Proteomic Verification of Signal Peptide Prediction Algorithms in *Rhodopseudomonas palustris*

W. Judson Hervey IV[†‡], Adam M. Tebbe[¶‡], Nathan C. VerBerkmoes[†‡], Manesh B. Shah[¶], Dale A. Pelletier[¶], Frank W. Larimer[¶], and Gregory B. Hurst[†*]

[†]Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory; [¶]Life Sciences Division, Oak Ridge National Laboratory; [‡]Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory

OVERVIEW

- ***Rhodopseudomonas palustris*, a gram-negative alpha proteobacteria, displays robust metabolism. As a result of its metabolic breadth, approximately 15% of its 5.4 Mb genome encodes genes related to transport.**
- **Systems biology examinations of this remarkable bacterial species are underway at ORNL. Ongoing proteomic studies have begun to illuminate the nature of the metabolic versatility of *R. palustris*. A complementary methodology to current studies is the determination of the subcellular localization of the proteome.**
- **Subcellular localization provides another dimension of information regarding the biological function of each member of the proteome. With respect to the proteome of *R. palustris*, enzymes, cofactors, and transport proteins functioning in metabolism are located within the bacterial periplasm, a subcellular compartment.**
- **To characterize this subset of the proteome and provide insight into the function of these proteins in the metabolism of *R. palustris*, a strategy for determining the subcellular localization within prokaryotes using metabolic labeling was devised.**

INTRODUCTION

- **Although relatively simplistic in comparison to eukaryotic organisms, bacteria are nonetheless complex creatures. An example of such complexity is the gram negative cellular membrane. This compartmentalized membrane system consists of the inner membrane, peptidoglycan layer, periplasm, and outer membrane.**
- **In general, the biological function of the membrane system is to provide an environment conducive to life. This task requires flexibility, as the system must provide protection against the harsh external environment, yet permit the flow of essential nutrients both into and out of the cell.**
- **The inner and outer membranes are constructed of phospholipid bilayers to shelter the internal biological processes from the external environment (Figure 1A). The thin peptidoglycan layer in Gram negative organisms also provides a degree of protection.**
- **To permit the transit of essential nutrients and small molecules in and out of the cell, a family of porous proteins is housed into both the inner and outer membrane. Known as integral membrane membrane proteins, these proteins provide safe passage through the membrane system in conjunction with other proteins and the expenditure of adenosine triphosphate (ATP).**
- **Integral membrane proteins, in association with periplasmic binding proteins (PBP), often work cooperatively in transport systems known as ATP Binding Cassette Transporters (ABC Transporters). As substrates diffuse through porins in the outer membrane, PBPs bind them in the periplasm and transport them intracellularly.**
- **By determining the subcellular localization of putative periplasmic proteins, such as ABC transporters, PBPs, protein cofactors, and proteins of unknown function, another dimension of the proteome can contribute to a more comprehensive understanding of the robust metabolism of *R. palustris*.**
- **The translocation of proteins across the inner membrane to the periplasm occurs during protein synthesis (Figure 1A) through the secretory (Sec) mechanism. In order for secretion to occur, an amino terminal signal peptide must be present on the nascent polypeptide (Figure 1B).**

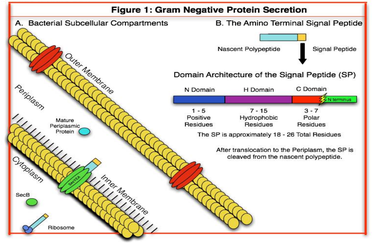


Figure 1: Gram Negative Protein Secretion

- A: Bacterial Subcellular Compartments - Consisting of the Cytoplasm, Inner Membrane, Peptidoglycan Layer, Periplasm, and Outer Membrane, the bacterial membrane system provides flexible protection from the external environment. The secretion of proteins through the inner membrane to the periplasm occurs during protein synthesis at the ribosome, whereby the chaperone SecB identifies a signal peptide within the nascent polypeptide. The polypeptide is then shuttled through the integral membrane protein complex SecYEA into the periplasm, where the signal peptide is cleaved and a mature periplasmic protein is produced.
- B: Amino Terminal Signal Peptide - Translocation across the inner membrane by the Sec mechanism is dependent upon the presence of a signal peptide consisting of three distinct domains: the N Domain, the H Domain, and the C Domain. Each domain displays distinct characteristics of its composition, as shown above. Although this domain structure does not have to be strictly followed for translocation, adherence to this pattern expedites processing by the Sec mechanism. Cleavage of the signal peptide C-terminally to the C Domain occurs in the periplasm by a signal peptidase, resulting in a mature periplasmic protein.

METHODS

Subcellular Localization Strategy using Metabolic Labeling

- Previous analyses of periplasmic proteins by MudPIT revealed contamination of the fraction by cytoplasmic proteins.
- To account for unavoidable residual cellular lysis during the harvest of periplasmic proteins, a strategy was devised to distinguish cytoplasmic proteins from periplasmic proteins (Figure 2). By integrating identifications of peptides within the periplasmic protein fractions with their metabolic enrichment, the subcellular location of the peptides can be determined.

