Paper # K-125

Evaluation of Periplasmic Proteome Enrichment by Differential Isotope Labeling and Mass Spectrometry

W. Judson Hervey IV[1,2], Chongle Pan[1,2], Dale A. Pelletier[2], and Gregory B. Hurst[2]

[1]UT-ORNL Graduate School of Genome Science and Technology, Oak Ridge, TN 37830 [2]Oak Ridge National Laboratory, Oak Ridge, TN 37831

Contact Information:
P.O. Box 2008 MS-6131
Oak Ridge, Tennessee 37831
p: (865) 576-4546 f: (865) 576-8559
email: herveywjiv@ornl.gov

ABSTRACT

- Isolation of selected cellular fractions is a prerequisite to more detailed proteomic characterization of microbial cells. Methods of enriching the periplasmic proteome have complicated its characterization, due to incomplete removal of cytosolic proteins.
- To assess the efficiency of periplasmic enrichment protocols in removing cytosolic proteins, we have developed a methodology that employs differential isotopic labeling of periplasm-enriched versus whole cell lysates from *E. coli*. Identical cultures were grown in separate ^{14}N- and ^{15}N-enriched media. Periplasmic enrichments and whole cell lysates were prepared from both "light" and "heavy" isotope-labeled cultures.
- Mixtures of differentially-labeled periplasmic and cell lysate at equal protein concentration were prepared. Tryptic digests of these protein mixtures were analyzed by LC-MS/MS. Tandem mass spectra were searched with the SEQUEST algorithm against the predicted *E. coli* proteome.
- Isotope abundance ratios were calculated at the peptide and protein levels using the program ProRata. An average of 87% of the identifications with periplasmic:cytoplasmic isotope ratios greater than 1 were found to be annotated as periplasmic in the SwissProt database. In replicate experiments, 39 known periplasmic proteins and 22 proteins with hypothetical or unknown function were identified as periplasmic-enriched.
- Our approach highlighted the identification and quantification of both known periplasmic proteins and signal peptide-containing hypothetical proteins at high abundance, suggesting it is effective at differentiating the periplasmic proteome from cytosolic proteins.
- 81% of the periplasmic-enriched protein quantifications are known periplasmic proteins, signal peptide-containing hypothetical proteins, and membrane proteins with peripheral periplasmic domains.

