., Hu, L.I., Sahu, A., Sorensen, D., Minasov, G., Lima, B.P., Scholle, M., Mrksich, M., Anderson, W.F., Gibson, B.W., Schilling, B., Wolfe, A.J.: Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. PLoS ONE **9**(4), 94816 (2014). doi:10.1371/journal.pone.0094816
33. Macek, B., Mann, M., Olsen, J.V.: Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol **49**, 199–221 (2009). doi:10.1146/annurev.pharmtox.011008.145606
34. Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S.: Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20**(18), 3551–67 (1999). doi:10.1002/(SICI)1522-2683(19991201)20:18¡3551::AID-ELPS3551¿3.0.CO;2-2
35. Eng, J.K., McCormack, A.L., Yates, J.R.: 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). doi:10.1016/1044-0305(94)80016-2
36. Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., Bryant, S.H.: Open mass spectrometry search algorithm. J Proteome Res **3**(5), 958–64 (2004). doi:10.1021/pr0499491
37. Craig, R., Beavis, R.C.: Tandem: matching proteins with tandem mass spectra. Bioinformatics **20**(9), 1466–7 (2004). doi:10.1093/bioinformatics/bth092
38. Na, S., Bandeira, N., Paek, E.: Fast multi-blind modification search through tandem mass spectrometry. Mol Cell Proteomics **11**(4), 111–010199 (2012). doi:10.1074/mcp.M111.010199
39. Houser, J.R., Barnhart, C., Boutz, D.R., Carroll, S.M., Dasgupta, A., Michener, J.K., Needham, B.D., Papoulas, O., Sridhara, V., Sydykova, D.K., Marx, C.J., Trent, M.S., Barrick, J.E., Marcotte, E.M., Wilke, C.O.: Controlled measurement and comparative analysis of cellular components in e. coli reveals broad regulatory changes in response to glucose starvation. PLoS Comput Biol **11**(8), 1004400 (2015). doi:10.1371/journal.pcbi.1004400
40. Wiśniewski, J.R., Rakus, D.: Quantitative analysis of the escherichia coli proteome. Data Brief **1**, 7–11 (2014). doi:10.1016/j.dib.2014.08.004
41. Zhang, K., Zheng, S., Yang, J.S., Chen, Y., Cheng, Z.: Comprehensive profiling of protein lysine acetylation in escherichia coli. J Proteome Res **12**(2), 844–851 (2013). doi:10.1021/pr300912q
42. Scherl, A., Shaffer, S.A., Taylor, G.K., Hernandez, P., Appel, R.D., Binz, P.-A.A., Goodlett, D.R.: On the benefits of acquiring peptide fragment ions at high measured mass accuracy. J Am Soc Mass Spectrom **19**(6), 891–901 (2008). doi:10.1016/j.jasms.2008.02.005
43. Reeh, S., Pedersen, S.: Post-translational modification of escherichia coli ribosomal protein s6. Mol Gen Genet **173**(2), 183–7 (1979)
44. Vila-Sanjurjo, A., Schuwirth, B.-S.S., Hau, C.W., Cate, J.H.D.: Structural basis for the control of translation initiation during stress. Nat Struct Mol Biol **11**(11), 1054–9 (2004). doi:10.1038/nsmb850
45. Polikanov, Y.S., Blaha, G.M., Steitz, T.A.: How hibernation factors rmf, hpf, and yfia turn off protein synthesis. Science **336**(6083), 915–8 (2012). doi:10.1126/science.1218538
46. Agafonov, D.E., Kolb, V.A., Spirin, A.S.: Ribosome-associated protein that inhibits translation at the aminoacyl-trna binding stage. EMBO Rep **2**(5), 399–402 (2001). doi:10.1093/embo-reports/kve091
47. Ueta, M., Yoshida, H., Wada, C., Baba, T., Mori, H., Wada, A.: Ribosome binding proteins yhbh and yfia have opposite functions during 100s formation in the stationary phase of escherichia coli. Genes Cells **10**(12), 1103–12 (2005). doi:10.1111/j.1365-2443.2005.00903.x
48. Fisher, R.A.: On the interpretation of χ 2 from contingency tables, and the calculation of p. Journal of the Royal Statistical Society **85**(1), 87–94 (1922)
49. Nesterchuk, M.V., Sergiev, P.V., Dontsova, O.A.: Posttranslational modifications of ribosomal proteins in escherichia coli. Acta Naturae **3**(2), 22–33 (2011)
50. Arai, K., Clark, B.F., Duffy, L., Jones, M.D., Kaziro, Y., Laursen, R.A., L'Italien, J., Miller, D.L., Nagarkatti, S., Nakamura, S., Nielsen, K.M., Petersen, T.E., Takahashi, K., Wade, M.: Primary structure of elongation factor tu from escherichia coli. Proc Natl Acad Sci U S A **77**(3), 1326–30 (1980)
51. Smith, V.F., Schwartz, B.L., Randall, L.L., Smith, R.D.: Electrospray mass spectrometric investigation of the chaperone secb. Protein Sci **5**(3), 488–494 (1996). doi:10.1002/pro.5560050310
52. Bernal-Perez, L.F., Sahyouni, F., Prokai, L., Ryu, Y.: Rimj-mediated context-dependent n-terminal acetylation of the recombinant z-domain protein in escherichia coli. Mol Biosyst **8**(4), 1128–30 (2012). doi:10.1039/c2mb05499j
53. Miao, L., Fang, H., Li, Y., Chen, H.: Studies of the in vitro nalpha-acetyltransferase activities of e. coli riml protein. Biochem Biophys Res Commun **357**(3), 641–647 (2007). doi:10.1016/j.bbrc.2007.03.171
54. Wu, J., Chang, S., Gong, X., Liu, D., Ma, Q.: Identification of n-terminal acetylation of recombinant human prothymosin alpha in escherichia coli. Biochim Biophys Acta **1760**(8), 1241–1247 (2006). doi:10.1016/j.bbagen.2006.04.001
55. Charbaut, E., Redeker, V., Rossier, J., Sobel, A.: N-terminal acetylation of ectopic recombinant proteins in escherichia coli. FEBS Lett **529**(2-3), 341–5 (2002)
56. Lindner, H., Helliger, W.: Age-dependent deamidation of asparagine residues in proteins. Exp Gerontol **36**(9), 1551–63 (2001)
57. Robinson, N.E., Robinson, A.B.: Molecular clocks. Proc Natl Acad Sci U S A **98**(3), 944–9 (2001). doi:10.1073/pnas.98.3.944
58. Stephenson, R.C., Clarke, S.: Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J Biol Chem **264**(11), 6164–70 (1989)
59. Hao, P., Ren, Y., Alpert, A.J., Sze, S.K.: Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. Molecular & Cellular Proteomics **10**(10), 111–009381 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

doi:10.1074/mcp.O111.009381. The School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