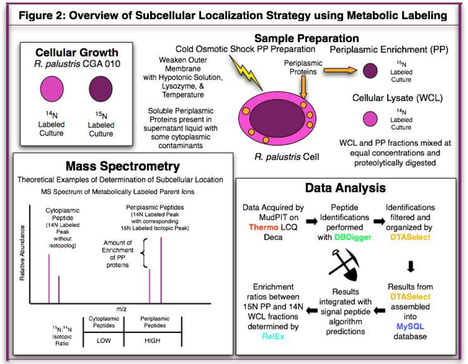
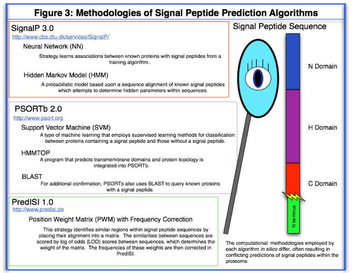


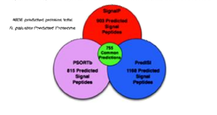
Figure 2: Overview of Subcellular Localization Strategy using Metabolic Labeling

- **Cellular Growth**
  - Cultures of *R. palustris* CGA010 were grown photoheterotrophically with succinate as a carbon source in PMS-10 medium. For metabolic labeling, cultures designated ^{14}N were grown in PMS-10 with no modification to the medium components and those cultures designated ^{15}N were given ^{15}N labeled ammonium sulfate (Isotec) as the sole source of nitrogen in culture.
- **Sample Preparation**
  - Following growth to a similar optical density (OD660), ^{14}N and ^{15}N cultures were harvested by centrifugation.
    - Proteins were quantified by the BCA Assay (Pierce). Equal protein concentrations from each protein fraction were combined and proteolytically digested with Trypsin (Promega).
- **Liquid Chromatography and Mass Spectrometry**
  - For liquid chromatography, two-dimensional MudPIT (Washburn *et al*) was used (reverse phase and strong cation exchange over 12 hours). MudPIT columns were placed in line to the mass spectrometer in front of an LC Packings Ultimate HPLC pump.
  - Mass spectral data were acquired by ThermoFinnigan LCQ Deca and Deca XP instruments.
- **Data Analysis**
  - Following the acquisition of mass spectra, charged states were assigned to peptides using MS2ZAssign (Tabb *et al*).
  - Peptide identifications from tandem mass spectra were performed by DBDigger (Tabb *et al*) using the annotated *R. palustris* proteome (Larimer *et al*). Resulting peptide identifications were filtered and organized with DTASelect (Tabb *et al*) and relevant information stored within a MySQL database.
  - Signal peptide predictions from the *R. palustris* proteome were performed with SignalP (Bendtsen *et al*), PredISI (Hiller *et al*), and PSORTb (Gardy et al). These predictions were integrated within the MySQL database for confirmation of subcellular location by the strategy.
  - Enrichment ratios between the ^{14}N cellular lysate and the ^{15}N periplasmic enrichment were determined by RelEx (MacCoss *et al*).

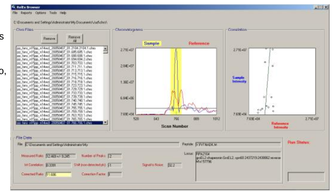
- **Data Analysis (continued)**
  - **Signal Peptide Prediction Algorihms**
    - To confirm the peptide identifications from the subcellular localization strategy, the predicted *R. palustris* proteome based upon an *in silico* translation of the annotated genome (Larimer *et al*) was searched with three different signal peptide prediction algorithms: SignalP (HMM and NN), PSORTb, and PredISI.
    - These three algorithms employ different algorithmic strategies in the prediction of signal peptides (Figure 3, below).
    - For classification as periplasmic according to the subcellular localization strategy, MudPIT protein identifications must contain a predicted signal peptide by all three algorithms.



    - The predicted *R. palustris* proteome based upon an *in silico* translation of the completed genome (Larimer *et al*) was submitted to each of the three signal peptide prediction algorithms (Figure 4, right).
    - PredISI predicted 1,168 signal peptides within the proteome.
    - PSORTb predicted 815 signal peptides within the proteome.
    - SignalP predicted 903 signal peptides within the proteome.
    - Among the three tools, 755 predictions were similar.
    - The predictions of each of the three tools were stored within the MySQL database along with protein identifications from the proteomics experiments.



- **Calculation of Isotopic Ratios**
  - The software tool RelEx was used to determine the isotopic ratio of ^{15}N-labeled periplasmic proteins to the ^{14}N whole cell lysate protein fraction.
  - Chromatographic data was extracted from proteomic data using the ExtractChro utility within RelEx. The ExtractChro utility parses the DTASelect-filter.txt file from DTASelect output and extracts the corresponding chromatograms from Xcalibur Raw files. The resulting chromatographic files were then filtered by isotopic ratio, standard deviation, and number of peptides per protein (>0.01, +/- 1 SD, and 2, respectively) using the RelEx browser.

