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

Viswanadham Sridhara[1], Daniel R. Boutz[2,3], Craig Barnhart[3], Jeffrey E. Barrick[1,2,3,4], Edward M. Marcotte[1,2,3,4,5], and Claus O. Wilke[1,2,3,5]

November 26, 2014

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

Abstract

A

1 Introduction

Most of the times, specific biological function of the proteins is identified by the post-translational modification associated with it. Much work has been done in identifying the protein coding genes, however, identification of the functional PTMs is still at its infancy. Mass-spectrometry (MS) based proteomics is the only technique that can characterize these PTMs in large scale. Peptide search algorithms are then run on these mass-spectrometry data sets to identify peptides and PTMs associated with it. Here we used MS on E. coli proteome grown under glucose-limiting conditions and then identified various PTMs in large scale.

Identifying a targeted list of PTMs (mostly up to 6) is a general norm with most of the sequence search algorithms. As the number of target PTMs increase, search space and the false positives generally increase that limits the number of input PTMs. However recently, software for unrestricted search of PTMs is starting to grow. We used MODa, a naive based multi-blind spectral alignment algorithm, to look for PTMs in this dataset.

Most of the studies till date limit to capturing RNA/proteome information to 1 or 2 time points during exponential and stationary phases for analysis. Here we collected MS data on E. coli proteome at 9 different time points ranging from exponential to long-stationary phases up to 2 weeks

under glucose-limiting conditions. To our knowledge this is a first attempt to look into time-course of PTMs over 2 weeks of a microbe.

[31] Protein oxidation and aging

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

[14] Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase,

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

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

[38] Origin and evolution of the protein-repairing enzymes methionine sulphoxide reductases

This fixation of MsrA is helpful [1] Enzymatic reduction of oxidized alpha-1-proteinase inhibitor restores biological activity

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

Mascot [25], Sequest [10], OMSSA [12], X!Tandem [6], TagRecon [7], MODa [21], and Byonic [2]

Large-scale analysis of PTMs [23]

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

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

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

Msr fixing MetSO to Met [3] Enzymatic reduction of protein-bound methionine sulfoxide

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

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

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

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

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

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

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

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

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

[37] Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12

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

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

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

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

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

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

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

2   Results

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

2.1   Running MODa on *E. coli* proteome

We did this analysis using MODa, a naive based spectral alignment search algorithm. Unlike other peptide identification search algorithms, MODa requires as an input the range of PTM masses instead of the actual PTMs. So, for our analysis we used a range of -200 to 200 Da which is typical to this search engine. We generated mass-spec raw data at different time points in exponential and stationary phases of E. coli growth. At each of the 9 time points, we have 3 biological replicates. For each of the replicates, we ran MODa. We asked some of the following questions while analyzing the data: (a) How much of the proteome is modified? (b) How do these modifications change over time? (c) Can we find any novel modifications? (d) Can we explain the pattern of the PTMs over time using other diverse kinds of data such as RNA-seq?

2.2   Modified *E. coli* proteome

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

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

Since MODa outputs mass-shift, we used UNIMOD database to map the mass-shift to the probable PTM. MODa outputs not only the mass-shifts, but also the amino-acids on which this mass-shift occurs. We focussed not only on the most common PTM-amino acid combination, but also on the PTMs irrespective of the amino acid. For example, we know acetylation is most common on n-term

and lysine, however we also looked at the time-course of all acetylations i.e., sum of acetylations on all amino acids. (Add other citations and results from literature)

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

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

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

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

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

2.5 Oxidative damage and repair in *E. coli*

Finally, we investigated oxidative damage by looking at the oxidations identified by MODa program i.e.., mass-shifts of +16Da. The oxidations seem to go down from exponential to stationary phases. This result seemed strange as we expected oxidative damage and hence the oxidations to go up as the E. coli reaches stationary and long-stationary phases. However may be the sulfoxide reductases MsrA and MsrB found in e. coli seem to fix methionine sulfoxide back to methionine. Since we also gather RNA-seq data on the same sample, we were able to retrieve the transcriptomic abundances of MsrA and MsrB along with proteomic abundances. There is not much change in

the profiles of transcripts of these reductases i.e., the abundance seem to go down from exponential to stationary and stayed flat during long stationary phase. However the proteomic abundance seem to go down initially from exponential to stationary phase, however it surged up during the long-stationary phase indicating the probable fixing of MetSO to Met. Previously it was shown that these reductases play a role, however in this study for the first time we showed the time course fixation of MetSO to Met.