60. Ghesquière, B., Gevaert, K.: Proteomics methods to study methionine oxidation. Mass Spectrom Rev **33**(2), 147–56 (2014). doi:10.1002/mas.21386
61. Starheim, K.K., Gevaert, K., Arnesen, T.: Protein n-terminal acetyltransferases: when the start matters. Trends Biochem Sci **37**(4), 152–161 (2012). doi:10.1016/j.tibs.2012.02.003
62. Clatterbuck Soper, S.F., Dator, R.P., Limbach, P.A., Woodson, S.A.: In vivo x-ray footprinting of pre-30s ribosomes reveals chaperone-dependent remodeling of late assembly intermediates. Mol Cell **52**(4), 506–16 (2013). doi:10.1016/j.molcel.2013.09.020
63. Martinez, A., Traverso, J.A., Valot, B., Ferro, M., Espagne, C., Ephritikhine, G., Zivy, M., Giglione, C., Meinnel, T.: Extent of n-terminal modifications in cytosolic proteins from eukaryotes. Proteomics **8**(14), 2809–31 (2008). doi:10.1002/pmic.200701191
64. Hwang, C.-S.S., Shemorry, A., Varshavsky, A.: N-terminal acetylation of cellular proteins creates specific degradation signals. Science **327**(5968), 973–7 (2010). doi:10.1126/science.1183147
65. Kang, W.K., Icho, T., Isono, S., Kitakawa, M., Isono, K.: Characterization of the gene rimk responsible for the addition of glutamic acid residues to the c-terminus of ribosomal protein s6 in escherichia coli k12. Mol Gen Genet **217**(2-3), 281–8 (1989)
66. Kade, B., Dabbs, E.R., Wittmann-Liebold, B.: Protein-chemical studies on escherichia coli mutants with altered ribosomal proteins s6 and s7. FEBS Letters **121**(2), 313–316 (1980)
67. Brown, M.E., Apirion, D.: Mapping a cluster of ribosomal genes in escherichia coli. Mol Gen Genet **133**(4), 317–27 (1974)
68. Koonin, E.V., Bork, P., Sander, C.: A novel rna-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res **22**(11), 2166–7 (1994)
69. Little, R.H., Grenga, L., Saalbach, G., Howat, A.M., Pfeilmeier, S., Trampari, E., Malone, J.G.: Adaptive remodeling of the bacterial proteome by specific ribosomal modification regulates pseudomonas infection and niche colonisation. PLoS Genet **12**(2), 1005837 (2016). doi:10.1371/journal.pgen.1005837
70. Kitakawa, M., Blumenthal, L., Isono, K.: Isolation and characterization of specialized transducing lambda phages carrying ribosomal protein genes of escherichia coli. Mol Gen Genet **180**(2), 343–9 (1980)
71. Kino, K., Arai, T., Arimura, Y.: Poly-alpha-glutamic acid synthesis using a novel catalytic activity of rimk from escherichia coli k-12. Appl Environ Microbiol **77**(6), 2019–25 (2011). doi:10.1128/AEM.02043-10
72. Robinson, N.E., Robinson, Z.W., Robinson, B.R., Robinson, A.L., Robinson, J.A., Robinson, M.L., Robinson, A.B.: Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. J Pept Res **63**(5), 426–36 (2004). doi:10.1111/j.1399-3011.2004.00151.x
73. Rivers, J., McDonald, L., Edwards, I.J., Beynon, R.J.: Asparagine deamidation and the role of higher order protein structure. J Proteome Res **7**(3), 921–7 (2008). doi:10.1021/pr070425l
74. Hains, P.G., Truscott, R.J.W.: Age-dependent deamidation of lifelong proteins in the human lens. Invest Ophthalmol Vis Sci **51**(6), 3107–14 (2010). doi:10.1167/iovs.09-4308
75. Rimsky, S., Travers, A.: Pervasive regulation of nucleoid structure and function by nucleoid-associated proteins. Curr Opin Microbiol **14**(2), 136–41 (2011). doi:10.1016/j.mib.2011.01.003
76. Murphy, G.E., Jensen, G.J.: Electron cryotomography of the e. coli pyruvate and 2-oxoglutarate dehydrogenase complexes. Structure **13**(12), 1765–73 (2005). doi:10.1016/j.str.2005.08.016
77. Chib, S., Mahadevan, S.: Involvement of the global regulator h-ns in the survival of escherichia coli in stationary phase. J Bacteriol **194**(19), 5285–93 (2012). doi:10.1128/JB.00840-12
78. Yang, H., Zubarev, R.A.: Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. Electrophoresis **31**(11), 1764–72 (2010). doi:10.1002/elps.201000027
79. Walsh, C.T., Garneau-Tsodikova, S., Gatto, G.J.: Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl **44**(45), 7342–7372 (2005). doi:10.1002/anie.200501023
80. Baeza, J., Dowell, J.A., Smallegan, M.J., Fan, J., Amador-Noguez, D., Khan, Z., Denu, J.M.: Stoichiometry of site-specific lysine acetylation in an entire proteome. J Biol Chem **289**(31), 21326–21338 (2014). doi:10.1074/jbc.M114.581843
81. Ribas de Pouplana, L., Santos, M.A.S., Zhu, J.-H.H., Farabaugh, P.J., Javid, B.: Protein mistranslation: friend or foe? Trends Biochem Sci **39**(8), 355–62 (2014). doi:10.1016/j.tibs.2014.06.002
82. Vizcaíno, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.-A.A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R.J., Kraus, H.-J.J., Albar, J.P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G.S., Martens, L., Jones, A.R., Hermjakob, H.: Proteomexchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol **32**(3), 223–6 (2014). doi:10.1038/nbt.2839
83. Jeong, H., Barbe, V., Vallenet, D., Choi, S.H., Lee, C.H., Lee, S.W., Vacherie, B., Yoon, S.H., Yu, D.S., Cattolico, L., Hur, C.G., Park, H.S., Segurens, B., Blot, M., Schneider, D., Studier, F.W., Oh, T.K., Lenski, R.E., Daegelen, P., Kim, J.F.: Complete genome sequence of escherichia coli (b) rel606. Unpublished (2009)
84. Van Der Walt, S., Colbert, S.C., Varoquaux, G.: The numpy array: a structure for efficient numerical computation. Computing in Science & Engineering **13**(2), 22–30 (2011)
85. Pérez, F., Granger, B.E.: Ipython: a system for interactive scientific computing. Computing in Science & Engineering **9**(3), 21–29 (2007)
86. McKinney, W.: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, vol. 445, pp. 51–56 (2010)
87. Hunter, J.D.: Matplotlib: A 2d graphics environment. Computing in Science & Engineering **9**(3), 90–95 (2007). doi:10.1109/MCSE.2007.55
88. Terhorst, C., Möller, W., Laursen, R., Wittmann-Liebold, B.: The primary structure of an acidic protein from 50-s ribosomes of escherichia coli which is involved in gtp hydrolysis dependent on elongation factors g and t. Eur J Biochem **34**(1), 138–52 (1973)
89. Young, C.C., Bernlohr, R.W.: Elongation factor tu is methylated in response to nutrient deprivation in

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

escherichia coli. J Bacteriol **173**(10), 3096–100 (1991)

90. Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y., Zhao, Y.: Identification of lysine succinylation as a new post-translational modification. Nat Chem Biol **7**(1), 58–63 (2011). doi:10.1038/nchembio.495
91. Dognin, M.J., Wittmann-Liebold, B.: Purification and primary structure determination of the n-terminal blocked protein, l11, from escherichia coli ribosomes. Eur J Biochem **112**(1), 131–51 (1980)

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

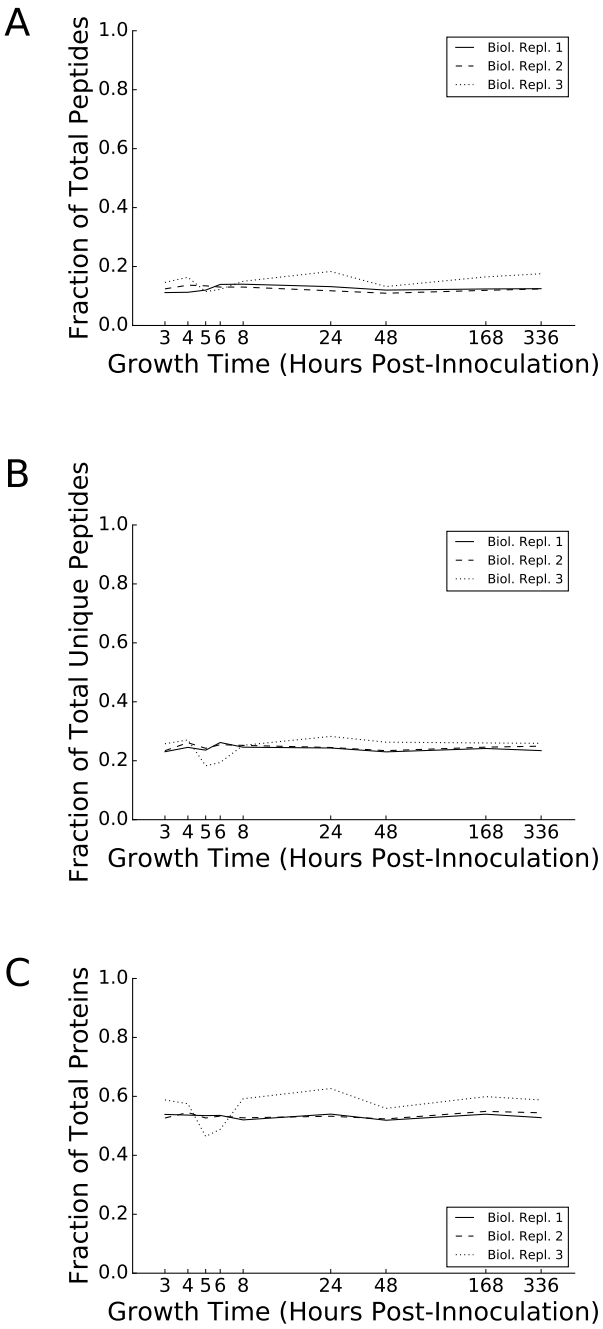

Figure 1 Global abundance of all modifications across growth timepoints.
Fraction of total counts of PSMs (A), unique peptides (B), and proteins (C) containing at least one mass shift passing the 1% FDR threshold at the timepoint indicated on the x-axis for biological replicates 1, 2 and 3 (solid, dashed, and dotted lines, respectively)

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

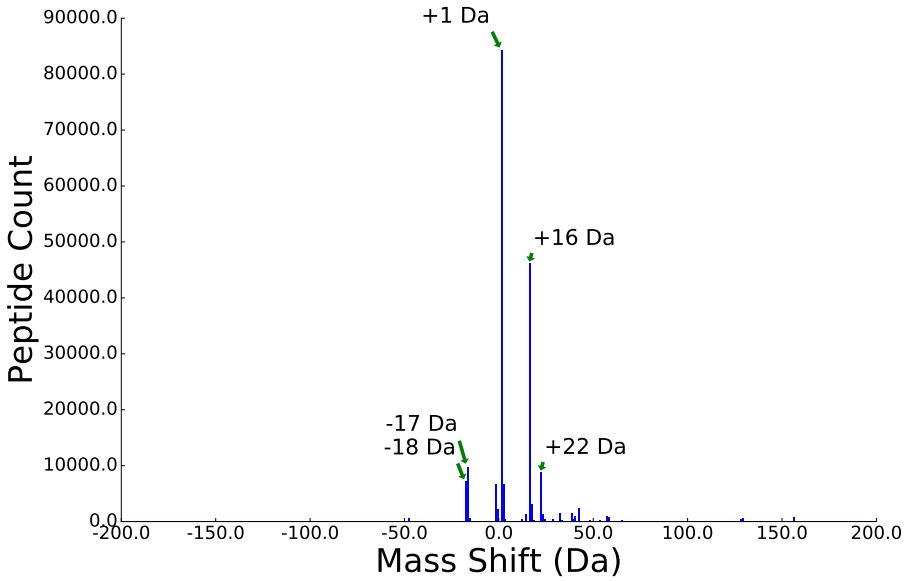

Figure 2 Mass-shift counts across all timepoints and datasets.
PSM counts for all mass shifts identified by MODa[38] between -200 Da and $+200$ Da, summed across all nine timepoints and all three biological replicates; labeled peaks are the top 5 most abundant (by raw count) mass shifts in the dataset.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

