# NEW ANALYTICAL TOOLS FOR SYSTEMS BIOLOGY

**By**

**XIAOTING TANG**

**A dissertation submitted in partial fulfillment of**
**the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

**WASHINGTON STATE UNIVERSITY**
**Department of Chemistry**

**December 2006**

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of XIAOTING TANG find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

_____

## ACKNOWLEDGMENT

This work was accomplished through support of many individuals, in particular, my advisors Dr. James E. Bruce and Dr. Herbert H. Hill.

First of all, I would like to express my sincere gratitude to my committee chair and advisor, Dr. Herbert H. Hill, for his continuous guidance, generous support, and scientific direction. He also deserves particular credit for creating an opportunity for me to conduct both biological mass spectrometry and ion mobility spectrometry instrumentation research during my Ph.D. studies, which certainly helped strengthen and diversify my scientific expertise. I would also like to thank my committee members, Dr. John J. Wyrick, Dr. Alexander DeQuan Li, Dr. William F. Siems, and my former committee member, Dr. Ralph G. Yount, for their valuable suggestions and advice throughout my graduate studies. The bioinformatics knowledge I learned from Dr. Wyrick was beneficial and essential to my research. In particular, I would also like to extend my gratitude to all the past and present members in both the Bruce group and Hill group.

I certainly owe a huge debt of gratitude to Dr. Gerhard R. Munske, who synthesized all the cross-linker compounds for this work. Every discussion with him has been inspiring and fruitful. His insightful thoughts and creative ideas helped promote the quality of this work. My thanks also go to our collaborators from Pacific Northwest National Laboratory, Nikola Tolic and Gordon A. Anderson, who provided enormous and critical informatics support for this research. I would also like to thank Dr. Steve Van Orden and

Dr. Christian Berg for their timely and generous help for troubleshooting the FTICR-MS instrument. In addition, I am grateful for the tremendous and invaluable support from Technical Services in WSU. Fred Schuetze helped build various electronics devices and George Henry helped fabricate all the components for the IMS project. None of this work would have been possible without their support.

I would also like to express my most heartfelt appreciation to my parents, for their unconditional love and encouragement, which have always been the driving force for fulfilling my dream. Though nothing I can do will ever match what they gave me, hopefully this achievement will be one of the best gifts I could ever give them. I am also grateful to my mom-in-law and my grandmas for their precious love and support.

Finally, my deepest acknowledgment goes to my husband, James E. Bruce, an outstanding scientist with exceptional creativity, insight and enthusiasm that I always admire. His love, encouragement, and companionship have made this journey worthwhile and memorable.

# NEW ANALYTICAL TOOLS FOR SYSTEMS BIOLOGY

## Abstract

By Xiaoting Tang, Ph.D.
Washington State University
December 2006

Chair: Herbert H. Hill Jr.

Systems biology imposes extremely challenging tasks for researchers because of complexity and dynamics of biological systems. In this research, we aim to tackle biological problems with the development of novel analytical tools including both chemical strategies and instrumental techniques.

Profiling membrane proteome provides critical information for both fundamental biology and applied biomedical research. However, membrane proteins are hard to study due to hydrophobicity and low abundance. We introduced a family of new cell-permeable chemical probes with biotin affinity tag and reactive groups for extensive labeling and detection of membrane proteins from intact cells. These hydrophobic chemical probes allowed close interactions with membranes and enabled identification of a substantial proportion of membrane proteins from total cell lysate without the need for specific membrane isolation. This research demonstrated the first extensive study of membrane proteome with labeling intact cells and revealed numerous proteins that are involved in metal reduction process in *Shewanella oneidensis*.

Current chemical cross-linking methods are commonly employed for mapping sites of interaction in purified, known protein complexes. When applied *in vivo*, the sites of interaction are usually unattainable. We employed novel cross-linkers that we call protein interaction reporters (PIRs) that have unique features to determine protein-protein interactions and surface topology concurrently. The utility of PIR strategy was successfully demonstrated with a known protein complex systems, ribonuclease S. Furthermore, this strategy was applied to *Shewanella oneidensis* bacteria and initial results demonstrate for the first time simultaneous identification of protein-protein interactions and sites of interaction *in vivo*.

Ion mobility spectrometry (IMS) has attracted growing interest in proteomics research because of its unique features of separating gas-phase ions in milliseconds and separating isomeric compounds. Atmospheric pressure IMS (AP-IMS) can be readily coupled to various mass spectrometers without any modification. Our study demonstrated AP-IMS can also be utilized to evaluate ionization efficiency for various atmospheric ionization sources under various conditions due to its unique constant ion transfer property. We constructed the first hybrid instrument that hyphenated AP-IMS with FTICR-MS with the use of dual gates and a flared inlet capillary interface and provided the capability of separating isomeric peptides for FTICR mass spectrometers.

# TABLE OF CONTENTS

# LIST OF TABLES

**Chapter 3: Tables**



**Chapter 4: Tables**



**Chapter 6: Tables**

# LIST OF FIGURES

**Chapter 2: Figures**



**Chapter 3: Figures**

**Chapter 4: Figures**

**Chapter 5: Figures**

**Chapter 6: Figures**

# DEDICATION

**This dissertation is dedicated to:**

**my father Xianwei Tang (汤显伟),**

**my mother Guifeng Wang (王桂凤),**

**and my husband James E. Bruce.**

# CHAPTER 1

# INTRODUCTION

## Research Objectives

Systems-level study of an organism has become essential to understand fundamental biological functions and processes. New emerging technologies have initiated and enabled large-scale genomics and proteomics research. However, the dynamically changing and highly complex biological systems are far from being completely understood with current techniques. The goal of the present research was to develop and apply chemical and instrumental methodologies and strategies for tackling biological problems with specific focus in protein-protein interactions. *Shewanella oneidensis* strain MR-1, a Gram-negative facultatively anaerobic bacterium, was employed as model systems.

Toward this goal, this research is envisioned with two primary research objectives: (i) development and application of novel chemical cross-linking strategies combined with shotgun LC/MS/MS and high performance FTICR-MS for detection and identification of protein-protein interactions of *S. oneidensis* MR-1 on proteome-wide scale; and (ii) development and application of a hybrid instrument that combines ion mobility spectrometer (IMS) with FTICR-MS to characterize biological mixtures, in particular, isomeric peptides.

**Systems Biology**

Biological molecules do not function individually, but mostly participate in complex assemblies and interaction networks to form the working of cells, tissues, organs, and organisms.[1] The interplay and interactions of all components of an organism give rise to its forms and functions. Systems biology[2-3] is the study of genes, proteins, metabolites, and their interaction networks of an organism at integrated level instead of analyzing individual components or aspects of the organism. Systems-level studies of an organism are keys to fundamental understanding of biological functions. Traditional biological research concentrates on a particular gene or protein – study of its interactions, its regulations, and its pathways. However, as opposed to conventional biology, systems biology demands high-throughput and universal methods and techniques that can take a global view of all interacting components at a time. Thus, systems biology imposes an extremely challenging task for researchers because of the complexity of biological systems. Numerous new technologies have been emerged and developed to study biology in a new way - at systems level during the past decade. For example, DNA microarray technology provides high throughput screening of thousands of genes simultaneously in a single array.[4] 2D-gel methods,[5] in particular 2D-DIGE[6,7] that uses fluorescent dye labeling, have been the work horse for quantitative profiling over thousand proteins at proteome-wide level. Alternatively, 2D-LC/MS[8-10] technologies in conjunction with stable isotope labeling such as ICAT,[11,12] or SILAC[13,14] have provided an efficient way for large-scale profiling complex protein mixtures. Furthermore, analytical methods such as GC/MS[15,16], LC/MS[17,18], CE/MS[19-21], NMR[22,23] have been

extensively applied for the metabolomics study. Although study of the expression levels of genes, proteins, or metabolites at a particular time and state provides important clue for investigating functions and processes of biological systems, determination of the interacting networks is more critical for understanding how biological systems works. Thus, our research interest is to develop and apply novel analytical tools to study proteins and protein-protein interactions at systems-level.

