A Novel Approach to the Proteomic Analysis of an Organism with an Unsequenced Genome

by

Shreya Chand

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the degree of

DOCTOR OF PHILOSOPHY

Major Subject: Chemistry

Approved by the
Examining Committee:

_____
Mark D. Platt, Thesis Adviser

_____
Linda B. McGown, Member

_____
Jane F. Koretz, Member

_____
Lee A. Ligon, Member

_____
Janet L. Paluh, Member

Rensselaer Polytechnic Institute
Troy, New York

April, 2012
(For Graduation May 2012)

ABSTRACT

The field of proteomics, with the stated goal of "identifying all of the proteins expressed by a genome" is an attempt to link gene expression and cellular function. While this is relatively straightforward for an organism with a sequenced genome, it becomes much more challenging when the proteins present in the sample do not exist in a well-annotated database. In the series of experiments detailed here, various aspects of the proteomics process have been modified and incorporated into an overall analytical approach which permits the sequencing of proteins from the important, unsequenced model organism *Spisula solidissima*.

In the first series of experiments, oocytes were obtained from *S. solidissima* and soluble proteins were extracted. Two-dimensional gel electrophoresis was used to resolve proteins from the whole cell lysate and in-gel digestion of the protein spots was carried out using different proteolytic enzymes. Peptide sequencing was accomplished using nano-flow reversed phase high performance liquid chromatography microelectrospray ionization mass spectrometry (nHPLC-μESI-MS) and a novel *de novo* sequencing algorithm referred to as the Comprehensive Peptide Database (CPDB) Search. Overlapping peptides were then aligned and the resulting amino acid sequences were compared against the non-redundant (nr) protein database at the National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool (BLAST) in order to identify known proteins having significant homology.

Following successful use of the procedure for the proteomic analysis of *S. solidissima* oocytes, the second set of experiments involved the identification of novel proteins from the centrosome, a valuable organelle for the study of the cell cycle. Using sucrose density gradient centrifugation, centrosomes were isolated from *S. solidissima* oocytes and their proteins were then separated using one-dimensional gel electrophoresis. In-gel digestion was carried out using different proteolytic enzymes and an isotopic labeling strategy was employed at the peptide level to aid in CPDB searching. Additionally, centrosomes were treated with potassium iodide in order to extract the salt insoluble protein matrix which is a part of the pericentriolar material (PCM). These salt insoluble PCM proteins were separated using one-dimensional gel electrophoresis and novel proteins were identified as mentioned above. These analyses

resulted in the identification of several novel proteins from *Spisula solidissima*, and represent the first unbiased proteomic analysis of the organism. Importantly, the methodology described here is applicable to the identification of proteins from any organism with an unsequenced genome.