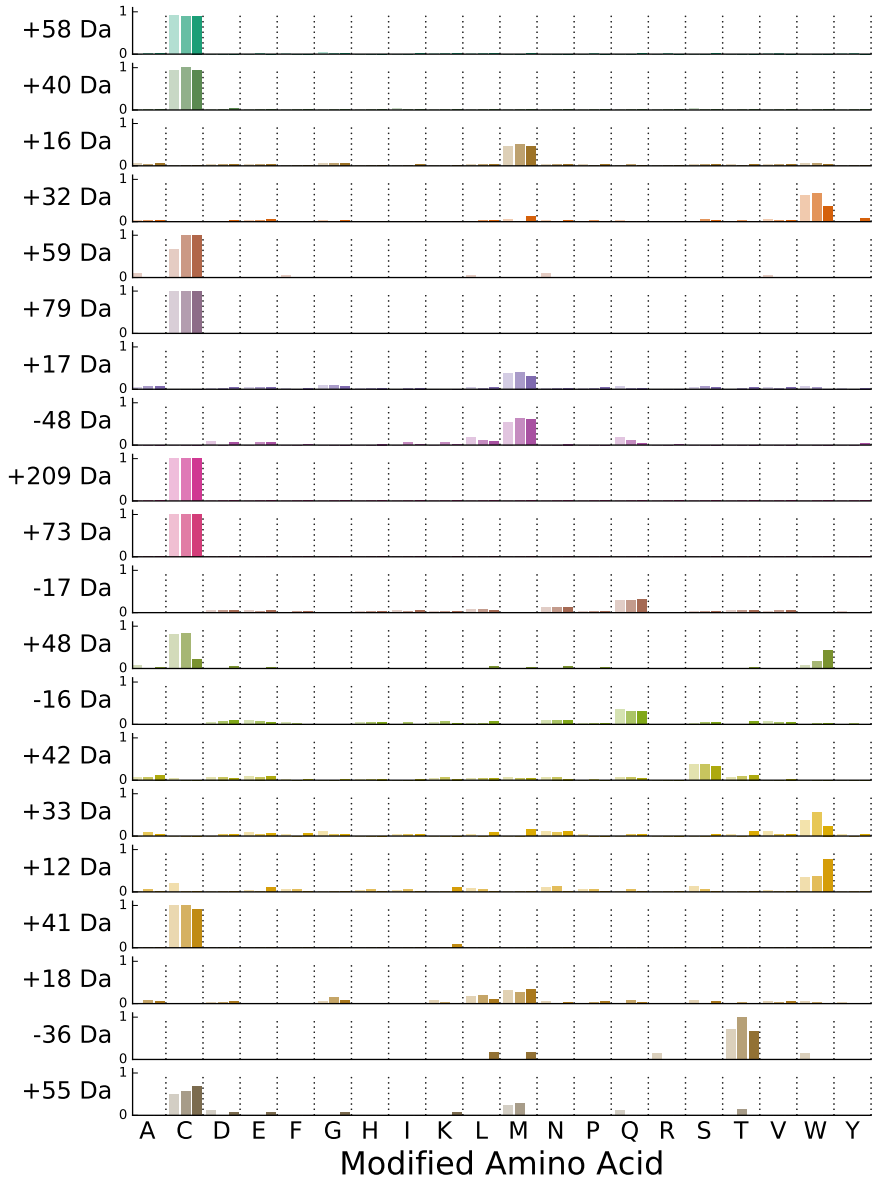

Figure 3 Distribution of selected mass shifts across amino acids.
Height of bars within each row represents the fraction of total AA positions for the mass shift (indicated on the y axis) that were identified on each amino acid residue type (columns). Individual bars within each column represent fractions for each biological replicate (replicates 1,2, and 3 from left to right within each column). Mass shifts are ordered by the single-AA bias score (the ratio of counts for the most commonly modified AA type to the mean of the counts for all other types; see Methods) with the highest score (most biased for a single AA) at the top; only the top 20 mass shifts are shown. Note that a constant mass shift of +57 Da was added to all cysteine modifications to correct for the presence of carbamidomethylation, meaning that a small number of cysteine modifications (e.g. +209 Da) fall outside of the mass range scanned by MODa (+/- 200 Da).

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

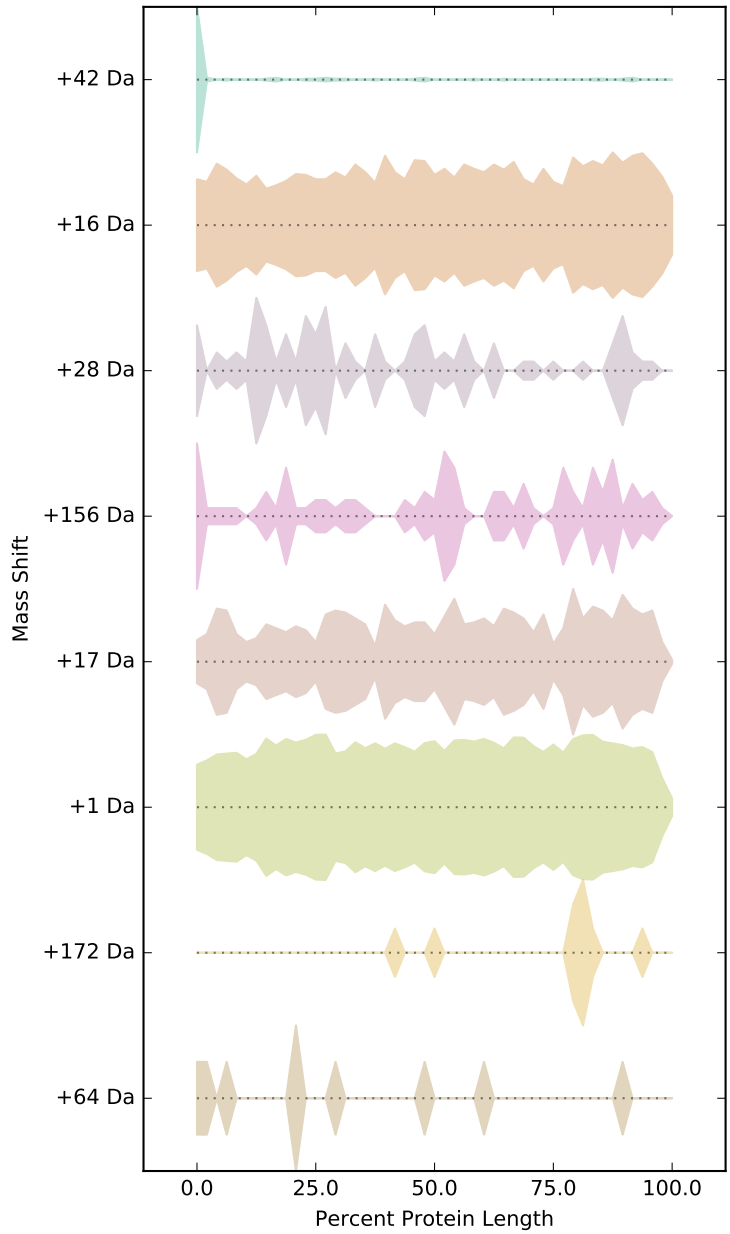

Figure 4 Distribution of Nt-biased mass shifts across positions in protein sequence. Width of traces within each row represent the density of unique positions identified for the mass shift indicated to the left along target proteins, normalized by protein length (x-axis). Traces are plotted symmetrically about the x-axis. Mass shifts are ranked from top to bottom by combined p-value from the Fisher's exact test for N-terminal modification enrichment across all three replicates (see section "N-terminal and C-terminal Modifications" and Table 4), with mass shifts having the strongest N-terminal enrichment at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