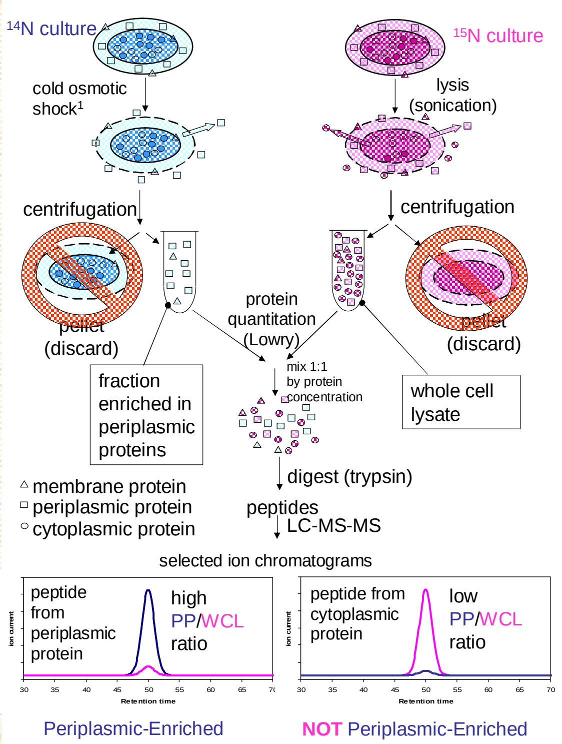
INTRODUCTION

- Periplasmic proteins play important roles in electron transport, binding of nutrients, cell wall biosynthesis, and modification of molecules that will eventually enter the cytoplasm[2]. Isolation of these proteins from the proteome is integral to a more comprehensive understanding of these cellular processes.
- The cold-osmotic shock method of enriching periplasmic proteins[1] often results in a residual amount of cellular lysis.
- Previous LC-MS/MS-based studies of the periplasmic proteins of the *E. coli* K12 proteome[3-6] have enriched periplasmic proteins by the cold-osmotic shock enrichment method. The presence of cytoplasmic proteins within the enrichment complicates the interpretation of the results.
- Primary sequence analysis of protein identifications present in the periplasmic enrichment has been a primary method of distinguishing periplasmic proteins from cytoplasmic proteins.
- Recently, quantitative proteomics methodologies have been applied to determine the localization of proteins in the Golgi apparatus[7], endoplasmic reticulum[7], and mitochondria[8] in eukaryotic organisms.
- Quantitative proteomics techniques provide the measurement of protein abundance. This measurement is often used to compare either the absolute or relative abundance of one protein in a cell under one state relative to the abundance of the same protein under another state. This measurement may also be used as additional source of data to infer subcellular localization[7-8].
- In this study, we have developed a quantitative proteomics methodology to evaluate the cold-osmotic shock method[1] of protein enrichment at liberating periplasmic proteins from the cell.
- The measurement of the relative abundance of proteins from cold-osmotic shock periplasmic enrichments to proteins from the whole cell lysate in differentially-labled protein mixtures provides a measurement that may be used to distinguish *bona fide* periplasmic proteins from cytoplasmic proteins.
- In this methodology, a protein's relative abundance in the differentially-labled protein mixtures is expressed as the periplasmic-to-whole cell lysate (PP:WCL).
- *Bona fide* periplasmic proteins should have a high PP:WCL ratio in the differentially-labeled protein mixtures, while cytoplasmic proteins originating from cellular lysis during the cold-osmotic shock should have a low PP:WCL ratio.

METHODS

- Identical *E. coli* K12 MG1655 cultures were grown separately in ^{14}N- and ^{15}N- enriched modified M9 Minimal Medium. Ammonium sulfate was substituted as the sole nitrogen source.
- Cold osmotic shock-based periplasmic enrichments[1] and whole cell lysates were prepared from both "light" and "heavy" isotope-labeled cultures.
- Three differentially-labeled isotopologous protein mixtures were created at equal ^{14}N:^{15}N protein concentrations as determined by the Lowry Assay:
  - ^{14}N periplasmic enrichment mixed with ^{15}N periplasmic enrichment (not shown)
  - ^{14}N periplasmic enrichment mixed with ^{15}N cell lysate (illustrated below)
  - ^{15}N periplasmic enrichment mixed with ^{14}N cell lysate (not shown)
- Enzymatic proteolysis of the protein mixtures was performed with trypsin. The resulting proteolytic peptides from each mixture were separated by a two-dimensional chromatography system (strong cation-exchange in the first phase and reverse phase in the second phase) coupled to a linear ion trap mass spectrometer (ThermoFinnigan LTQ). Duplicate LC-MS/MS experiments were conducted for each protein mixture.
- Tandem mass spectra were searched with the SEQUEST[9] algorithm against the 4,337 proteins of the predicted *E. coli* K12 proteome[10]. Identifications were filtered and sorted by the DTASelect[11] program.
- Isotope abundance ratios were calculated at the peptide and protein level by the ProRata[12] program.
- Amino terminal signal peptide predictions were obtained by primary sequence analysis with the PredISI[13], SignalP[14], PSORTb[15], and TATFIND[16] algorithms.
- The experimental overview is outlined below in Figure 1.