2.6   Novel phosphoserinegluconylation on ribosomal protein S6

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

2.7   in *E. coli*

3   Discussion

Limitations: Large-scale PTM analysis is still at its infancy because of both experimental enrichment techniques of PTMs as well as the computational search algorithms that have to look at different combinations to identify the PTMs. Here we used MODa, a naive based search algorithm on E. coli proteome data obtained under glucose-limiting conditions. Our analysis showed that 1/3rd of the peptide spectral matches found are modified. We also saw a decrease in oxidation levels from exponential to stationary phase, with sulfoxide reductases playing a role in fixing Methionine oxide back to methionine. n-term acetylations are frequent, with serine n-term acetylation happening 75% of all n-term acetylations, all the times on the 2nd amino acids of the protein, after methionine excision. Phosphorylation increased from exponential to stationary phases, while carboxylations and nitrosylations remained constant throughout the time course of E. coli growth. Na and K adducts seem to happen 3-4% of the time. Finally, a novel serine modification i.e., phosphogluocnylation of serine seems to happen on ribosomal protein S6 frequently.

Most of the peptide identification search algorithms required a list of PTMs to search for and usually this list is limited to 6. However databases like UNIMOD, RESID have listed at least 400 known PTMs present. So, we used a naive based search algorithm that outputs mass-shifts, instead of the PTMs in the peptide-spectral match. Then we use UNIMOD to identify the PTM that matches the mass-shift. However we did not try and match all the mass-shifts, but investigated in detail the frequent mass-shifts identified by MODa. One limitation is we used the default mass-range search between -200 and 200 Da and one other search with -100 and 300 Da range. So, we miss out on larger PTMs i.e., polyubiquitination tails etc. This resulted in analysis of +1 Da mass-shift (C13 peak detection), oxidation, acetylation, Na and K adducts, along with some widely studied PTMs in phosphorylation, carboxylation and nitrosylation. This program is previously used for similar large-scale analysis with urinary proteomics and they identified novel PTMs. However the study used 2 programs and considered the overlap of PTMs as highly confident. In-

stead here, our focus is not to identify highly-confident PTMs later used as biomarkers, but to get a wider coverage at a lower FDR of 1% and look at the time course or evolution of these PTMs during the entire 2 weeks of E. coli growth.

To avoid false positives and as recommended by the original MODa paper, we ran search engine looking for only 1 possible modification on the peptide. Generally the PTMs are validated by a 2nd round of identifying highly confident hits. This is usually done by using programs like Ascore etc. However here, we did not do any 2nd round of validating peptides. Also, we looked at the global level of PTMs, instead of counting which peptide and hence protein is modified. We did not do any PTM level quantitation or try to undestand the specific PTM stoichiometry, as these require sophisticated experimental instrumentation, protocol and the algorithms to characterize the PTMs associated with proteins.