### ***Shewanella oneidensisis*** **MR-1**

The model systems we chose for this research is *Shewanella oneidensis* MR-1. *S. oneidensis*[24] is a metabolically versatile bacterium that can grow both aerobically and anaerobically. In absence of oxygen, *S. oneidensis* can respire on a diversity of other electron acceptors - fumurate, nitrate, nitrile, thiosulfate, sulfur, iron, uranium, manganese, and chromium, to obtain energy for its growth and survival.[25-26] Many of these compounds and metals that *S. oneidensis* can use for energy and growth are toxic to humans and other organisms when concentrated in environment. *S. oneidensis* can convert a variety of toxic metal ions in solution to their precipitate forms thus prevent from spreading. This metal-reducing ability makes *S. oneidensis* an important bacterium organism for potential bioremediation. The diverse respiratory capabilities of *S. oneidensis* are resultant from its complex multi-component branched electron transport system composed of cytochromes, reductases, iron-sulfur proteins and quinones.[25,27] Little is known about the mechanisms of participating in these multicomponent processes. Numerous efforts have been actively involved in profiling gene expression

and protein expression of *S. oneidensis* on global scale using DNA microarrays,[28-30] 2D gels in combination with mass spectrometry,[31-33] liquid chromatography with FTICR.[34-35] However, no systems-level studies of protein interaction networks of *S. oneidensis* to help better understanding of its versatile electron transport processes have yet been reported.

**Protein-Protein Interaction and Cross-linking Method**

Most proteins function through protein complex assemblies. Bruce Alberts, the president of National Academy of Sciences, states that "nearly every major process in a cell is carried out by assemblies of 10 or more protein molecules. And, as it carries out its biological functions, each of these assemblies interacts with several other large complexes of proteins. Indeed, the entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines".[36] The ability to measure protein-protein interactions in biological systems has undergone significant advances in last decade due to emerge and growth of new molecular biology and mass spectrometry technologies. Yeast two-hybrid,[37-38] fluorescence resonant energy transfer (FRET),[39,40] and co-immunoprecipitation are genetic approaches for detection and characterization of protein-protein interactions. A key feature of all these approaches is that certain prior knowledge of the interacting systems is a prerequisite so that a bait sequence can be chosen for fishing interacting partners.

Chemical cross-linking in combination with mass spectrometric techniques is a proteomic approach for studying protein-protein interactions[41-44] and topological structures of protein complexes.[45-46] This approach works by first covalently linking the interacting partners with two reactive groups of a chemical cross-linker, followed by a series of well-established proteomics protocols - 1D or 2D gels, in-gel digestion, data dependent LC/MS/MS, and database search, for identification of the cross-linked proteins. It is believed that the individual components in a protein complex assembly stay in close proximity; that is, within the reach of the two reactive groups of a cross-linker. Thus chemical cross-linking reactions are considered specific for linking proteins which are only in close proximity. Cross-linking strategies provide a number of advantages. (i) no prior knowledge on the biological systems is required; (ii) it works for both targeted and whole picture study; (iii) covalently cross-linked complexes are stable and tolerate subsequent harsh handling conditions; (iv) in addition to identifying what proteins interact within a complex, cross-linking approaches can also pinpoint where proteins interact; (v) in-vivo cross-linking is feasible (vi) rapid cross-linking reactions can snap-shot temporal and spatial interactions. However, in spite of these advantages and strong interest in profiling protein-protein interactions, only a limited number of reports have illustrated successful application of cross-linkers to study protein complexes and the publications of cross-linking strategies for profiling protein interactions on global level are virtually not existent. Several factors of cross-linking strategy have restricted its successful applications on a wider scale. First most chemical cross-linking reactions have low yield and large number of undesirable side products. In addition to 1:1 linkage of the interacting proteins, so called inter-cross-link, cross-linking reactions can result in

another two types of products, dead-end labeling and intra-cross-link. Dead-end labeling sometimes provides useful information in term of studying protein surface chemistry. Intra-cross-linking has been extensively employed for determining 3-D structure of a protein since the length of the spacer chain defines the through space of two cross-linked amino acid residues. Due to the low yield of cross-linking reactions, a large excessive amount of cross-linkers are often used thus resulting in multiple labeling on one sequence. Furthermore, as commonly used in current proteomics protocol, an enzymatic digestion of cross-linked complex is performed prior to analysis by mass spectrometry in order to identify interacting proteins and interacting regions of proteins. Proteolysis further increases the complexity of the situation because the digestion mixtures include predominately unmodified peptides and various types of singly and multiply modified peptides, and non-specific modified peptides. And lastly, MS/MS fragmentation spectra of an intra- or inter-cross-linked species are more complicated than MS/MS of simply modified sequences; much fragmentation can not be interpreted by normal MS/MS nomenclature since two residues are modified through one bridge. To take advantage of cross-linking strategy and overcome the shortcomings of this approach, a novel and unique type of cross-linkers that incorporates specific mass spectrometry cleavable bonds, mass-coded tag, and affinity tag will be developed in this research. A two-stage mass spectrometric strategy capitalizing the novel cross-linking strategy will be developed as well to achieve profiling protein interactions at systems-level.

**Ion Mobility Spectrometry (IMS) and FTICR-MS**

Ion mobility spectrometry (IMS), in principle, is an instrument that measures mobility of ions in gas phase. Ion mobility measurements are performed in a constant electric field formed in a drift tube that contains the inert buffer gas. The drift tube is composed of a series of stacked ring electrodes which provide uniform electric field. Ions are accelerated by electric field until they collide with the neutral molecules and get decelerated. In low field regime used for IMS, the continuous acceleration and deceleration result in a constant drift velocity, $v_d$, which is directly proportional to the electric field, $E$. The ratio of the drift velocity and electric field is defined as the ion mobility, $K$.[47]

$$K = \frac{v_d}{E} = \frac{L^2}{t_d V} \tag{1}$$

Where L is the length of the drift tube, $t_d$ is the drift time, and $V$ is the voltage drop across the drift tube. The ion mobility is usually expressed as the reduced mobility, $K_0$, which is normalized to standard conditions of temperature $T$ (Kelvin) and pressure $P$ (Torr).

$$K_0 = K \frac{273}{T} \frac{P}{760} \tag{2}$$

Ion mobility is directly proportional to the charge of an ion, $q$, and reversely proportional to the collision cross section of an ion, $\Omega$. Therefore IMS separates ions based on $\Omega/q$ rather than $m/z$ that is used by MS. Collision cross section can be calculated using Mason-Schamp equation[47,48]

$$\Omega = \frac{3q}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{1/2} K \tag{3}$$

Where N is the number density of the drift gas at standard conditions, $\mu$ ($\mu = mM/(m+M)$) is the reduced mass of an ion ($m$) and the neutral drift gas ($M$), and $k_B$ is Boltzmann's constant. The $\Omega/q$ rather than $m/z$ separation mechanism of IMS makes it possible to separate isomeric compounds and conformers that have different geometries.

Revercomb and Mason have defined the resolving power of IMS in 1974.[47]

$$R = \frac{t_d}{\Delta t} = \left( \frac{LEze}{16 k_B T \ln 2} \right)^{1/2} \qquad (4)$$

Where $R$ is resolving power, $ze$ is the charge on the ion. According to this equation, to achieve higher resolving power it is necessary to increase the length of the drift tube, increase the electric field, and decrease temperature.

IMS has been extensively used to detect drugs,[49-52] chemical warfare agents,[53-56] explosives,[57-59] environmental pollutants,[60,61] amino acids,[62-65] peptides,[66-73] proteins,[74-79] oligonucleotides,[80,81] and carbohydrates.[82,83] IMS has higher resolutions than either GC or LC and separation can be achieved in millisecond scale. A variety of atmospheric ionization sources such as $^{63}$Ni, corona discharge, ESI, and atmospheric MALDI have been applied to IMS. Because IMS separates ions based on size to charge ratio as contrast to mass spectrometer that separates ion based on mass to charge ratio, hyphenation of these two orthogonal separation mechanisms can result in additional gains in resolving powers. Low pressure (1-10 Torr) or atmospheric pressure IMS has been coupled with a variety of mass spectrometer[84] such as quadrupole,[49,85-88] TOF,[55,67,74,89] ion

trap,[78,90-92] and FTICR.[93] However, the combination of atmospheric pressure IMS with FTICR has not been reported.