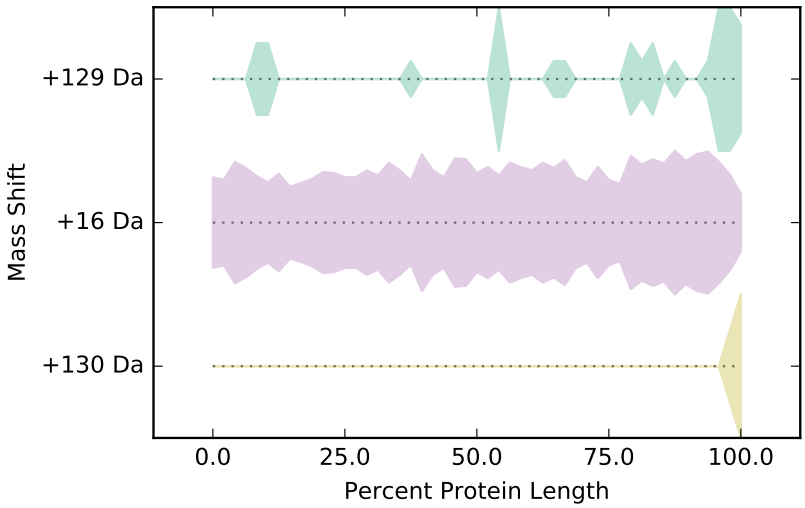

Figure 5 Distribution of Ct-biased mass shifts across positions in protein sequence. Width of traces within each row represent the density of unique modified positions (i.e. positions with more than one modified PSM; each position is counted once per protein) identified for each mass shift (indicated to the left) along target proteins, normalized by protein length (x-axis). Traces are plotted symmetrically about the x-axis. Mass shifts are ranked from top to bottom by combined p-value from the Fisher's exact test for N-terminal modification enrichment across all three replicates (see section "N-terminal and C-terminal Modifications" and Table 5), with mass shifts having the strongest C-terminal enrichment at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

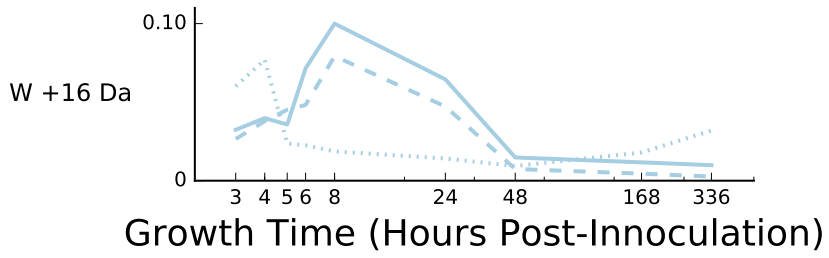

Figure 6 Abundance across all growth timepoints of tryptophan monooxidation, the sole mass shift with stronger modification in exponential phase.
Plot shows the fractional modification $N_{\text{mod}}/(N_{\text{mod}} + N_{\text{unmod}})$ across all nine time points, for positions having at least one W +16 Da modification at any time point. Individual traces show results for individual biological replicates 1 (solid lines), 2 (dashed lines), and 3 (dotted lines).

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

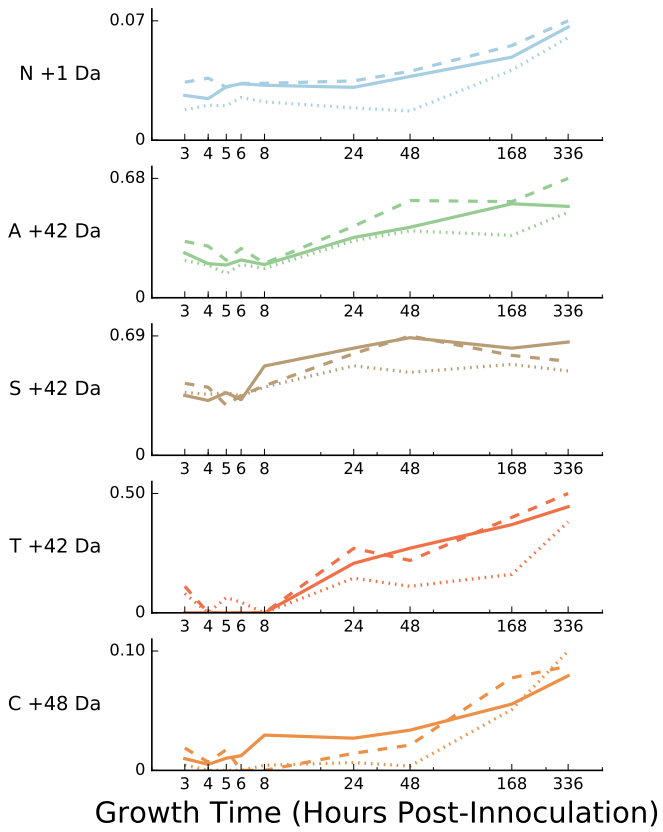

Figure 7 Abundance across all growth timepoints of mass shifts with stronger modification in stationary phase.

Each plot shows the fractional modification $N_{\text{mod}}/(N_{\text{mod}} + N_{\text{unmod}})$ across all nine time points, for positions having at least one modification of the indicated type at any time point. Individual traces within each plot show results for individual biological replicates 1 (solid lines), 2 (dashed lines), and 3 (dotted lines). Mass shift are ranked from top to bottom by p-value from the Fisher's exact test for modification enrichment in exponential phase (STA $>$ EXP; see Section "Temporal patterns" and Table 7), averaged across all three replicates, with the most stationary-phase-enriched (lowest p-values) at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

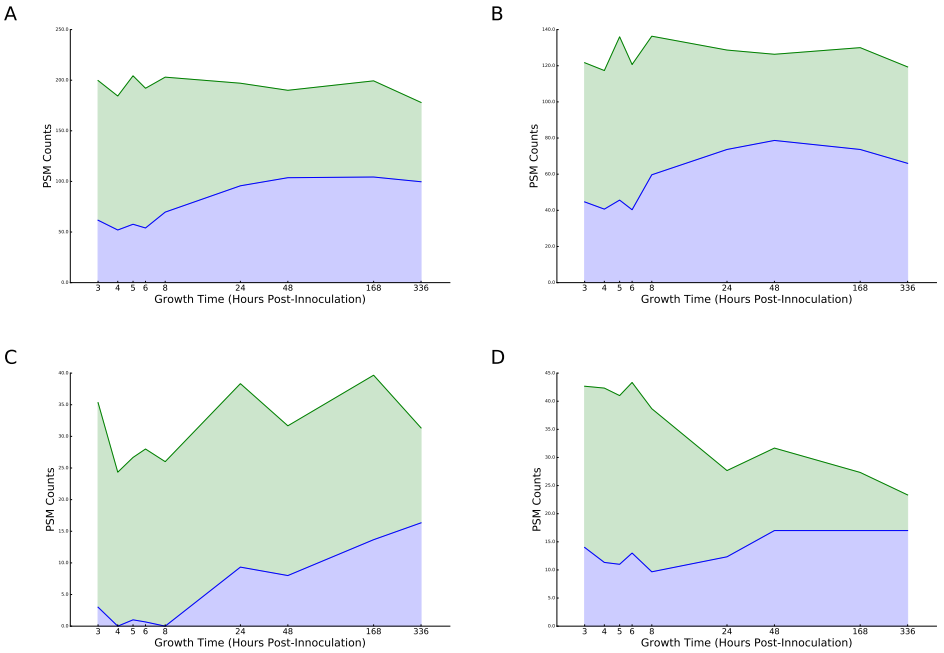

Figure 8 N-terminal +42 Da modified proteins are preferentially retained in stationary phase. Plots show unmodified (green) and +42 Da modified (blue) PSM counts for all N-terminal positions possessing at least one +42 Da modification at any time point, averaged across the three biological replicates. Shown are total counts (A), counts for peptides with a penultimate (i.e. following a cleaved N-terminal methionine) serine residue (B), a penultimate threonine residue (C), and a penultimate alanine residue (D).

Tables

Table 1 Counts of PSMs, Unique Peptides, and Proteins for Unfiltered, 1% FDR, and 5% FDR Datasets

		Total PSMs	Error Sum	FDR	Unique Peptides	Proteins
Unfiltered		2,527,135	44,4225.241	0.176	32,755	3,544
	Modified	608,357	19,0915.030	0.314	25,362	3,478
	Unmodified	1,918,778	253,310.212	0.132	7,393	66
1% FDR		1,473,636	14,736.377	0.010	19,265	2,121
	Modified	198,277	3,224.088	0.016	8,369	1,690
	Unmodified	1,275,359	11,512.288	0.009	10,896	431
5% FDR		1,980,884	99,044.212	0.050	22,776	2,445
	Modified	362,291	31,912.966	0.088	13,299	2,188
	Unmodified	1,618,593	67,131.246	0.041	9,477	257

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2 Previously Identified Post-Translational modifications recovered in our analysis