Figure 1: Experimental Overview



RESULTS

Overview

- Isotopic ratios of LC-MS/MS-based peptide identifications were estimated at both the protein and peptide level by the ProRata[12] program in two different types of differentially-labeled protein mixtures:
  - ^{14}N Periplasmic Enrichment mixed with a ^{15}N Periplasmic Enrichment (Control)
    - 96 Protein Quantifications
    - Distribution of protein quantifications is shown in Figure 2A
  - Differentially-Labeled Mixtures of Periplasmic Enrichments and Whole Cell Lysate
    - 667 Protein Quantifications
    - Distribution of protein quantifications is shown in Figure 2B
- Periplasmic to whole cell lysate (PP:WCL) isotopic ratios for 667 protein identifications were estimated by the ProRata[12] program from LC-MS/MS-based peptide identifications. These protein quantifications were categorized by SwissProt functional annotation[17], amino-terminal signal peptide predictions by primary sequence analysis tools[13-14], and the observed value of the PP:WCL ratio. The relationships among the 667 observed protein quantifications are shown in Figure 3.
- The estimated PP:WCL ratios, confidence intervals, and number of observed tandem mass spectra are shown in Figure 4 for the twelve proteins with the largest PP:WCL ratios.
- Of the 667 protein quantifications, twelve annotated periplasmic proteins[17] were estimated with low PP:WCL ratios. These twleve proteins are listed in Table 1 with their corresponding PP:WCL ratio and confidence intervals observed from one LC-MS/MS experiment.

Distributions of Protein Quantifications

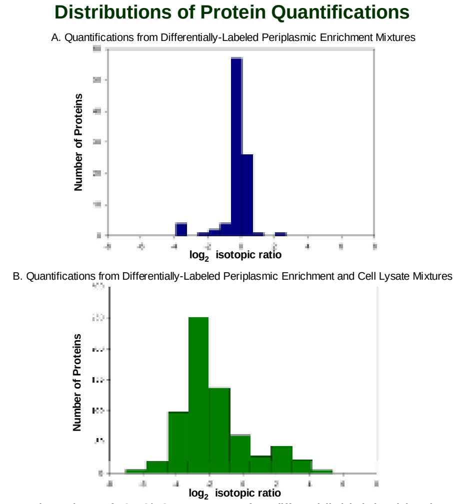


Figure 2: The control experiment of LC-MS/MS measurements from differentially-labeled periplasmic enrichments is shown in 2A. In this experiment, 96 proteins were quantified with $\log_2$ isotopic ratios ranging from 0.6 to -1.4. Figure 2B depicts the distribution of the $\log_2$ isotopic ratios among the 667 protein quantifications in the differentially-labeled mixture of periplasmic-enriched proteins and cell lysate. This ratio is an measurement of abundance of the protein in the periplasmic enrichment *relative* to the abundance in the cell lysate. The values of the $\log_2$ isotopic ratios shown in Figure 2B range from 5.38 to -7.

Table 1: Known Periplasmic Proteins with Low PP:WCL Ratios

SwissProt ID	Description	PP:WCL Ratio	Lower CI	Upper CI
[illegible]	Thiol peroxidase	-0.15	-1.075	0.8
[illegible]	D-ribose binding protein	-0.25	-0.85	0.35
[illegible]	Protein yhjJ	-0.45	-3.25	3.15
[illegible]	Glucans biosynthesis protein D	-0.5	-1.6	0.6
[illegible]	Beta-lactamase precursor	-0.8	-4.1	0.4
[illegible]	Glucans biosynthesis protein G	-1.5	-3.27	0.325
[illegible]	Vitamin B12 transport periplasmic protein	-2.1	-5.2	-1.25
[illegible]	Me-accepting chemotaxis protein	-2.35	-3.25	-1.6
[illegible]	Tail-specific protease	-3.4	-5.6	-2.26
[illegible]	Biosynthetic arginine decarboxylase	-3.45	-6.7	-2.05
[illegible]	Me-accepting chemotaxis protein II	-3.6	-4.4	-2.8
[illegible]	Protease degP Precursor	-4.1	-6.7	-3.1