FTICR-MS is a high performance instrument with superior analytical figures of merit.[94,95] FTICR-MS has the highest resolving power of all mass analyzers (over 1,000,000) and mass accuracy (less than 1 ppm). It is capable of analyzing highly complex mixture directly without any prior separation. The use of an external quadrupole prior to ICR cell to selectively accumulate ions has further enhanced the sensitivity and dynamic range of FTICR-MS.[96] FTICR-MS is a mass analyzer that is capable of various types of MS/MS and $MS^n$. The collisional induced dissociation (CID) can take place both in the ICR cell and the external quadrupole. Infrared multiphoton dissociation (IRMPD) and electron capture dissociation (ECD) can yield different fragmentation patterns which are complementary to CID. Thus we believe union of IMS with FTICR-MS will provide unparallel instrumental capabilities for analysis of complex mixtures.

**Attributions**

Chapters 2-6 were written based on the format requirement for publications in each respective journal. The work described in Chapter 2 was published in the journal *Analytical Chemistry* (Xiaoting Tang, Gerhard R Munske, William F Siems, and James E Bruce, *Anal Chem*, **2005** 77(1):311-318). Tang designed and performed all the cross-linking labeling and LC/MS/MS experiments and wrote the paper. Munske synthesized the cross-linker compounds and provided valuable ideas and advice on cross-linker

design and labeling experiments. Siems provided critical suggestions for this paper. This project was supported both scientifically and financially by Bruce who contributed to the novel ideas presented in this paper and provided overall scientific direction throughout the project.

The work for Chapter 3 was written by Tang in the format required by *Journal of Proteome Research* (Xiaoting Tang, Wei Yi, Devi P Adhikari, Gerhard R Munske, and James E Bruce, *Journal of Proteome Research*, submitted, June 2006). Tang participated in the initial method development for the labeling experiments and conducted all the 2D-LC/MS/MS experiments, data analysis, and protein function characterization. Yi performed the electron microscopy experiments and Western blot analysis. Adhikari conducted the cell culture and labeling experiments. Munske made all three cross-linker compounds used in this research. Bruce was the advisor for this project and provided scientific direction and funding support for all aspects of this research.

The research work described in Chapter 4 was written in the format required by the journal *Nature Chemical Biology* (Xiaoting Tang, Wei Yi, Gerhard R Munske, Nikola Tolic, Natalia L Zakharova, Devi P Adhikari, Gordon A Anderson, and James E Bruce, *Nature Chemical Biology*, submitted, July 2006). Tang performed all the multiplexed LC/FTICR experiments, 2D-LC/MS/MS experiments, data analysis for both stage 1 and stage 2, and wrote the manuscript. Tang also worked with Adhikari for optimizing the labeling experiment protocols. Yi conducted all molecular biology experiments such as immuno-gold EM experiments and anti-SecA Western Blot experiments. Munske made

the cross-linker compound for this project. Adhikari, Yi, and Zakharov involved in cell culture, labeling, and sample preparation for various stages of the project. Tolic and Anderson from Pacific Northwest National Laboratory (PNNL) are the collaborators who provided all the informatics support for this research. Bruce was the overall project leader who conceived the novel ideas for this research and provided advice and directions throughout.

The research work described in Chapter 5 was written by Tang and submitted to the journal *Analytical Chemistry* (Xiaoting Tang, James E Bruce, and Herbert H Hill Jr, *Anal Chem*, submitted, July 2006). Hill was the project advisor who provided valuable scientific suggestions and insight throughout the project including the instrument design, problem solving, data interpretation, etc. Bruce provided valuable suggestions on experimental design and data analysis. Tang conducted all the experiments for this research and put together the manuscript.

Chapter 6 was prepared according to the format necessary for publication in the journal *Rapid Communications in Mass Spectrometry* (Xiaoting Tang, James E Bruce, and Herbert H Hill Jr, *Rapid Comm Mass Spectrom*, to be submitted in Aug 2006). All the experiments and data analysis were performed by Tang. Bruce helped configure the interface between IMS and FTICR-MS, troubleshoot the instruments, and provide numerous valuable suggestions. Hill ensured the project progress and provided consistent and valuable scientific advice throughout.

## References


1. Collins, F.S., Green, E.D., Guttmacher, A.E. & Guyer, M.S. A vision for the future of genomics research. *Nature* **422**, 835-47 (2003).
2. Aderem, A. Systems biology: its practice and challenges. *Cell* **121**, 511-3 (2005).
3. Kirschner, M.W. The meaning of systems biology. *Cell* **121**, 503-4 (2005).
4. DeRisi, J.L. & Iyer, V.R. Genomics and array technology. *Curr Opin Oncol* **11**, 76-9 (1999).
5. Fey, S.J. & Larsen, P.M. 2D or not 2D. *Current Opinion in Chemical Biology* **5**, 26-33 (2001).
6. Unlu, M., Morgan, M.E. & Minden, J.S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-7 (1997).
7. Swatton, J.E., Prabakaran, S., Karp, N.A., Lilley, K.S. & Bahn, S. Protein profiling of human postmortem brain using 2-dimensional fluorescence difference gel electrophoresis (2-D DIGE). *Mol Psychiatry* **9**, 128-43 (2004).
8. Wolters, D.A., Washburn, M.P. & Yates, J.R., 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-90 (2001).
9. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7 (2001).

10. Wang, H. & Hanash, S. Multi-dimensional liquid phase based separations in proteomics. *Journal of Chromatography BProteomic Databases Part II* **787**, 11-18 (2003).

11. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-9 (1999).

12. Li, J., Steen, H. & Gygi, S.P. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics* **2**, 1198-204 (2003).

13. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-86 (2002).

14. Everley, P.A., Krijgsveld, J., Zetter, B.R. & Gygi, S.P. Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol Cell Proteomics* **3**, 729-35 (2004).

15. Fiehn, O. et al. Metabolite profiling for plant functional genomics. *Nat Biotechnol* **18**, 1157-61 (2000).

16. Jonsson, P. et al. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem* **76**, 1738-45 (2004).

17. Tolstikov, V.V. & Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal Biochem* **301**, 298-307 (2002).

18. Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N. & Fiehn, O. Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal Chem* **75**, 6737-40 (2003).

19. Soga, T. et al. Pressure-assisted capillary electrophoresis electrospray ionization mass spectrometry for analysis of multivalent anions. *Anal Chem* **74**, 6224-9 (2002).

20. Soga, T. et al. Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal Chem* **74**, 2233-9 (2002).

21. Soga, T. et al. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* **2**, 488-94 (2003).

22. Reo, N.V. NMR-based metabolomics. *Drug Chem Toxicol* **25**, 375-82 (2002).

23. Griffin, J.L. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* **7**, 648-54 (2003).

24. Venkateswaran, K. et al. Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov. *Int J Syst Bacteriol* **49 Pt 2**, 705-24 (1999).

25. Tiedje, J.M. Shewanella--the environmentally versatile genome. *Nat Biotechnol* **20**, 1093-4 (2002).

26. Thompson, D.K. et al. Transcriptional and proteomic analysis of a ferric uptake regulator (fur) mutant of Shewanella oneidensis: possible involvement of fur in

energy metabolism, transcriptional regulation, and oxidative stress. *Appl Environ Microbiol* **68**, 881-92 (2002).

27. Heidelberg, J.F. et al. Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. *Nat Biotechnol* **20**, 1118-23 (2002).

28. Murray, A.E. et al. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc Natl Acad Sci U S A* **98**, 9853-8 (2001).

29. Qiu, X., Sundin, G.W., Wu, L., Zhou, J. & Tiedje, J.M. Comparative analysis of differentially expressed genes in Shewanella oneidensis MR-1 following exposure to UVC, UVB, and UVA radiation. *J Bacteriol* **187**, 3556-64 (2005).

30. Liu, Y. et al. Transcriptome analysis of Shewanella oneidensis MR-1 in response to elevated salt conditions. *J Bacteriol* **187**, 2501-7 (2005).