Locus	Position	AA	Mass Shift	PTM	Biol. Repl. 1	Biol. Repl. 2	Biol. Repl. 3	Ref.
rplL	82	K	+14 Da	Monomethylation	38 / 56 (71%)	13 / 18 (72%)	71 / 87 (81%)	[88]
tufB	57	K	+14 Da	Monomethylation	0 / 3 (0%)	0 / 6 (0%)	97 / 192 (50%)	[50, 89]
			+28 Da	Dimethylation	0 / 3 (0%)	0 / 6 (0%)	41 / 95 (43%)	
pgm	146	S	+81 Da	Phosphorylation	1/25 (4%)	0 / 11 (0%)	0 / 34 (0%)	UniProt version 2015-08 released on 2015-07-22 (UniProt consortium)
gapA	124	K	+42 Da	Acetylation	0 / 39 (0%)	1 / 71 (1.4%)	0 / 239 (0%)	[90]
	213	K	+42 Da	Acetylation	0 / 0 (0%)	0 / 3 (0%)	7 / 25 (28%)	[90]
icdA	242	K	+42 Da	Acetylation	1/198 (0.5%)	2 / 164 (1.2%)	0 / 184 (0%)	[90]
glyA	346	K	+42 Da	Acetylation	0 / 0 (0%)	1 / 1 (100%)	0 / 17 (0%)	[90]
fbaA	326	K	+42 Da	Acetylation	0 / 9 (0%)	1 / 17 (5.9%)	0 / 40 (0%)	[90]
rplK	40	K	+42 Da	Trimethylation	1 / 16 (6.3%)	0 / 1 (0%)	0 / 16 (0%)	[91]
rpsF	131	E	+129	Glutamylation	122 / 219 (55.7%)	152 / 415 (37%)	64 / 169 (38%)	[43]
secB	2	S	+42	Acetylation	184 / 221 (83%)	169 / 215 (78%)	115 / 135 (85%)	[51]
rpsE	2	A	+42	Acetylation	0 / 0 (0%)	0 / 0 (0%)	5 / 5 (100%)	[49]

Table 3 Most Commonly Modified AA Residues

Residue	Modified PSMs	Percentage
M	42457	21.41
Q	13990	7.06
A	13682	6.90
N	13500	6.81
L	12581	6.35
V	11309	5.70
E	10872	5.48
G	10623	5.36
S	9528	4.81
T	8396	4.23
I	8267	4.17
P	8139	4.10
D	7859	3.96
K	5325	2.69
W	4964	2.50
Y	4274	2.16
C	4211	2.12
H	4084	2.06
F	3261	1.64
R	955	0.48

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 4 Mass Shifts occuring more frequently on N-terminal ends of proteins

Protein-N-terminal-Biased Mass Shifts				
Mass Shift	**Biol. Replicate**	**N-terminal**	**Non-N-Terminal**	**p-value**
+42 Da	1	759 / 1869 (40.61%)	132 / 7860 (1.68%)	0.0
	2	657 / 1561 (42.09%)	111 / 6814 (1.63%)	0.0
	3	682 / 1930 (35.34%)	147 / 7681 (1.91%)	0.0
+16 Da	1	1187 / 8016 (14.81%)	13633 / 510869 (2.67%)	0.0
	2	972 / 6971 (13.94%)	9350 / 385682 (2.42%)	0.0
	3	531 / 5216 (10.18%)	20571 / 423856 (4.85%)	7.36×10^{-55}
+28 Da	1	16 / 128 (12.50%)	54 / 31261 (0.17%)	4.56×10^{-24}
	2	27 / 136 (19.85%)	11 / 28221 (0.04%)	1.78×10^{-55}
	3	45 / 140 (32.14%)	305 / 33420 (0.91%)	1.65×10^{-54}
+156 Da	1	20 / 1051 (1.90%)	277 / 54832 (0.51%)	1.17×10^{-6}
	2	14 / 832 (1.68%)	154 / 58376 (0.26%)	1.34×10^{-7}
	3	26 / 674 (3.86%)	313 / 43231 (0.72%)	2.73×10^{-11}
+17 Da	1	60 / 3107 (1.93%)	585 / 178417 (0.33%)	8.94×10^{-26}
	2	45 / 2730 (1.65%)	512 / 116280 (0.44%)	6.38×10^{-13}
	3	54 / 2070 (2.61%)	1960 / 147203 (1.33%)	7.64×10^{-6}
+1 Da	1	138 / 6868 (2.01%)	27945 / 2042671 (1.37%)	1.88×10^{-5}
	2	151 / 5934 (2.54%)	25111 / 1455121 (1.73%)	5.28×10^{-6}
	3	204 / 6226 (3.28%)	30808 / 1810086 (1.70%)	1.52×10^{-17}
+172 Da	1	15 / 15 (100.00%)	0 / 9431 (0.00%)	3.11×10^{-48}
	2	1 / 1 (100.00%)	1 / 11114 (0.01%)	1.8×10^{-4}
	3	14 / 14 (100.00%)	11 / 7365 (0.15%)	2.77×10^{-37}
+64 Da	1	13 / 58 (22.41%)	13 / 1620 (0.80%)	1.84×10^{-13}
	2	4 / 48 (8.33%)	6 / 1340 (0.45%)	2.27×10^{-4}
	3	13 / 103 (12.62%)	19 / 2119 (0.90%)	3.58×10^{-10}

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 5 Mass Shifts occuring more frequently on C-terminal ends of proteins

C-terminal-Biased Mass Shifts				
Mass Shift	**Biol. Replicate**	**C-terminal**	**Non-C-terminal**	**p-value**
+129 Da	1	165 / 703 (23.47%)	78 / 10189 (0.77%)	2.00×10^{-142}
	2	212 / 886 (23.93%)	49 / 8541 (0.57%)	1.98×10^{-177}
	3	82 / 502 (16.33%)	59 / 8478 (0.70%)	4.36×10^{-67}
+16 Da	1	39 / 205 (19.02%)	14781 / 518680 (2.85%)	7.88×10^{-21}
	2	42 / 245 (17.14%)	10280 / 392408 (2.62%)	7.76×10^{-22}
	3	22 / 103 (21.36%)	21080 / 428969 (4.91%)	5.13×10^{-09}
+130 Da	1	56 / 427 (13.11%)	2 / 366 (0.55%)	1.18×10^{-14}
	2	75 / 542 (13.84%)	0 / 549 (0.00%)	1.02×10^{-24}
	3	18 / 276 (6.52%)	0 / 196 (0.00%)	8.29×10^{-05}

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 6 Mass-shift–amino-acid pairs with elevated frequency in exponential phase

Mass Shift	Amino Acid	Biol. Replicate	Exponential	Stationary	p-value
+16 Da	**W**	1	285/6836 (4.17%)	248/8775 (2.83%)	5.55×10^{-06}
		2	238/6182 (3.85%)	141/7242 (1.95%)	4.31×10^{-11}
		3	282/6182 (4.56%)	113/6414 (1.76%)	9.13×10^{-20}

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 7 Mass-shift–amino-acid pairs with elevated frequency in stationary phase

Mass Shift	Amino Acid	Biol. Replicate	Exponential	Stationary	p-value
+1 Da	**N**	1	1273/45738 (2.78%)	2260/52371 (4.32%)	2.49×10^{-38}
		2	1133/31975 (3.54%)	1663/36568 (4.55%)	3.02×10^{-11}
		3	988/40658 (2.43%)	1313/43285 (3.03%)	8.55×10^{8}
+42 Da	**S**	1	176/585 (30.09%)	322/534 (60.30%)	2.08×10^{-24}
		2	143/461 (31.02%)	287/509 (56.39%)	2.02×10^{-15}
		3	195/597 (32.66%)	269/570 (47.19%)	4.77×10^{-07}
+42 Da	**A**	1	55/261 (21.07%)	59/135 (43.70%)	4.15×10^{-06}
		2	55/202 (27.23%)	68/125 (54.40%)	1.22×10^{-06}
		3	38/210 (18.10%)	63/165 (38.18%)	2.08×10^{-05}
+48 Da	**C**	1	15/1604 (0.94%)	51/1217 (4.19%)	1.18×10^{-08}
		2	14/1285 (1.09%)	30/770 (3.90%)	4.80×10^{-05}
		3	1/964 (0.10%)	22/707 (3.11%)	6.93×10^{-08}
+42 Da	**T**	1	0/244 (0.00%)	57/363 (15.70%)	3.54×10^{-14}
		2	5/263 (1.90%)	51/321 (15.89%)	1.31×10^{-9}
		3	11/326 (3.37%)	40/423 (9.46%)	1.11×10^{-3}

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 8 Overlapping N-terminal Nα-acetylation targets between current data and *P. aeruginosa* [13]

E. coli Locus Tag	Locus	E. coli Peptide	P. aeruginosa Locus Tag	P. aeruginosa Peptide	Description
ECB_00686	sucB	SSVDILVPDLPESVADATVATWHKK	PA14_44000	MAIEIK	dihydrolipoamide succinyltransferase
ECB_03391	dppF	STQEATLQQPLLQAIDLKK	PA14_58490	METVLTAR	dipeptide transporter ATP-binding subunit
ECB_00915	rpsA	TESFAQLFEESLKE	PA14_23330	SESFAELFEESLK	30S ribosomal protein S1
ECB_00155	yadR	SDDVALPLEFTDAAANKV	PA14_08510	SIETFTPTPLLFTPGAANK	iron-sulfur cluster insertion protein ErpA
ECB_00183	accA	SLNFLDFEQPIAELEAKI	PA14_23860	SNWLVDKLIPSIMR	acetyl-CoA carboxylase carboxyltransferase subunit alpha
ECB_03467	secB	SEQNNTEMTFQIQRI	PA14_67720	TEQATNGAADEQQPQFSLQR	preprotein translocase subunit SecB

