Two-Dimensional Liquid Chromatography Combined with High-Resolution ES-FTICR-MS for Top-Down Characterization of Bacterial Proteomes*

Robert Hettich[1], Kyle Ellrott[2], Nathan VerBerkmoes[1,2], Greg Hurst[1], Patricia Lankford[3], Dale Pelletier[3], and Frank Larimer[3]

[1] Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN
[2] Genome Science and Technology Graduate School, ORNL-UTK, Knoxville, TN
[3] Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Introduction

The characterization of bacterial proteomes provides detailed insight into the functional state and activities of these simple organisms. Such information enables an understanding of how such organisms might play a role in carbon sequestration or bioremediation applications. We have undertaken such a study for the organism *Rhodopseudomonas palustris*, a phototrophic bacterium that has a high metabolic diversity and is able to degrade aromatic hydrocarbons such as lignin. In order to enhance protein detection and identification, we have designed an experimental protocol that employs two-dimensional liquid chromatography (LC) combined with high-resolution Fourier transform mass spectrometry (FTMS). This top-down experimental approach provides high-resolution and sequence information about the molecular nature of the intact proteins, which can be compared with cataloging information from the bottom-up technique to yield comprehensive proteomic information.

Methods

R. palustris was grown to mid-log phase under photoheterotrophic conditions. Cells were harvested, washed twice with buffer, and disrupted with ultrasonication. Unbroken cells were removed with low speed centrifugation and four crude protein fractions were created by ultracentrifugation (100,000g for 1 hour). The first stage of chromatography was off-line anion exchange (tris, pH 8, 0-1M NaCl gradient) fractionation of the intact proteins into ~30 simplified mixtures. Each of these fractions was then examined with on-line reverse phase liquid chromatography (Vydac C4 capillary column) directly into an FTICR-MS instrument (IonSpec HiRes electrospray FTICR-MS, with 9.4T magnet), where high-resolution and accurate mass measurements were conducted on the intact proteins. Bioinformatic algorithms were developed to query the protein database with the intact protein information from either annotated or unannotated genome information. These results were compared with proteins identified by bottom-up technique, in order to search for and identify protein post-translational modifications.

Results

Each of the ~20 anion-exchange fractions was examined by capillary, reverse phase LC-FTICRMS. Figure 1 illustrates the reverse phase chromatogram from FPLC fraction #4. Even though the chromatographic resolution is somewhat limited, this method provided enough spatial resolution to enhance the mass spectrometry measurements. Figure 2 shows the mass spectrum obtained for the peak at 20.1 min. Note the appearance of at least six unique proteins, each one measured under high resolution conditions. By querying the high resolution molecular masses against the tabulated protein database (which was generated from the genome sequence), it was possible to identify several of these species. The appearance of abundant proteins, such as the ribosomal species, is not surprising; however, this mass spectrum also contains three abundant hypothetical species. Two of the hypothetical proteins (ORFs 6829 and 1420) correlate quite well with the expected gene products, whereas the third conserved hypothetical (ORF 1314) is N-terminally truncated.

The multi-dimensional LC-FTICR-MS experiment provided the ability to measure more than 250 intact proteins for *R. palustris*, although some of these were measured redundantly in separate fractions. In total, about 200 discrete proteins have been measured at high resolution. Not only are highly abundant proteins such as the ribosomal species identified, but information is also obtained for several of the other functional categories, including several hypothetical proteins. For example, two different proteins, each corresponding to a distinct GroES species, were measured and characterized. It is somewhat unexpected that *R. palustris* would express both versions in a single growth condition. One GroES

species was observed to correspond to the predicted gene product, whereas the other species was N-terminal methionine truncated. This represents the power of the measurement of intact proteins by the top-down MS technique, in that protein post-translational processing can be identified and characterized. The top-down MS technique is especially sensitive to measurement of the smaller proteins. This is vital experimental information, as many of the gene-calling programs are not very accurate in their prediction of small ORFs. Not surprising, about one-half of the proteins measured could not be matched directly to the predicted gene products. This suggests a substantial amount of post-translational processing in the expressed proteins. The Initial work was focused on examining this data for common alterations such as N-terminal methionine truncation and signal peptide truncation.

A bioinformatic algorithm has been developed to compare the experimental data with the predicted gene products to search for several known protein alterations. The availability of high resolution data for the intact proteins is essential for comparison to the bottom-up results for locating these important protein modifications. A particular goal of this approach is to identify signal peptides from hypothetical proteins, thereby providing important information that can be used to unravel their functions. An alternate bioinformatics approach has been developed to attempt to utilize the measured protein masses to identify proteins from the unannotated genomic information. This genome scanning technique, first developed by Michael Giddins for peptides (**ref???**), searches the translated genome with the high resolution mass spectrometry data. Once a putative match is found, it can be confirmed with tandem mass spectrometry techniques. This approach does not rely on gene-calling algorithms, which may have some errors in the gene assignments. One additional advantage is the ability to readily identify truncations, including the identification of signal peptides. We have tested the algorithm with several known and predicted proteins, to verify the accuracy of the identifications. At present, we are beginning to screen extensive experimental data, in an effort to verify (or correct) gene start sites and truncation products.

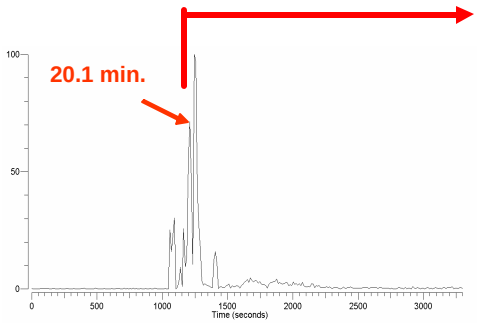

Figure 1. LC-FTICR-MS of anion-exchange fraction #4 for *R. palustris*.

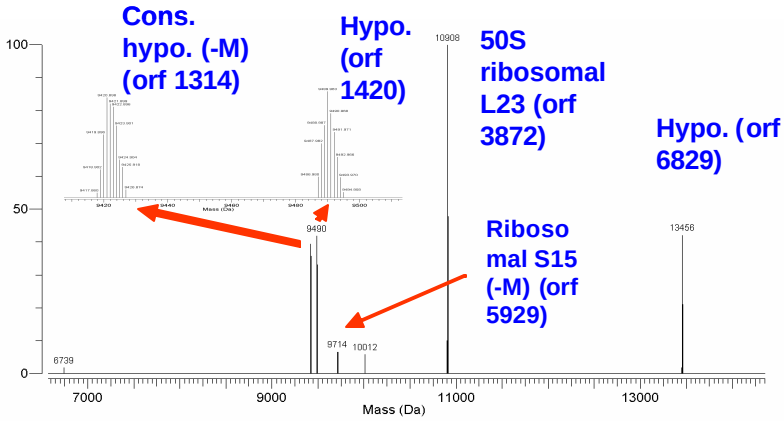

Figure 2. High resolution FTICR-mass spectrum of the peak at 20.1 min. from Figure 1. Each of the molecular species could be measured under high resolution conditions, as shown in the insets for two hypothetical proteins.

* Research sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research. Oak Ridge National Laboratory is operated and managed by the University of Tennessee-Battelle, L.L.C., for the U.S. Department of Energy under contract DE-AC05-00OR22725.