31. Devreese, B., Vanrobaeys, F. & Van Beeumen, J. Automated nanoflow liquid chromatography/tandem mass spectrometric identification of proteins from Shewanella putrefaciens separated by two-dimensional polyacrylamide gel electrophoresis. *Rapid Commun Mass Spectrom* **15**, 50-6 (2001).

32. Vanrobaeys, F. et al. Proteomics of the dissimilatory iron-reducing bacterium Shewanella oneidensis MR-1, using a matrix-assisted laser desorption/ionization-tandem-time of flight mass spectrometer. *Proteomics* **3**, 2249-57 (2003).

33. Giometti, C.S. et al. Analysis of the Shewanella oneidensis proteome by two-dimensional gel electrophoresis under nondenaturing conditions. *Proteomics* **3**, 777-85 (2003).

34. VerBerkmoes, N.C. et al. Integrating 'top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of Shewanella oneidensis. *J Proteome Res* **1**, 239-52 (2002).

35. Romine, M.F. et al. Validation of Shewanella oneidensis MR-1 small proteins by AMT tag-based proteome analysis. *Omics* **8**, 239-54 (2004).

36. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-4 (1998).

37. Fields, S. & Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet* **10**, 286-92 (1994).

38. Miller, J. & Stagljar, I. Using the yeast two-hybrid system to identify interacting proteins. *Methods Mol Biol* **261**, 247-62 (2004).

39. Matyus, L. Fluorescence resonance energy transfer measurements on cell surfaces. A spectroscopic tool for determining protein interactions. *J Photochem Photobiol B* **12**, 323-37 (1992).

40. Hink, M.A., Bisselin, T. & Visser, A.J. Imaging protein-protein interactions in living cells. *Plant Mol Biol* **50**, 871-83 (2002).

41. Rappsilber, J., Siniossoglou, S., Hurt, E.C. & Mann, M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* **72**, 267-75 (2000).

42. Back, J.W. et al. Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling. *Anal Chem* **74**, 4417-22 (2002).

43. Trester-Zedlitz, M. et al. A modular cross-linking approach for exploring protein interactions. *J Am Chem Soc* **125**, 2416-25 (2003).

44. Trakselis, M.A., Alley, S.C. & Ishmael, F.T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug Chem* **16**, 741-50 (2005).

45. Back, J.W., de Jong, L., Muijsers, A.O. & de Koster, C.G. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* **331**, 303-13 (2003).

46. Novak, P. et al. Unambiguous assignment of intramolecular chemical cross-links in modified Mammalian membrane proteins by fourier transform-tandem mass spectrometry. *Anal Chem* **77**, 5101-6 (2005).

47. Revercomb, H.E.M., E. A. Theory of plasma chromatography/gaseous electrophoresis. Review. *Analytical Chemistry* **47**, 970-83 (1975).

48. Mason, E.A.M., E. W. *Transport properties of Ions in Gases*, (John Wiley & Sons, New York, 1988).

49. Karasek, F.W., Hill, H.H., Jr. & Kim, S.H. Plasma chromatography of heroin and cocaine with mass-identified mobility spectra. *J Chromatogr* **117**, 327-36 (1976).

50. Lawrence, A.H. & Nanji, A.A. Ion mobility spectrometry and ion mobility spectrometry/mass spectrometric characterization of dimenhydrinate. *Biomed Environ Mass Spectrom* **16**, 345-7 (1988).

51. Fytche, L.M., Hupe, M., Kovar, J.B. & Pilon, P. Ion mobility spectrometry of drugs of abuse in customs scenarios: concentration and temperature study. *J Forensic Sci* **37**, 1550-66 (1992).

52. Wu, C., Siems, W.F. & Hill, H.H., Jr. Secondary electrospray ionization ion mobility spectrometry/mass spectrometry of illicit drugs. *Anal Chem* **72**, 396-403 (2000).

53. Bocos-Bintintan, V., Brittain, A. & Thomas, C.L. Characterisation of the phosgene response of a membrane inlet 63Ni ion mobility spectrometer. *Analyst* **127**, 1211-7 (2002).

54. Steiner, W.E., Clowers, B.H., Matz, L.M., Siems, W.F. & Hill, H.H., Jr. Rapid screening of aqueous chemical warfare agent degradation products: ambient pressure ion mobility mass spectrometry. *Anal Chem* **74**, 4343-52 (2002).

55. Steiner, W.E., Clowers, B.H., Haigh, P.E. & Hill, H.H. Secondary ionization of chemical warfare agent simulants: atmospheric pressure ion mobility time-of-flight mass spectrometry. *Anal Chem* **75**, 6068-76 (2003).

56. Steiner, W.E., Klopsch, S.J., English, W.A., Clowers, B.H. & Hill, H.H. Detection of a Chemical Warfare Agent Simulant in Various Aerosol Matrixes by Ion Mobility Time-of-Flight Mass Spectrometry. *Anal Chem* **77**, 4792-9 (2005).

57. Buxton, T.L. & Harrington Pde, B. Trace explosive detection in aqueous samples by solid-phase extraction ion mobility spectrometry (SPE-IMS). *Appl Spectrosc* **57**, 223-32 (2003).

58. Buryakov, I.A. Express analysis of explosives, chemical warfare agents and drugs with multicapillary column gas chromatography and ion mobility increment spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **800**, 75-82 (2004).

59. Perr, J.M., Furton, K.G. & Almirall, J.R. Solid phase microextraction ion mobility spectrometer interface for explosive and taggant detection. *J Sep Sci* **28**, 177-83 (2005).

60. Puskar, M.A. & Plese, M.R. Evaluation of real-time techniques to measure hydrogen peroxide in air at the permissible exposure limit. *Am Ind Hyg Assoc J* **57**, 843-8 (1996).

61. Jung, C.M., Newcombe, D.A., Crawford, D.L. & Crawford, R.L. Detection and decontamination of residual energetics from ordnance and explosives scrap. *Biodegradation* **15**, 41-8 (2004).

62. Beegle, L.W., Kanik, I., Matz, L. & Hill, H.H., Jr. Electrospray ionization high-resolution ion mobility spectrometry for the detection of organic compounds, 1. Amino acids. *Anal Chem* **73**, 3028-34 (2001).

63. Bramwell, C.J., Colgrave, M.L., Creaser, C.S. & Dennis, R. Development and evaluation of a nano-electrospray ionisation source for atmospheric pressure ion mobility spectrometry. *Analyst* **127**, 1467-70 (2002).

64. Asbury, G.R. & Hill, H.H., Jr. Separation of amino acids by ion mobility spectrometry. *J Chromatogr A* **902**, 433-7 (2000).

65. Steiner, W.E., Clowers, B.H. & Hill, H.H., Jr. Rapid separation of phenylthiohydantoin amino acids: ambient pressure ion-mobility mass spectrometry (IMMS). *Anal Bioanal Chem* **375**, 99-102 (2003).

66. Valentine, S.J., Counterman, A.E., Hoaglund, C.S., Reilly, J.P. & Clemmer, D.E. Gas-phase separations of protease digests. *J Am Soc Mass Spectrom* **9**, 1213-6 (1998).

67. Hoaglund, C.S., Valentine, S.J., Sporleder, C.R., Reilly, J.P. & Clemmer, D.E. Three-dimensional ion mobility/TOFMS analysis of electrosprayed biomolecules. *Anal Chem* **70**, 2236-42 (1998).

68. Wu, C., Klasmeier, J. & Hill, H.H., Jr. Atmospheric pressure ion mobility spectrometry of protonated and sodiated peptides. *Rapid Commun Mass Spectrom* **13**, 1138-42 (1999).

69. Asbury, G.R. & Hill, H.H., Jr. Using different drift gases to change separation factors (alpha) in ion mobility spectrometry. *Anal Chem* **72**, 580-4 (2000).

70. Wu, C., Siems, W.F., Klasmeier, J. & Hill, H.H., Jr. Separation of isomeric peptides using electrospray ionization/high-resolution ion mobility spectrometry. *Anal Chem* **72**, 391-5 (2000).