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Additional Files

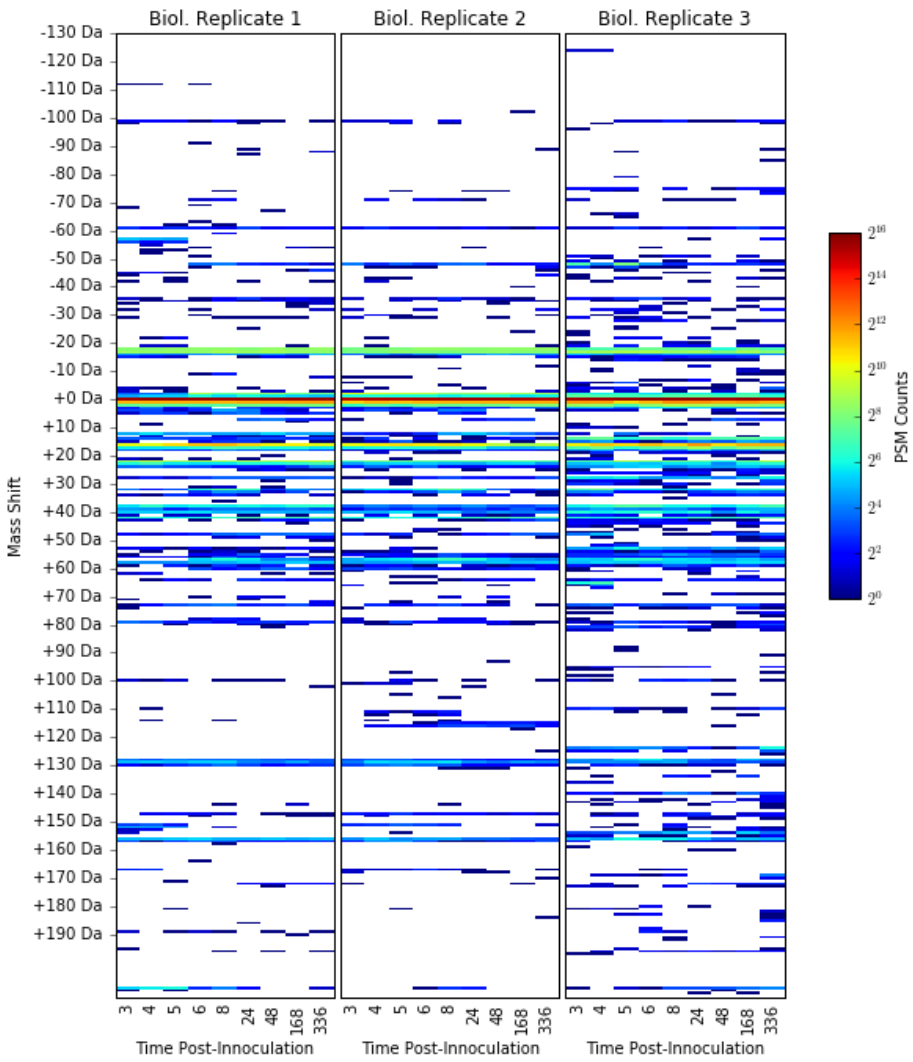

Figure S1 Abundance of all observed mass shifts across all 9 timepoints and 3 biological replicates.

Color of heatmap corresponds to the $\log_2$-transformed count of MODa-called modified PSMs in the 1% FDR set bearing the mass shift indicated on the y-axis for each of the nine timepoints (x-axis), for biological replicates 1, 2, and 3 (left, center, and right panels respectively). Although the MODa analysis was conducted for the mass window from -200 to $+200$ Da, no modifications were identified with mass shifts below -130 Da or above $+196$ Da.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

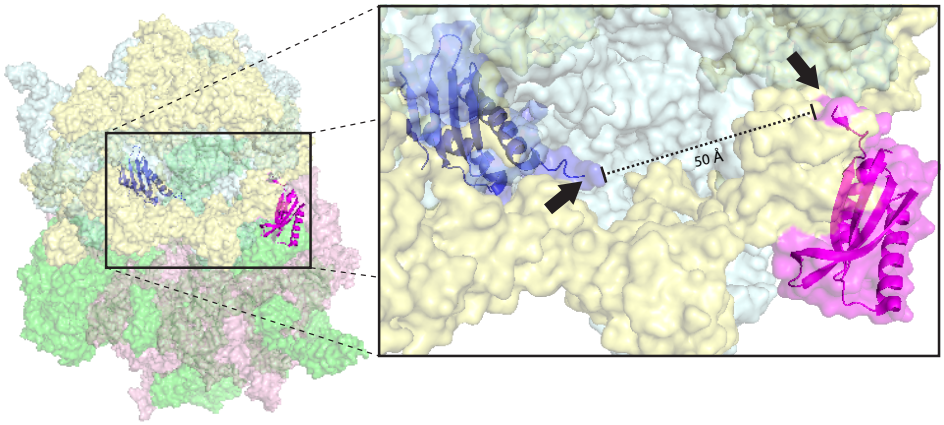

Figure S2 Location of YfiA and S6 in the 70S ribosome.
Relative locations of YfiA (blue) and native *T. thermophilus* S6 (magenta) proteins in crystal structure of *E. coli* YfiA bound to the *T. thermophilus* 70S ribosome (PDB ID 4V8I [45]). YfiA is positioned within the 30S subunit mRNA tunnel, and S6 on the outer surface of the 30S subunit; the C-terminal tails of both proteins (black arrows) point toward the same region of the 16S rRNA (light blue). The 17 C-terminal residues for YfiA, including the terminal glutamate residues, were not resolved in the crystal structure; the *T. thermophilus* S6 protein coding sequence ends at residue 101, lacking the 30-AA unstructured C-terminal domain present in *E. coli* S6. 16S rRNA is shown in light blue; 30S ribosomal proteins (other than S6) are shown in light yellow; 50S ribosomal proteins are shown in green; and 23S rRNA is shown in pink.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

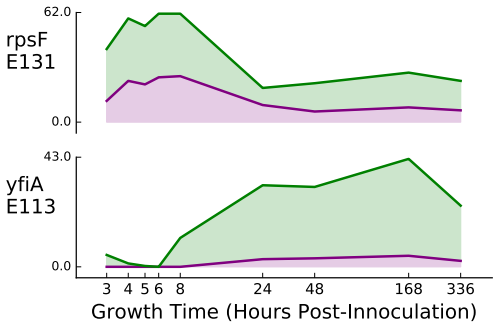

Figure S3 Modified and unmodified PSM counts for each AA position with a C-terminal +129 Da modification across all timepoints.
Plots show unmodified (green) and +129 Da modified (purple) PSM counts across all nine timepoints (x-axis) for the C-terminal position of the two proteins that have at least one PSM identified by MODa as containing a C-terminal +129 Da modification. Counts represent the average of the three biologcial replicates.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

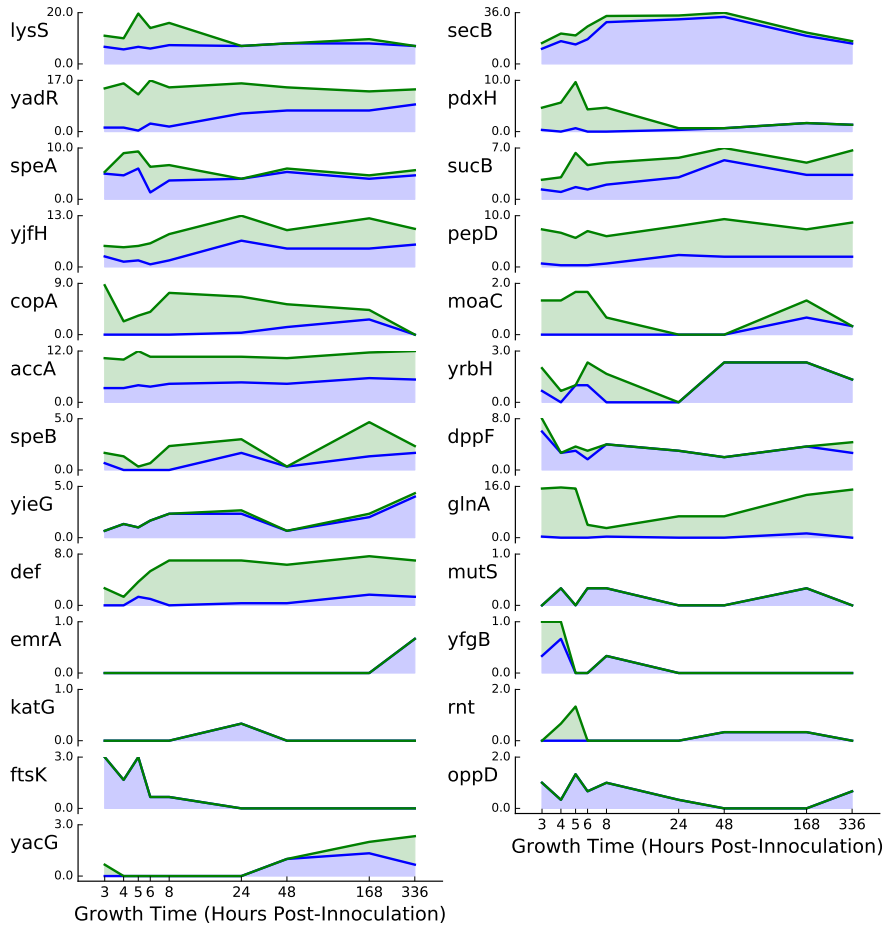

Figure S4 Temporally variable modification for individual proteins with an N-terminal serine possessing a +42 Da modification.