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

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

Copied from the same paper At the onset of glucose starvation, cyclic AMP and its receptor protein (cAMP-CRP) were found to play important roles in the expression of a number of genes. An early 2-DE study identified five glucose-responsive outer membrane proteins (four upregulated and one downregulated) (186). A comparison with membrane proteins from mutant strains revealed that two of the upregulated proteins were the receptors for lambda and T6, and coelectrophoresis of the outer membrane fraction identified the downregulated protein as OmpA. The glucose starvation stimulon was further examined using 2-DE followed by comparison to the E. coli gene-protein database (218). Members of this stimulon were found to include enzymes of the Embden-Meyerhof-Parnas pathway, phosphotransacetylase (Pta) and acetate kinase (AckA) in the acetic acid pathway, and formate transacetylase. Trichloroacetic acid cycle enzymes were repressed, whereas enzymes involved in acetate and formate production and the Embden-Meyerhof-Parnas pathway were induced. These modulations suggest that a glucose-starved cell increases the relative flow of carbon through the Pta-AckA pathway. Indeed, pta and pta-ackA mutants were found to be impaired in their abilities to survive glucose starvation, indicating that the capacity to synthesize acetyl phosphate, an intermediate of this pathway, is indispensable for glucose-starved cells. The pta mutant failed to induce several proteins of the glucose starvation stimulon. More recently, proteome studies revealed that glucose limitation upregulates the levels of proteins such as AceA, AldA, ArgT, AtpA, DppA, GatY, LivJ, MalE, MglB, RbsB, UgpB, and YdcS (311). Of these, ArgT, DppA, LivJ, MalE, MglB, RbsB, UgpB, and YdcS are periplasmic binding proteins of the ABC transporters, suggesting that in addition to the central metabolism proteins, periplasmic binding proteins are involved in the carbohydrate and amino acid uptakes that are important during glucose limitation.

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

Total proteins could be the same - should I plot the total number of proteins over time? absolute

protein abundances calculated from APEX protocol?

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

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

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

4   Conclusions

We have found

5   Materials and Methods

5.1   *E. coli* growth

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

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

5.2   Mass-spectrometry of *E. coli* proteome

Dan Boutz part (Copied from John's paper): Frozen pellets where re-suspended in 300 uL of buffer (50mM Tris-HCL pH 8.1,100mM KCL, 5mM MgCl2). 50UL of sample was removed for preparation, treated with 50uL trifluoroethanol (TFE), and placed on ice for 15min. DTT was then added

to a final concentration of 5mM and incubated at 55C for 45 min. Samples where alkylated by addition of Iodoacetamide (IAM) to a final concentration of 15mM. Trypsin digest was performed by addition of 800uL of 50mM Tris pH 8.0, 2mM CaCl2 followed by 2ug of trypsin. Digestion took place at 37C for 4-5 hrs and was stopped by 10uL of formic acid (1% vol./vol). Samples where then ultrafiltrated to remove insoluble and undigested material using Amicon Ultra MWCO 10kD spin-caps and finally concentrated and purified by C18 filtration. Liquid chormotography and mass spectrometry (LC/MS) was carried out on a LTQ-Orbitrap (Thermo Fisher). ¡particular settings need to be filled in¿

5.3 Post-translational modification identification and analysis

We used MODa (cite) to identify peptides and characterize novel PTMs. MODa outputs mass-shifts on the amino acids instead of the post-translational modifications. So, we used UNIMOD database to map the mass-shift to the appropriate PTM or a mutation. We ran separate searches for each of the 9 time points and at each time point, there were 3 biological replicates. So, there were in total 27 MODa searches. We used TACC for computing resources. The enzyme used in the searches is trypsin with fully-tryptic and no proline rule. The missed cleavages allowed are 2. Since the fragmentation technique used is CID, we looked for b/y ions. The mass-tolerance of the precursor is 10ppm, while the mass-tolerance of the product ion is 0.5 Da. We used car-bamidomethylation of cysteine as a static or fixed modification. As mentioned earlier, MODa requires a mass range to search for variable modifications, so we tried 2 scenarios: (a) mass range between -200 to 200Da and (b) second search with mass range between -100 to 300Da. We used REL606 sequence library from NCBI sequence database. To identify high-confidence hits, we used target-decoy approach (cite). In this approach, we reverse the original REL sequences and concatenate to the original sequence database to form a database that is twice as much as of the original sequence database. The idea is that there are as many false positive hits to that of the original database as that of the decoy database. We used a 1% FDR in this approach which is a general norm in mass-spectrometry based proteomics searches.

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

5.4 Raw data and analysis scripts

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

6 Author Contributions

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

7 Acknowledgments

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Figures

Figure 1. Flowchart. We use .

Figure 2. Misclassification . For j.

Figure 3. Probability . For .

Figure 4. Predictive . Reactions .

Figure 5. Predictive . Reactions .

Supplementary Figures

Figure S1: Scatter . We .

Figure S2: . We .

Figure S3: . For .

Supplementary Tables

Table S1: . These .

Table S2: . These .