71. Taraszka, J.A., Counterman, A.E. & Clemmer, D.E. Gas-phase separations of complex tryptic peptide mixtures. *Fresenius J Anal Chem* **369**, 234-45 (2001).

72. Woods, A.S. et al. A study of peptide-peptide interactions using MALDI ion mobility o-TOF and ESI mass spectrometry. *J Am Soc Mass Spectrom* **13**, 166-9 (2002).

73. Hoaglund-Hyzer, C.S., Lee, Y.J., Counterman, A.E. & Clemmer, D.E. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures. *Anal Chem* **74**, 992-1006 (2002).

74. Henderson, S.C., Valentine, S.J., Counterman, A.E. & Clemmer, D.E. ESI/ion trap/ion mobility/time-of-flight mass spectrometry for rapid and sensitive analysis of biomolecular mixtures. *Anal Chem* **71**, 291-301 (1999).

75. Jarrold, M.F. Peptides and proteins in the vapor phase. *Annu Rev Phys Chem* **51**, 179-207 (2000).

76. Badman, E.R., Hoaglund-Hyzer, C.S. & Clemmer, D.E. Monitoring structural changes of proteins in an ion trap over approximately 10-200 ms: unfolding transitions in cytochrome c ions. *Anal Chem* **73**, 6000-7 (2001).

77. Matz, L.M., Asbury, G.R. & Hill, H.H., Jr. Two-dimensional separations with electrospray ionization ambient pressure high-resolution ion mobility spectrometry/quadrupole mass spectrometry. *Rapid Commun Mass Spectrom* **16**, 670-5 (2002).

78. Sowell, R.A., Koeniger, S.L., Valentine, S.J., Moon, M.H. & Clemmer, D.E. Nanoflow LC/IMS-MS and LC/IMS-CID/MS of protein mixtures. *J Am Soc Mass Spectrom* **15**, 1341-53 (2004).

79. Liu, X., Plasencia, M., Ragg, S., Valentine, S.J. & Clemmer, D.E. Development of high throughput dispersive LC-ion mobility-TOFMS techniques for analysing the human plasma proteome. *Brief Funct Genomic Proteomic* **3**, 177-86 (2004).

80. Koomen, J.M. et al. Oligonucleotide analysis with MALDI-ion-mobility-TOFMS. *Anal Bioanal Chem* **373**, 612-7 (2002).

81. Gidden, J., Ferzoco, A., Baker, E.S. & Bowers, M.T. Duplex formation and the onset of helicity in poly d(CG)n oligonucleotides in a solvent-free environment. *J Am Chem Soc* **126**, 15132-40 (2004).

82. Leavell, M.D., Gaucher, S.P., Leary, J.A., Taraszka, J.A. & Clemmer, D.E. Conformational studies of Zn-ligand-hexose diastereomers using ion mobility

measurements and density functional theory calculations. *J Am Soc Mass Spectrom* **13**, 284-93 (2002).

83. Clowers, B.H., Dwivedi, P., Steiner, W.E., Hill, H.H., Jr. & Bendiak, B. Separation of sodiated isobaric disaccharides and trisaccharides using electrospray ionization-atmospheric pressure ion mobility-time of flight mass spectrometry. *J Am Soc Mass Spectrom* **16**, 660-9 (2005).

84. Collins, D.C. & Lee, M.L. Developments in ion mobility spectrometry-mass spectrometry. *Anal Bioanal Chem* **372**, 66-73 (2002).

85. Liu, Y.C., David E. Characterizing oligosaccharides using injected-ion mobility/mass spectrometry. *Analytical Chemistry* **69**, 2504-2509 (1997).

86. Hill, H.H., Jr., Siems, W.F., St Louis, R.H. & McMinn, D.G. Ion mobility spectrometry. *Anal Chem* **62**, 1201A-1209A (1990).

87. Wu, C.S., W. F.; Asbury, G. R.; Hill, H. H. Jr. Electrospray Ionization High-Resolution Ion Mobility Spectrometry-Mass Spectrometry. *Analytical Chemistry* **70**, 4929-4938 (1998).

88. Dugourd, P.H., R. R.; Clemmer, D. E.; Jarrold, M. F. High-resolution ion mobility measurements. *Review of Scientific Instruments* **68**, 1122-1129 (1997).

89. Srebalus, B., Hilderbrand, A.E., Valentine, S.J. & Clemmer, D.E. Resolving isomeric peptide mixtures: a combined HPLC/ion mobility-TOFMS analysis of a 4000-component combinatorial library. *Anal Chem* **74**, 26-36 (2002).

90. Liu, Y.V., Stephen J.; Counterman, Anne E.; Hoaglund, Cherokee S.; Clemmer, David E. Injected-ion mobility analysis of biomolecules. *Analytical Chemistry* **69**, 728A-735A (1997).

91. Creaser, C.S., Benyezzar, M., Griffiths, J.R. & Stygall, J.W. A tandem ion trap/ion mobility spectrometer. *Anal Chem* **72**, 2724-9 (2000).

92. Clowers, B.H.H., Herbert H., Jr. Mass Analysis of Mobility-Selected Ion Populations Using Dual Gate, Ion Mobility, Quadrupole Ion Trap Mass Spectrometry. *Analytical Chemistry* (2005).

93. Bluhm, B.K.G., Kent J.; Russell, David H. Development of a Fourier-transform ion cyclotron resonance mass spectrometer-ion mobility spectrometer. *Review of Scientific Instruments* **71**, 4078-4086 (2000).

94. Shi, S.D., Hendrickson, C.L. & Marshall, A.G. Counting individual sulfur atoms in a protein by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry: experimental resolution of isotopic fine structure in proteins. *Proc Natl Acad Sci U S A* **95**, 11532-7 (1998).

95. He, F., Hendrickson, C.L. & Marshall, A.G. Baseline mass resolution of peptide isobars: a record for molecular mass resolution. *Anal Chem* **73**, 647-50 (2001).

96. Belov, M.E. et al. Electrospray ionization-Fourier transform ion cyclotron mass spectrometry using ion preselection and external accumulation for ultrahigh sensitivity. *J Am Soc Mass Spectrom* **12**, 38-48 (2001).

## CHAPTER 2

## A New Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein-Protein Interactions


### Abstract

A new mass spectrometry identifiable cross-linking strategy has been developed to study protein-protein interactions. The new cross-linker was designed to have two low energy MS/MS cleavable bonds in the spacer chain to provide three primary benefits. First, a reporter tag can be released from cross-link due to cleavage of the two labile bonds in the spacer chain. Second, a relatively simple MS/MS spectrum can be generated owing to favorable cleavage of labile bonds. And finally, the cross-linked peptide chains are dissociated from each other and each then can be fragmented separately to get sequence information. Therefore, this novel type of cross-linker was named Protein Interaction Reporter (PIR). To this end, two RINK groups were utilized to make our first-generation cross-linker using solid phase peptide synthesis chemistry. The RINK group contains a bond more labile than peptide bonds during low energy activation. The new cross-linker was applied to cross-link ribonuclease S (RNase S), a non-covalent complex of S-peptide and S-protein. The results demonstrated that the new cross-linker effectively reacted with RNase S to generate various types of cross-linked products. More importantly, the cross-linked peptides successfully released reporter ions during selective MS/MS conditions and the dissociated peptide chains remained intact during $MS^2$, thus enabling

$MS^3$ to be performed subsequently. In addition, dead-end, intra-, and inter-cross-linked peptides can be distinguished by analyzing MS/MS spectra.