Relationships Among Protein Quantifications

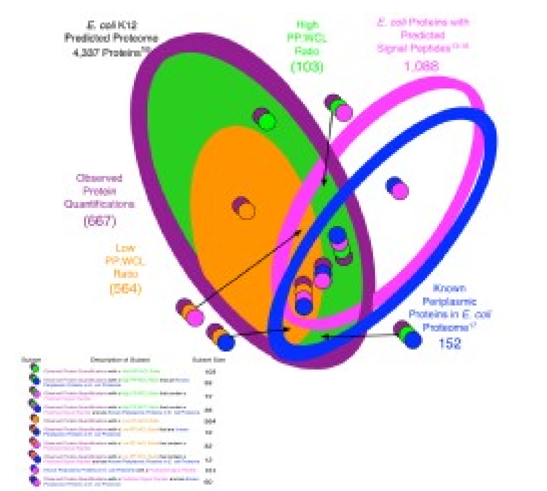


Figure 3: Venn Diagram of *E. coli* K12 Protein Quantifications: High PP:WCL Ratio is defined as $\log_2$ periplasmic:whole cell lysate (PP:WCL) ratio > 0, indicating enrichment of a protein in the cold osmotic shock procedure. Low PP:WCL Ratio is defined as a $\log_2$ periplasmic: whole cell lysate (PP:WCL) ratio < 0, indicating no enrichment of a protein in the cold osmotic shock procedure. Proteins with predicted signal peptides contain an amino-terminal signal peptide predicted by either the PredISI[13], SignalP NN[14], SignalP HMM[14], PSORTb[15], or TATFIND[16] algorithms. Finally, the known periplasmic proteins subset is based upon the SwissProt annotation[17] of the predicted E. coli K12 proteome.

Protein Quantifications with Highest PP:WCL Ratios

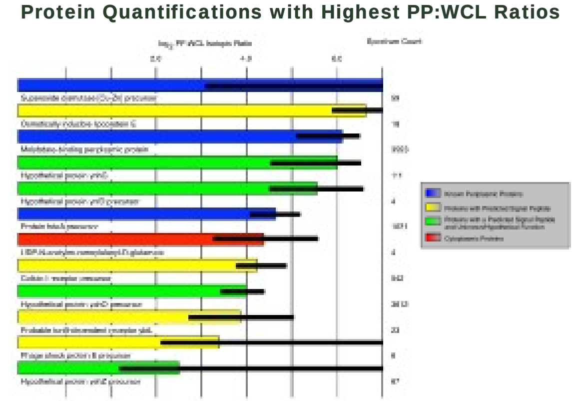


Figure 4: The twelve proteins with the highest estimated PP:WCL ratios (in $\log_2$ scale) from both types of differentially-labeled mixtures are listed. The values of the PP:WCL ratios and confidence intervals (represented in black) are from one LC-MS/MS experiment (the protein mixture of ^{14}N Periplasmic Enrichment and ^{15}N Whole Cell Lysate). The color of the graph representing the PP:WCL ratio indicates the functional annotation of each protein based upon experimental evidence[17] or analysis of the primary protein sequence by signal peptide prediction algorithms[13-16]. The values for the spectrum count for each protein from this LC-MS/MS experiment are represented to the right of each bar. This value represents the total number of tandem mass spectra identified for each protein.