Plots show unmodified (green) and +42 Da Modified (blue) PSM counts across all nine timepoints (x-axis) for the N-terminal position of all proteins that have both (i) at least one PSM identified by MODa as containing an N-terminal +42 Da modification and (ii) having a penultimate serine (AA position 2; i.e. the N-terminal residue following N-terminal methionine cleavage). Counts represent the average of the three biological replicates. Plots are ordered from top to bottom by the mean p value of the Fisher's exact test for preferential modification (see text) from left-to-right within each row, and top-to-bottom across rows, with the most significant protein at the top left.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

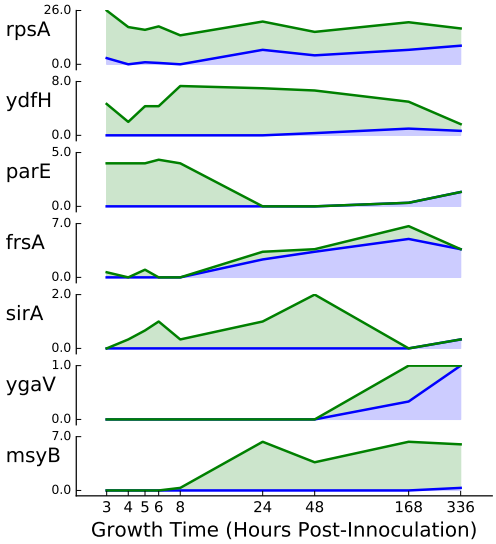

Figure S5 Temporally variable modification for individual proteins with an N-terminal threonine possessing a +42 Da modification.

Plots show unmodified (green) and +42 Da Modified (blue) PSM counts across all nine timepoints (x-axis) for the N-terminal position of all proteins that have both (i) at least one PSM identified by MODa as containing an N-terminal +42 Da modification and (ii) having a penultimate threonine (AA position 2; i.e. the N-terminal residue following N-terminal methionine cleavage). Counts represent the average of the three biological replicates. Plots are ordered from top to bottom by the mean p value of the Fisher's exact test for preferential modification (see text) from top to bottom, with the most significant protein at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

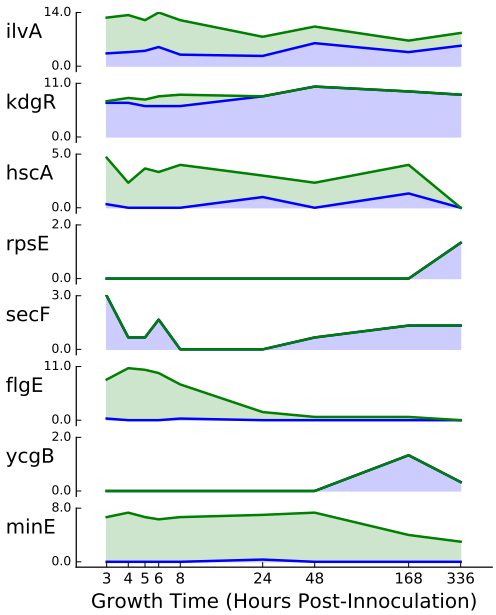

Figure S6 Fraction of total peptides across timepoints with an N-terminal alanine possessing a +42 Da modification.

Plots show unmodified (green) and +42 Da modified (blue) PSM counts across all nine timepoints (x-axis) for the N-terminal position of all proteins that have both (i) at least one PSM identified by MODa as containing an N-terminal +42 Da modification and (ii) having a penultimate Alanine (AA position 2; i.e. the N-terminal residue following N-terminal methionine cleavage). Counts represent the average of the three biologcial replicates. Plots are ordered from top to bottom by the mean p value of the Fisher's exact test for preferential modification (see text) from top to bottom, with the most significant protein at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

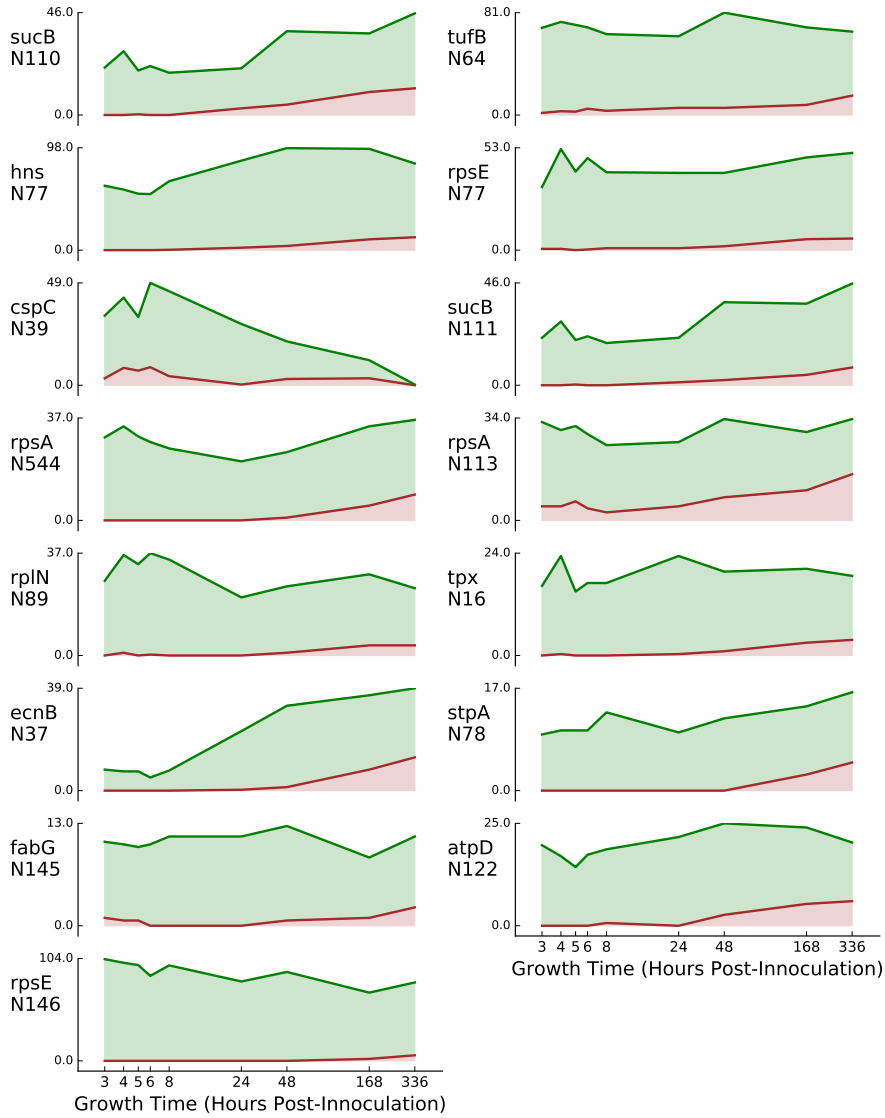

Figure S7 Modified and unmodified PSM counts for each AA position with a significantly stationary-phase biased +1 Da modification to asparagine.
Plots show unmodified (green) and +1 Da modified (brown) PSM counts across all nine timepoints (x-axis) for the 10 asparagine residues with the most significant p-values across all three biological replicates. Counts represent the average of the three biologcial replicates. Plots are ordered by the mean p value of the Fisher's exact test for preferential modification from left-to-right within each row, and from top-to-bottom across rows, with the most significant position at the top left.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

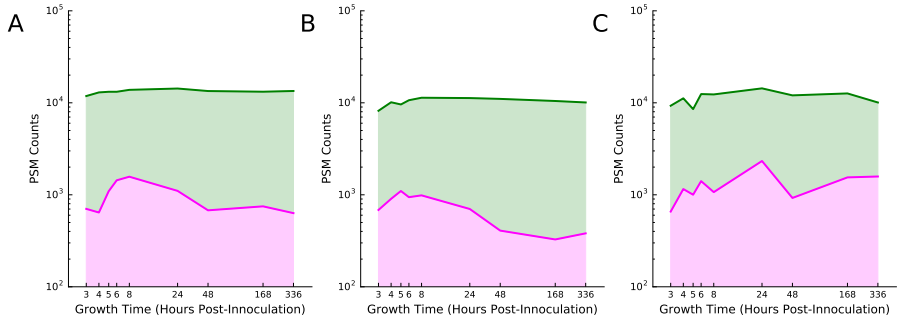

Figure S8 Modified and unmodified counts across timepoints for all AA positions with a +16 Da modification to methionine, pooled by biological replicate.