KEYWORDS: cross-linking, cross-linker, protein interactions, mass spectrometry, MS/MS, LC/MS/MS

## Introduction

Most cellular functions are carried out through large multi-protein complexes rather than individual proteins. Mapping of proteome-wide protein interactions is essential for understanding the processes in biological systems. However, characterization of protein interactions poses an extremely challenging task for the reasons that many protein interactions are not stable, most interactions are transient, and multi-protein complexes possess no common factors or physical properties that can be used as an analytical handle. Despite these difficulties, numerous efforts have been undertaken to profile large-scale protein interaction networks within a given cell.[1-4] Cross-linking in combination with mass spectrometric techniques is one commonly used approach for studying protein-protein interactions[5-7] and topological structures of protein complexes.[8-11] In addition to identifying what proteins interact within a complex, cross-linking approaches can also pinpoint where proteins interact. However, in spite of much interest in profiling protein-protein interactions using cross-linking strategies and the availability of many commercial cross-linkers,[12] few reports have illustrated successful application of

cross-linkers on a proteome-wide scale, mostly due to inherent complexity in cross-linking reaction mixtures which include large amounts of undesirable or unexpected products[13] in addition to 1:1 linkage of interacting proteins. An enzymatic digestion of cross-linked complex is commonly performed prior to analysis by mass spectrometry to identify interacting proteins and interacting regions of proteins. Proteolysis further increases the complexity since the digestion mixture includes predominately unmodified peptides, various types of singly modified peptides (so called dead-end, intra-, and inter-cross-linked peptides)[14], multiply modified peptides, and non-specific modified peptides.[13] As a result, identification of cross-linked peptides can be much like looking for a needle in a haystack. To facilitate identification of cross-linked peptides (needles) in complex mixtures (haystacks), researchers have developed a number of strategies that have been recently reviewed.[10,11] Most strategies focused on reducing the complexity of cross-linking mixtures by introducing a tag or a pattern associated with the cross-links. For example, the incorporation of an affinity tag such as biotin in the cross-linker allowed enrichment of cross-linked peptides.[5,15,16] Bennett et al.[17] and Back et al.[18] reported the use of cleavable cross-linkers and comparison of mass shifts in the spectra before and after cleavage of the cross-linker. The use of isotope labeling either the cross-linkers[19,20] or proteins[21] were shown to create a signature pattern for cross-linked peptides in mass spectra. Back et al[22] reported the use of $^{18}O$ enriched water for proteolysis to generate a mass shift of 8 Da for inter-cross-linked peptides as opposed to a 4 Da mass shift for other peptides. Finally, the same group also reported a new cross-linker which has a benzyl group attached in the spacer chain. The benzyl group can be cleaved by low energy MS/MS and was thus used to track cross-linked peptides.[23] However, the

challenge of the cross-linking approach is not only resultant from complexity of cross-linking reaction and digestion mixtures, but also related to the complexity of MS/MS mass spectra of cross-linked peptides. Our aim is to develop a new mass spectrometry identifiable cross-linking strategy to address both problems. The new type of cross-linker, which we called Protein Interaction Reporter (PIR), was designed to include two labile bonds. This allows specific release of reporter ions by low energy MS/MS, while also allowing specific cleavage of the spacer arm to release cross-linked peptide chains, without disrupting peptide backbones. The individual peptide chains can be then sequenced separately with additional stages of MS/MS. This type of structural property can significantly extend the ability to identify cross-linked proteins and peptides with mass spectrometry by reducing the complexity of both cross-linking reaction mixtures and MS/MS spectra. It has been noticed from our previous experience that RINK, a compound commonly used for amidation in peptide synthesis,[24] is preferably cleaved by mass spectrometry at low energy. Thus, two RINK groups were incorporated in our first-generation PIR which was made with solid phase peptide synthesis chemistry. Ribonuclease S (RNase S), a complex of S-peptide 1-20 and S-protein 21-124, was used to evaluate our new cross-linker. The results from our initial study demonstrated that PIR effectively reacted with RNase S to generate a cross-linked complex. Proteolysis of the cross-linked complex followed by nano-LC/MS/MS analysis demonstrated that cross-linked peptides successfully released reporter ions and intact peptide chains during low energy MS/MS. The dissociated peptide chains were able to be sequenced by another stage of MS/MS. In addition, careful analysis of MS/MS spectra showed that our cross-linking strategy was able to differentiate dead-end, intra-, and inter-cross-linked peptides.

## Experimental Section

**Chemicals.** RNase S was purchased from Sigma (St. Louis, MO) and used without further purification. Fmoc-RINK, Fmoc-Lys (Fmoc)-OH, and HMPB-MBHA resins used for synthesis of cross-linker were purchased from Novabiochem (San Diego, CA). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Water used for preparing solution and solvent was 18 MΩ deionized water made by Barnstead Nanopure Water Systems.

**Synthesis of Cross-Linker.** The cross-linker was synthesized using 431A Peptide Synthesizer (Applied Biosystem, Foster City, CA). Fmoc-Lys (Fmoc)-OH was coupled to HMPB-MBHA resin (4-hydroxymethyl-3-methoxyphenoxybutyric acid MBHA resin) using the standard symmetric anhydride method. Two RINK groups were coupled to Lys by treating the Fmoc-RINK linkers as Fmoc amino acid. Carboxyl groups were then introduced by reacting the primary amines of the RINK groups with succinyl anhydride. Subsequently, the two carboxyl groups were activated by forming the esters with N-hydroxysuccinimide (NHS). The final product was cleaved from the super acid sensitive resin with 0.5% TFA in chloroform.

**Cross-Linking Reaction.** The purified final product was dissolved in DMSO to make 100 mM stock solution. RNase S was dissolved in water to make 1 mM stock solution. For cross-linking reaction, 1 mM RNase S was diluted to 10 μM in phosphate saline buffer (100 mM sodium phosphate, 150 mM NaCl, pH 7.2) and the cross-linker was added in 50-, 100-, and 150-fold molar excess over RNase S. The reaction mixture

was incubated at room temperature and an aliquot was taken at 15, 30, 60, and 120 min and quenched by 1 M Tris, pH 7.5 (final concentration 40 mM).

**SDS-PAGE and In-Gel Digestion.** An aliquot of each reaction mixture was loaded onto SDS-PAGE for separation by using precast 4-12% gel from Invitrogen (Carlsbad, CA). The gels were stained with Coomassie blue (Bio-Rad, Hercules, CA) and imaged by Densitometer (Molecular Dynamics). The gel bands of interest were excised and destained by 50% methanol and 5% acetic acid. Performic acid oxidation method was used to dissociate di-sulfide bonds of the proteins. Performic acid was freshly made by mixing 95/5 (v/v) formic acid/30% $H_2O_2$ and incubating the mixture at room temperature for 20 min followed by incubating on ice for 5 min. 20-30 μL performic acid was used to cover the gel pieces and oxidize proteins on ice for 1 hr. Then the gels were washed and dried by acetonitrile. 20 ng/μL trypsin in 50 mM ammonium bicarbonate was used for digesting proteins at 37$^{o}$C overnight.

**Nano-LC/MS/MS and Data Analysis.** An electrospray-ion trap (Esquire *HCT*, Bruker Daltonics, Billerica, MA) mass spectrometer coupled with a nano-HPLC was used for nano-LC/MS/MS analysis of cross-linked RNase S complex. Nano-HPLC separation of tryptic peptides was performed with an LC packings Ultimate Nano-HPLC system equipped with a Famos micro autosampler and a Swichos micro column switching module (Dionex, Sunnyvale, CA). Samples were first injected by the autosampler and loaded onto a micro trap column (C18 PepMap, 300 μm × 1 mm, 5 μm, LC Packings) at a flow rate of 50 μL/min with solvent A (0.1% TFA in water). The loaded samples were continuously washed with solvent A for 3 minutes to remove salts. Peptides were then eluted at a flow rate of 300 nL/min to an analytical column (C18

PepMap, 75 μm × 150 mm, 3 μm, 100 Å, LC Packings) and separated using the following gradient: 0% B for 0-3 min, 20-70% B for 3-45 min, 90% B for 45-55 min, and 0% B for 55-65 min. Solvent B was 0.1% TFA in 95% acetonitrile. The eluant from the analytical column was sprayed on-line with a nano-spray emitter to the Esquire *HCT* mass spectrometer. The nano-spray emitter was made in house by etching fused silica capillary (20 μm × 360 μm) with HF. The spray potential was set at 1300-1500 volts. LC/MS/MS mass spectra were acquired using HyStar software (version 2.3, Bruker Daltonics). MS/MS data acquisition was set in automatic mode with active exclusion based on peak intensity and a selection of exclusion peak lists. Two precursor ions were selected from each MS scan and excluded after two MS/MS scans. $MS^3$ was performed manually by selecting specific precursor ions from MS/MS scan at specific retention time based on previous LC/MS/MS analysis of the same set of samples. For analysis of cross-linking reaction mixtures, samples were precipitated by TCA to remove the large amount of excessive unreacted cross-linker prior to loading to the nano-HPLC column. Calibration of the instrument was achieved using a solution of tuning mix (ES tuning mix, Agilent) composed of ions at *m/z* 322.05, 622.03, 922.01, 1521.93, and 2121.93. Data analysis and processing were performed using Bruker Daltonics DataAnalysis software (version 3.1). GPMAW software version 6.00 (Lighthouse Data, Odense, Denmark) was used to calculate the *m/z* of cross-linked peptides.