DISCUSSION

- The ProRata[12] program estimated the periplasmic to whole cell lysate (PP:WCL) isotopic ratios for peptide identifications from the following types of differentially-labeled protein mixtures:
  - ^{14}N periplasmic enrichment mixed with ^{15}N cell lysate
  - ^{15}N periplasmic enrichment mixed with ^{14}N cell lysate
  - Each protein mixture was analyzed by LC-MS/MS in two technical replicates.
- PP:WCL isotopic ratios of proteins with a minimum of two peptide identifications in at least one experimental replicate were averaged, resulting in computed PP:WCL isotopic ratios (on $\log_2$ scale) for 667 protein identifications. In general, PP:WCL isotopic ratios for periplasmic proteins enriched during the osmotic shock-based method should be larger than proteins arising from unwanted cellular lysis. The relationships among all 667 protein quantifications are shown in Figure 3.
- Of the 667 proteins for which PP:WCL isotopic ratios were estimated:
  - 103 of 667 proteins displayed $\log_2$ isotopic ratios > 0 (high PP:WCL ratio), indicating enrichment in the periplasmic protein isolation protocol. The 103 protein quantifications displaying high PP:WCL ratios were categorized on the basis of SwissProt annotation[17] and previous LC-MS/MS-based identification[3-6], amino-terminal signal peptide predictions from signal peptide prediction algorithms[13-16], and membrane topology sequence analysis by the PSORTb algorithm[15].
  - When these 103 proteins are compared to the 152 experimentally verified or predicted periplasmic proteins in the SwissProt database[17]:
    - 39 of the 103 proteins were functionally annotated as periplasmic proteins.
      - Among proteins exhibiting large PP:WCL isotopic ratios were the periplasmic oligopeptide binding protein (OppA), the periplasmic dipeptide transport protein (DppA), the lysine-arginine-ornithine binding periplasmic protein (ArgT), the periplasmic dipeptide transport protein (DppA), and the molybdate-binding periplasmic protein (ModA); all of which have been identified by other LC-MS/MS-based studies[3-6].
    - 43 of the 103 proteins displaying high PP:WCL ratios are *not* functionally annotated as periplasmic, but *do contain a predicted amino-terminal signal peptide* by either the PredISI[13], SignalP[14], PSORTb[15], or TATFIND[16] algorithms.
      - Further sequence analysis was conducted with the PSORTb program to search the primary sequence of the 43 proteins for membrane topology in order to identify co-enriching membrane proteins within the periplasmic enrichment. This strategy allowed the identification of 21 known membrane proteins, including OmpA and OmpX, from 22 potentially novel periplasmic proteins.
    - 21 of the 103 proteins displaying high PP:WCL ratios are functionally annotated as cytoplasmic proteins, most likely present within the periplasmic enrichment due to lysis of the inner membrane during cold osmotic shock.
      - Many cytoplasmic proteins in the periplasmic enrichment are abundant cellular proteins consistently observed in each LC-MS/MS experiment, such as the 50S ribosomal protein L4.
  - 564 of the 667 total protein quantifications displayed $\log_2$ isotopic ratios < 0, indicating a higher abundance of these proteins in the whole cell lysate relative to the proteins liberated from the periplasm by cold osmotic shock. These proteins were designated "not periplasmic-enriched."
    - 12 of the 564 proteins with low PP:WCL ratios are known periplasmic proteins (Table 1). The observed low PP:WCL ratios for many of these proteins are reasonable in some cases.
      - For instance, the D-ribose binding protein, RBP, is a periplasmic binding protein but also functions in chemotaxis in concert with other membrane proteins[18]. It is possible that other copies of the RBP protein were sequestered at the membrane in this function rather than free in the periplasm to bind ribose. The interaction between the RBP protein and the membrane-bound chemotaxis receptor is a plausible explanation for the observed low PP:WCL ratio.
      - The beta-lactamase precursor, AmpC, is a another protein with a low PP:WCL ratio. Expression of AmpC is generally low but inducible under lactam stimuli[19]. Low expression of this protein under the growth conditions of this study likely attributed to the estimation of an inaccurate PP:WCL ratio for this periplasmic protein. The observed spectrum count for this protein is 2, the minimum cutoff value for both identification and quantification.
      - The BtuE protein is likely a protein that interfaces with membrane transporters in B12 transport and the true periplasmic binding protein is the BtuF protein[20] for the Vitamin B12 ABC transporter. As BtuF, not BtuE, is the periplasmic binding protein for the Vitamin B12 ABC transporter, the observed low PP:WCL ratio for BtuE is correct.
  - As Figure 2 shows, we observed a continuous range of isotope ratio values. If the periplasmic enrichment were perfect, one would expect all periplasmic proteins to have uniformly high PP:WCL isotope ratios. Our results suggest that the enrichment is protein dependent, and that cytoplasmic and membrane proteins can contaminate the isolated periplasmic fraction.