Plots show the total unmodified (green) and +16 Da modified (magenta) PSM counts across all nine timepoints (x-axis) for methionine residues that have at least one +16 Da modification at any time point in any replicate. The three panels show counts for each of the three biological replicates (A), 2 (B) and 3 (C). Note that the y-axis is plotted on a logarithmic (base 10) scale due to the high number of total counts relative to modified counts.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

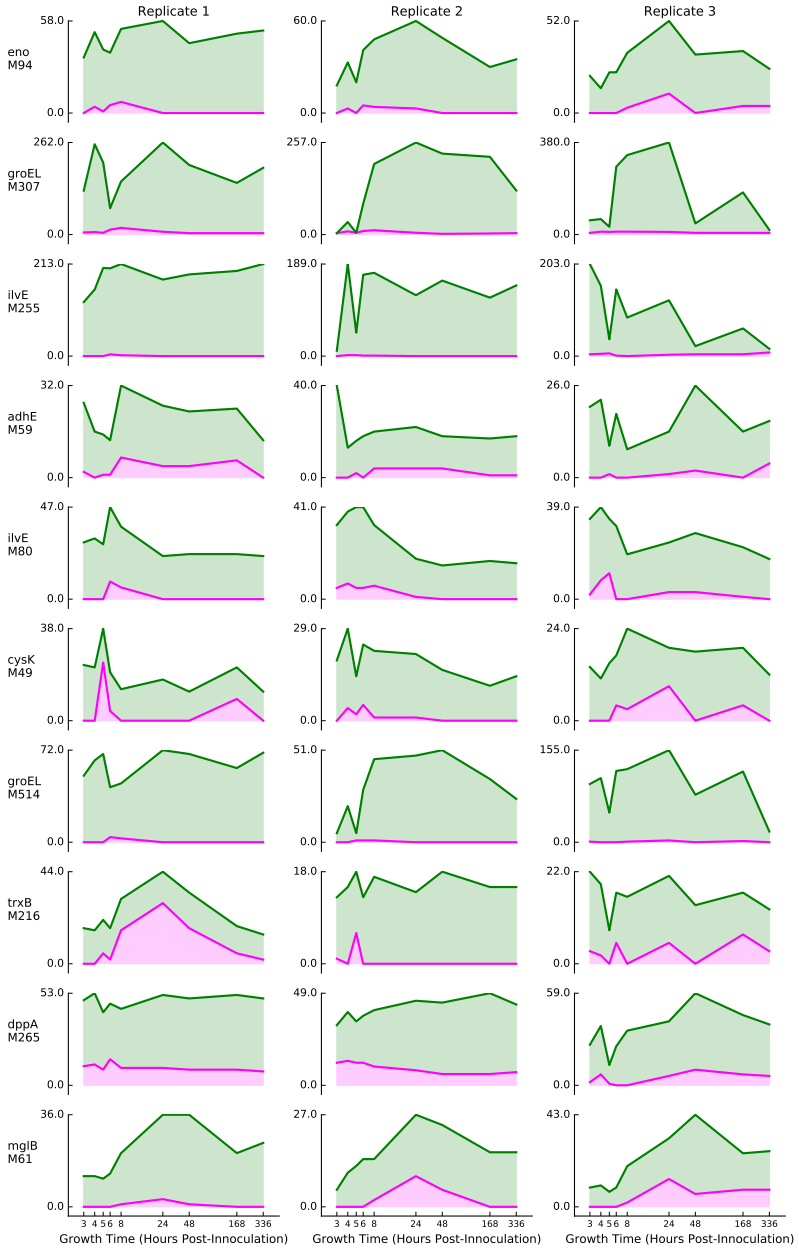

Figure S9 Modified and unmodified counts across timepoints for the top 10 exponential-enriched AA positions with a +16 Da modification to methionine.

Plots show unmodified (green) and +16 Da modified (magenta) methionine PSM counts across all nine timepoints (x-axis) for the protein and position indicated. Plots in columns correspond to the three biological replicates 1 (left column), 2 (center column), and 3 (right column). Counts represent the average of the three biologcial replicates. Plots are ordered from top to bottom by the mean p value of the Fisher's exact test for preferential modification (see text), with the most significant protein at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

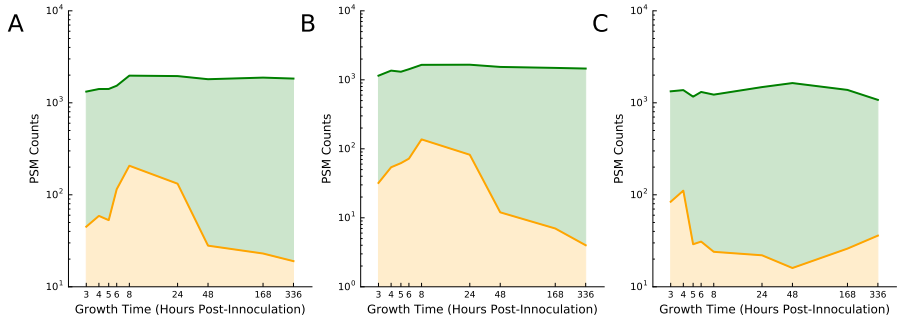

Figure S10 Modified and unmodified counts across timepoints for all AA positions with a +16 Da modification to tryptophan, pooled by biological replicate.
Plots show the total unmodified (green) and +16 Da modified (orange) PSM counts across all nine timepoints (x-axis) for tryptophan residues that have at least one +16 Da modification at any time point in any replicate. The three panels show counts for each of the three biological replicates 1 (A), 2 (B) and 3 (C). Note that the y-axis is plotted on a logarithmic (base 10) scale due to the high number of total counts relative to modified counts.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

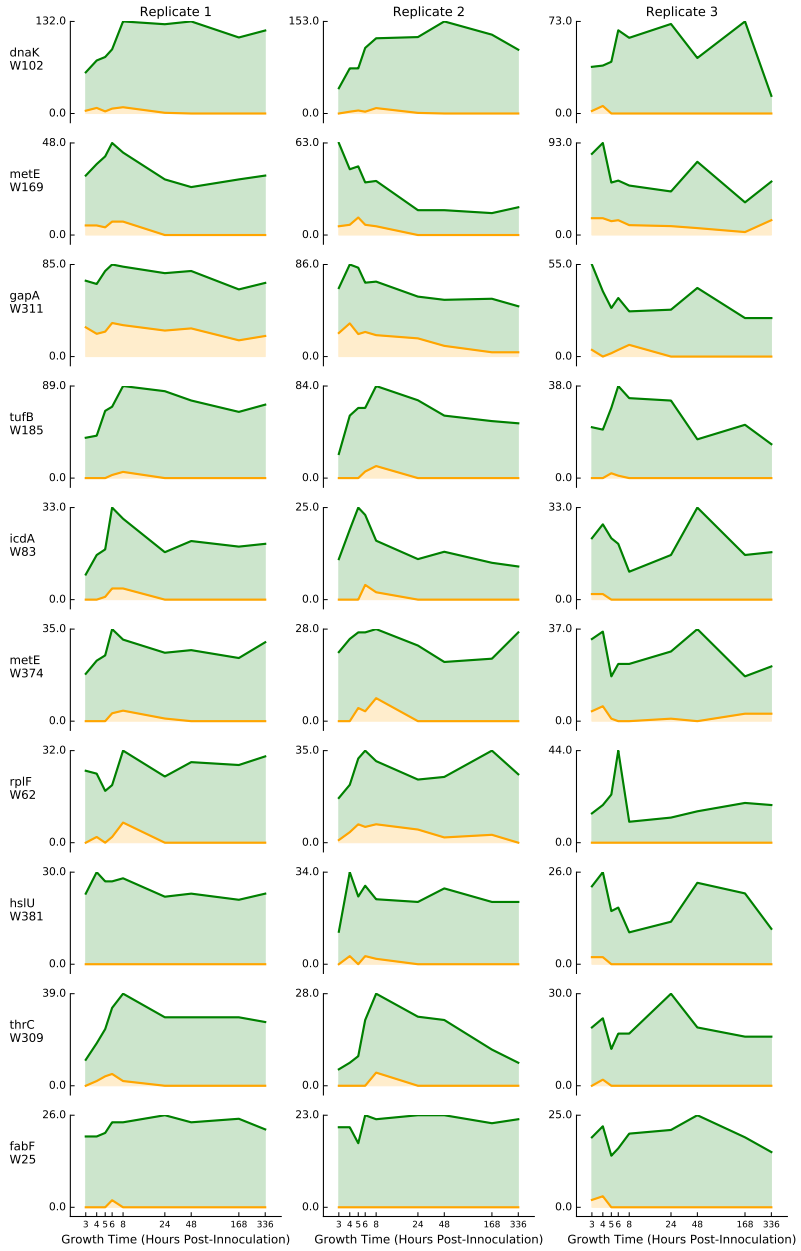

Figure S11 Modified and unmodified counts across timepoints for the top 15 exponential-enriched AA positions with a +16 Da modification to tryptophan.

Plots show unmodified (green) and +16 Da modified (orange) tryptophan PSM counts across all nine timepoints (x-axis) for the protein and position indicated. Plots in columns correspond to the three biological replicates 1 (left column), 2 (center column), and 3 (right column). Counts represent the average of the three biologcial replicates. Plots are ordered from top to bottom by the mean p value of the Fisher's exact test for preferential modification (see text), with the most significant protein at the top.

bioRxiv preprint doi: https://doi.org/10.1101/051185; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Additional file 12—Data Tables

Zip file containing several data tables in tab-separated format, as well as a readme file that explains the contents of each data file.