**Results and Discussion**

**Cross-Linker Design and Synthesis.** In spite of the great enthusiasm for using cross-linking approaches, successful reports for studying protein-protein interactions on a proteome-wide scale have been scarce, due in part to the complexity of mass spectral information and the need for additional cross-linker features. Our goal of developing mass spectrometry identifiable cross-linking strategy using PIR cross-linkers is to simplify MS/MS spectrum of cross-linked peptides and to guide data analysis by including a specific reporter. The conceptual components of the chemistry development for new cross-linker reagents are presented in Figure 1. Such a cross-linker should have low-energy mass spectrometry cleavable bonds which are more labile than peptide bonds. This feature is important to improved capabilities of cross-linking approach, since it can significantly benefit spectral interpretation by allowing additional stages of MS/MS to be carried out. Mass measurement of the intact cross-linked species, followed by MS/MS analysis of each of the peptide chains will provide both, the identities of cross-linked proteins, as well as information regarding sites of interaction. Based on our previous experience, we first coupled a RINK group to a 5-mer peptide and tested its specific cleavage property by low energy MS/MS. As expected, a majority of the fragmentation occurred at the amine position between two phenyl groups in the RINK, and the peptide chain remained intact (data not shown). This fragmentation was accomplished with activation energy much below that needed for peptide backbone fragmentation. Therefore two RINK groups were used to provide two labile bonds for our new cross-linker. Cleavage of the two labile bonds results in a specific reporter as shown in Figure 1. Our first cross-linker was made using solid phase Fmoc peptide synthesis chemistry. Two RINK groups were treated as amino acid analogs and coupled to two primary

amines in a lysine residue. For the time being, two NHS esters were subsequently incorporated to make our initial cross-linker. However, the proposed modular system can readily adapt new features by using different building blocks to incorporate an affinity tag or other functional groups as shown in Figure 1. A hetero-bifunctional cross-linker with one reactive group being NHS ester and the other one being benzophenone photoreactive group has been made based on this modular system in our group (data not shown). Figure 2 shows the structure of the new cross-linker and proposed reaction scheme. The MS/MS cleavage takes place at the amine bond between two phenyl groups in RINK as indicated by dashed lines in Figure 2. The carbonyl cation ion formed after cleavage was delocalized among two phenyl groups and methoxyl groups and thus highly stable. Further study showed that the cleavage of labile bonds only took place by MS/MS in positive ion mode and not in negative ion mode, indicating protonation of parent compound was critical for desired fragmentation. Moreover, the same cleavage can occur if this compound is treated with 90% TFA, which is commonly used for peptide amide synthesis. The acid cleavage feature can be used for tracking cross-links by comparing mass shift prior to and following cleavage of cross-links as any other cleavable cross-linkers such as DTSSP.[17,18] This feature was not used for the present studies, since the mass spectrometric cleavage features of the cross-linkers provide more benefits. As shown in Figure 2, if both vulnerable bonds in the cross-linker are cleaved, a reporter ion at *m/z* 711 will be released and detected. This reporter ion can be used to track peptides which are either inter- or intra-cross-linked. If one end of the cross-linker is hydrolyzed and the labile bond on the other end is cleaved, as in the case that a peptide is dead-end modified, a reporter ion at *m/z* 828 will be detected. This reporter ion can be

further cleaved to generate ion at *m/z* 711. To evaluate the cross-linker we have made and its mass spectrometric dissociation properties, the purified final product was dissolved in acetonitrile to make a 10 μM solution and directly infused to an ESI-ion trap mass spectrometer. Figure 3 shows the mass spectra of the cross-linker. As shown in Figure 3(a), at low capillary exit voltage, the cross-linker remained mostly intact at *m/z* 1139.6. Some cross-linker was hydrolyzed at one end during electrospray process generating ions at *m/z* 1042.5. Ions at *m/z* 925.6 were the product from cleavage of one labile bond of the cross-linker. With increased capillary exit voltage, more cleavage was observed. As shown in Figure 3 (b), reporter ions at *m/z* 711.4 and 828.5 resultant from activation in the region of capillary and skimmer were detected, indicating the cross-linker can successfully release reporter ions under low energy MS/MS conditions.

**Cross-Linking of RNase S Complex.** Ribonuclease S (RNase S) is a complex of S-peptide 1-20 and S-protein 21-124 (redefined 1-104) that are the hydrolysis products from ribonuclease A (RNase A) as a result of cleavage between residue 20 (Ala) and 21 (Ser) by subtilisin. S-peptide and S-protein have been found to form a weak non-covalent complex that retains similar enzymatic activity as RNase A. RNase S complex provides an ideal model system for testing our new cross-linker. RNase S complex is believed to be heterogeneous;[25] other S-peptides and S-proteins were also found in the complex but only the first 15 residues of S-peptide are involved in non-covalent binding with S-protein.[26] The RNase S used in this research was found to have S-peptides 1-20, 1-19, 1-18, 1-17, and 1-16 and S-peptide 1-19 was most abundant. Nevertheless, RNase S was used as is without further purification. With the cross-linker and RNase S ready, our first attempt was to optimize cross-linking reaction condition. It was carried out by

varying molar ratios of RNase S to the cross-linker at 1:50, 1:100, and 1:150, and reaction time at 15, 30, 60, and 120 min for each reaction ratio. RNase S with no cross-linker added in the reaction buffer was used as a control. The resulting cross-linking reaction mixtures were separated by 4-12% SDS-PAGE followed by Coomassie blue staining. Figure 4 shows the SDS-PAGE separation of cross-linking reaction mixture with molar ratios of RNase S and cross-linker at 1:50 and 1:100. The molecular masses of S-protein, S-peptide, and the cross-link are 11534 Da, 2166 Da, and 908 Da, respectively. Thus, the 1:1 cross-linked complex is located at around 14.6 kDa. As shown in Figure 4, in the control lane only S-protein was detected at expected region but not any S-peptide was observed probably due to the fact it was too small to be retained on the gel. With the cross-linker included in the reaction mixture, a distinctive band above S-protein band was clearly observed on the gel even at the first reaction time point, 15 min, at a reaction ratio of 1:50. The upper bands observed on the gel were believed to be cross-linked complex of S-peptide and S-protein. They appeared at higher molecular weight than 14.6 kDa for the reason that the cross-linked complex may migrate slower than it is expected to be. With 100-fold excess of the cross-linker used, all upper bands appeared more intense than lower bands. As compared to a reaction ratio of 1:50, the upper bands in reaction ratio of 1:100 located at slightly higher position indicating that more dead-end modification or intra-cross-linking of proteins have occurred. Since 1:1 linkage of interacting proteins is most desirable, reaction ratio at 1:50 at incubation time 15 min was used for the following experiments. For in-gel trypsin digestion, the gel bands were excised carefully to include most proteins with minimum volume. The tryptic digest was further analyzed by nano-LC/MS/MS.