CONCLUSION

- We have presented a quantitative proteomics methodology to assist the characterization of the periplasmic proteome using differential isotopic labeling and mass spectrometry.
- Estimated PP:WCL isotopic ratios determined from protein mixtures of osmotic-shock-based periplasmic enrichments and whole cell lysates assists in differentiating authentic periplasmic proteins from co-enriching membrane proteins and cytosolic proteins resulting from unwanted cellular lysis.
- Known periplasmic proteins, signal peptide-containing hypothetical proteins, and membrane proteins with peripheral periplasmic domains accounted for 80% of the enriched protein identifications in differential isotopically-labeled protein mixtures of periplasmic enrichments and whole cell lysates.
- These data suggest that our methodology is capable of identifying some well-characterized periplasmic proteins. When the PP:WCL isotopic ratio is combined with primary sequence analysis, our methodology may also be used to infer subcellular locations for hypothetical and unknown proteins.
- Although our methodology was able to differentiate some well-characterized periplasmic proteins from co-enriching cytoplasmic contaminants, the continuous range of isotope ratio values suggests that enrichment is protein dependent.
- As the number of completed microbial genome sequences grows, it is likely that the information our methodology yields will prove valuable to the functional annotation of proteins.

REFERENCES

1. Neu HC and Heppel LA *J Biol Chem* 240: 3685-3692. 1965
2. Ferguson SJ "The Periplasm" in *Prokaryotic Structure and Function: A New Perspective*, S Mohan, C Dow and JA Coles, eds. Cambridge University Press, 310-339. 1991
3. Link AJ *et al* *Int J Mass Spec and Ion Processes* 160: 303-316. 1997
4. Link AJ *et al* *Electrophoresis* 18: 1259-1313. 1997
5. Lopez-Campistrous A *et al* *Mol Cell Proteomics* 4: 1205-1209. 2005
6. Corbin RW *et al* *PNAS* 100: 9232-9237. 2003
7. Dunkley, TP *et al* *Mol Cell Proteomics* 3: 1128-1134. 2004
8. Jiang, X.*et al* *Mol Cell Proteomics* 3:441-455. 2004
9. Yates JR III *et al* *Analytical Chemistry* 67: 1426-1436. 1995
10. Blattner FR *et al* *Science* 277: 1453-1474. 1997
11. Tabb DL *et al* *J Proteome Res* 1: 21-26. 2002
12. a. Pan C *et al* "Novel Algorithms for Determination of Peptide Abundance Ratios in Quantitative Proteomics with Stable Isotope Labeling." Oral Presentation. *53rd ASMS Conference on Mass Spectrometry and Allied Topics*. June 5, 2005, San Antonio, TX.
13. b. Pan C *et al* *Anal Chem* (SUBMITTED) 2006
14. Hiller K *et al* *Nucleic Acids Res* 32: W375-379. 2004
15. Bendtsen JD *et al* *J Mol Biol* 340: 783-795. 2004
16. Rey S *et al* *Nucleic Acids Res* 33: D164-168. 2005
17. Rose RW *et al* *Mol Microbiol* 45: 943-950. 2002
18. Boeckmann B *et al* *Comp Rendus Bio* 328: 882-899. 2005
19. Hanson, ND *et al* *Curr Pharm Des* 5:881-894. 1999
20. Groarke, JM *et al* *J Biol Chem* 258: 12952-12956. 1983
21. Borths, EL *et al* *Biochemistry* 44:16301-16309. 2005

ACKNOWLEDGMENTS

- Research sponsored by the ORNL Laboratory Director's Research and Development Program, and the Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.