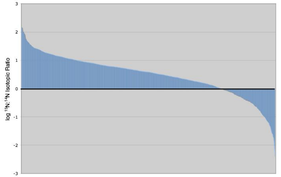


    - **Figure 6:** Chromatographic data for GroEL, a chaperonin protein, is shown to the right. The calculated ^{15}N:^{14}N ratio of 11.686 is representative of a cytoplasmic protein.

RESULTS

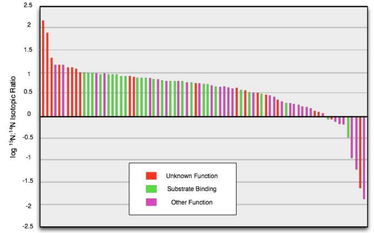
- **Identification of Proteins from the Periplasmic Subproteome**
  - Following collection of tandem mass spectra and database searching with the DBDigger search algorithm, a total of 84 proteins from 1,508 total protein identifications were identified as periplasmic according to our strategy, based upon the following conservative criteria:
    - Identification of at least two peptides by the DBDigger search algorithm with scores above 23.76, 28.07, and 43.64 for +1, +2, and +3 charged states, respectively;
    - A score of ≥ 0.50 from the PredISI signal peptide search algorithm;
    - A score of ≥ 0.50 from the SignalP NN signal peptide search algorithm;
    - A score of ≥ 0.75 from the SignalP HMM signal peptide search algorithm;
    - A score of ≥ 0.5 from the PSORTb signal peptide search algorithm.
  - Of the 84 proteins fulfilling these criteria, all proteins were either periplasmic binding proteins (PBPs), ABC transporters, metabolic enzymes, transmembrane proteins, or uncharacterized (hypothetical, unknown, or putative proteins).
  - **Interestingly, the 84 protein identifications determined as periplasmic by our strategy contained at least one ^{15}N labeled peptide and one ^{14}N labeled peptide.**
  - After extraction of chromatographic data with RelEx, protein identifications were filtered as described in the Methods section. The logarithmic values of the ^{15}N:^{14}N isotopic ratios of the 1,463 total MudPIT protein identifications from *R. palustris* are shown below (Figure 6).

Figure 6: Isotopic Ratios of Proteins Identified by MudPIT



  - **Of the 84 proteins identified as periplasmic by our strategy, 79 were found to contain ^{15}N:^{14}N isotopic ratios by RelEx (Figure 7).**

Figure 7: Isotopic Ratios of 79 Proteins with Predicted Signal Peptides



  - 69 of the 79 periplasmic proteins displayed a positive value for the logarithm of the ^{15}N:^{14}N isotopic ratio, which illustrates enrichment of the 15N-labeled periplasmic protein fraction.
  - As shown above, the annotated functions of the 79 proteins represent the functions of known periplasmic proteins. 27 of the protein identifications are annotated as substrate binding proteins, which are known to perform their biological function within the periplasm.
  - In addition, the strategy provides preliminary evidence for the function of 32 proteins of unknown function within the periplasm.

CONCLUSIONS

- **The subcellular localization of the proteome provides yet another dimension of information when integrated with other systems-wide datasets, such as shotgun proteomics data.**
- Despite endeavors in higher biological systems, a comprehensive methodology for determining subcellular localization in prokaryotes has escaped development.
- To overcome this deficiency in the field, a strategy was devised using the metabolically robust bacteria *R. palustris*, whose proteome contains an abundance of transport proteins.
- **Through the integration of MudPIT protein identifications, the isotopic ratios of their respective peptides, and signal peptide predictions by three algorithms, 79 proteins were identified as periplasmic proteins from 1,463 total MudPIT identifications.**
- **Proteins functioning in substrate binding, transport, and metabolism were found among the 79 identifications.**
- The metabolically labeled ^{15}N periplasmic fraction contained an enrichment of periplasmic proteins.
- **The comprehensive strategy provided preliminary evidence into the function of unknown proteins within the periplasm.**
- **The increasing number of completed bacterial genomes will ensure that this strategy becomes more effective in the future.**

REFERENCES

- Hiller K et al *Nucleic Acid Res* **32**:W375-9
- Bendtsen JD et al *J Mol Biol* **340**:783-95
- Gardy JL et al *Bioinformatics* **21**:617-23
- Washburn MP et al *Nat Biotechnol* **19**:242-7
- MacCoss MJ et al *Anal Chem* **75**:6912-21
- Tabb DL et al *J Proteome Res* **1**:21-6
- Tabb DL et al *Anal Chem* **77**:2464-74
- Link et al *Int J Mass Spec and Ion Processes* **160**:303-316
- Larimer et al Nat Biotechnol **22**: 55-61

ACKNOWLEDGMENTS

- Jennifer L. Gardy, Simon Frasier University, Burnaby, BC
- WJH gratefully acknowledges the following members of the OBMS group for their assistance and invaluable discussions: David L. Tabb, W. Hayes McDonald, M. Brad Strader, Chongle Pan, Alon Savidor, Melissa R. Thompson, Heather M. Connelly, Keiji G. Asano, Patricia K. Lankford, Becky Maggard and Christine T. Shook.
- WJH and AMT acknowledge the UT-ORNL Graduate School of Genome Science and Technology.
- Research sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.