**Analysis of Cross-Linked Proteins.** The in-gel tryptic digest of cross-linked complex was further analyzed by nano-LC/MS/MS to locate the cross-links. LC/MS/MS data were acquired with data-dependent mode. As described above, during LC/MS/MS process, detection of ions at *m/z* 711 indicates the potential existence of the precursor ions carrying cross-links, and observation of ions at *m/z* 828 is a signal that precursor ion is a dead-end modified peptide. Therefore, extracted ion chromatograms (EICs) are built for ions at *m/z* 711 and 828 and the peaks in EICs provide useful clues for identifying cross-link containing peptide ions. This information can be extremely helpful since it narrows down the search window for cross-links dramatically. Figure 5 presents an example illustrating the significance of reporter ions for tracking cross-links from a complex tryptic digest mixture. Figure 5 (a) is the base peak chromatogram which is filtered to have MS scan data only. Figure 5 (b) and (c) are the reconstructed EICs for ions at *m/z* 828 and 711 during MS/MS scans, respectively. The peaks in these EICs indicate that cross-link containing species eluted during retention time 48-51 min. As guided by these peaks, further analysis focused on these MS/MS spectra which have ions at *m/z* 828 and 711. Due to the specific cleavage features of the PIR, MS/MS spectra of cross-link containing peptides give relatively simple fragmentation patterns, i.e., most intense peaks are generated from cleavage of labile bonds in the cross-linker. The simple fragmentation pattern can be further utilized to differentiate dead-end, intra-, or inter-cross-linked peptides as illustrated in Figure 6. For dead-end modified peptides, the cleavage of the precursor ions generates the reporter ion at *m/z* 828 and the intact peptide chain with the remaining tag. The *m/z* of this peptide chain can be deduced by subtracting 828 from the mass of the precursor ion. For cleavage of intra-cross-linked

peptides, it is expected that the reporter ions at *m/z* 711 and ions of the intact peptide chain with two remaining tags will be observed. In this case, the *m/z* of this peptide chain is the difference of the mass of the precursor ion and 711. If an inter-cross-linked peptide complex is fragmented, two separate intact peptide chains with the remaining tags are generated in addition to reporter ions at *m/z* 711. Thus, the sum of masses of the two peptide chains should equal the difference of the mass of precursor ion and 711.

As directed by EICs of the reporter ions, an inter-cross-linked peptide was readily identified as shown in Figure 7. The cross-link was identified to be between tryptic peptide fragment (1-1), Lys 1 of S-peptide and fragment (14-19), NLTKDR of S-protein. The doubly charged precursor ions at *m/z* $900.9^{2+}$ were isolated and fragmented. The MS/MS spectrum shows the efficient cleavage of the labile bonds in the spacer chain to generate ions at *m/z* $245.9^{+}$, $711.3^{+}$, $778.4^{2+}$, $778.4^{+}$, and $845.6^{+}$ as indicated in Figure 7 (a). By analyzing the MS/MS spectrum, it can be further confirmed that the precursor ions were from an inter-cross-linked peptide owing to the fact that the sum of 711, neutral mass 244.9 from one peptide chain, and neutral mass 844.6 from the other peptide chain is equal to the *m/z* of singly charged precursor ions. The undisrupted peptide chain at *m/z* 845.6 was further fragmented by another stage of MS/MS. As shown in Figure 7 (b), the fragment ions generated by $MS^3$ supported the assignment of the peptide sequence with the remaining tag attached at Lys 17. Figure 8 shows X-ray crystal structure of RNase S and sequences of S-peptide and S-protein. Because the peptides were generated by trypsin digestion, the cross-link was assigned to the N-terminal amine instead of the ε-amine of Lys 1 in S-peptide since trypsin would likely not cut at this lysine otherwise. The inter-cross-linked peptide from a bigger tryptic peptide fragment (1-7) with one

missed cleavage, KETAAAK of S-peptide, was not observed in the mass spectra. The reason for this observation could be because the hydrophobic nature of larger cross-linked peptides makes it difficult for recovery from gels, as has been observed by other researchers.[10] In our future studies, a biotin group will be incorporated in the cross-linker thus affinity purification and other non-gel based separations can be employed to purify cross-linked peptides from in-solution enzymatic digest. No other cross-links were identified at this point except for a dead-end modification that was observed on Lys 17 of S-protein (data not shown). However, it was easier to comprehend why the cross-link was formed between N-terminus of S-peptide and Lys 17 of S-protein when the cross-link was transposed upon the space-filling model of crystal structure of RNase S as indicated in Figure 8. The N-terminus of S-peptide and Lys 17 of S-protein are well exposed to solvent and thus, have good accessibility for cross-linking reaction. This is consistent with reported observation that residues 3-13 of the S-peptide and 96-100 of S-protein are involved in binding.[26,27]

In addition, the cross-linking reaction mixture was analyzed directly in the absence of tryptic digestion using nano-LC/MS/MS to check for modifications of S-peptide. An aliquot of cross-linking reaction mixture with 100-fold excess of cross-linker at incubation time of 30 min was precipitated by TCA to remove salts and most unreacted cross-linker. The precipitated peptides were resuspended in 10% acetonitrile solution and loaded to the nano-LC column for LC/MS/MS analysis. Both forms of S-peptide 1-19 and 1-20 were observed to be modified by cross-linking reactions to generate dead-end and intra-cross-linked peptides. As an example, Figure 9 shows mass spectra of the dead-end modified S-peptide 1-19. The triply charged precursor ions at

*m/z* 1007.3 was isolated and fragmented by MS/MS. The majority of fragmentation took place at two labile bonds in the spacer chain to generate a simple $MS^2$ spectrum. As shown in Figure 9 (a), ions at *m/z* $711.6^{+}$, $1097.6^{2+}$, and $969.5^{3+}$ were dominant in the spectrum. The observation of ions at *m/z* $1097.6^{2+}$ further confirmed that the precursor was a dead-end modified peptide, since they were produced from the precursor ions by losing the reporter ions at *m/z* 828. To get sequence information of the intact peptide chain, ions at *m/z* $1097.6^{2+}$ were isolated for $MS^3$. As shown in Figure 9 (b), the observation of a series of b ions ($b_{10}$, $b_{11}$, $b_{12}$, $b_{13}$, $b_{14}$, $b_{17}$, and $b_{18}$) provided the sequencing information for the modified S-peptide. In particular, the existence of fragment ions at $y_{13}$ and $y_{14}$ confirmed that the modification occurred at N-terminus or ε-amine of Lys 1 of S-peptide but not Lys 7. An intra-cross-linked S-peptide 1-19 was also identified. The precursor ions at *m/z* $1001.3^{3+}$ were isolated and fragmented by MS/MS. Figure 10 shows the fragmentation results. Due to the cleavage of two labile bonds in the spacer chain, ions at *m/z* $711.4^{+}$ and $1147.1^{2+}$ were detected, which also gave an indication that the precursor ions were from an intra-cross-linked peptide since the sum of the neutral masses of these two ions equals the neutral mass of the precursor ion. Ions at *m/z* $956.3^{3+}$ were produced from cleavage of one labile bond in the spacer chain with additional loss of a di-methoxyl phenyl group. Ions at *m/z* $969.3^{3+}$ were formed by cleaving one labile bond and the remaining tag in the peptide. These ions were further cleaved at the other labile bond to generate ions at *m/z* $1097.6^{2+}$. The peptide ions at *m/z* $1147.1^{2+}$ were isolated for further fragmentation by $MS^3$. The fragmentation pattern was observed to be similar among S-peptide 1-19 with two tags as shown in Figure 10 (b), one tag as shown in Figure 9 (b), and S-peptide 1-19 itself (data not shown), indicating

the remaining tags did not interfere the fragmentation of the peptide chain. The fragment ions observed at $b_5$, $b_6$, and $y_{13}$ in Figure 10 (b) confirmed that the intra-cross-link was formed between α-amine at N-terminus and ε-amine of Lys 1 and Lys 7 in S-peptide was not involved.

The three $MS^2$ and $MS^3$ spectra presented here illustrate that our cross-linking strategy can effectively distinguish dead-end, intra-, and inter-cross-linked peptides. The results also demonstrated that the bulky spacer chain in our cross-linker did not affect its cross-linking capability. Nevertheless, the use of the bulky spacer was able to generate reasonable size of reporter ions to be detected by any type of tandem mass spectrometer. The release of the bulky group did not only provide reporter ions but also made it convenient for efficient fragmentation of peptide chains with short tags during $MS^3$. In addition, the maximum UV absorption wavelength of the cross-linker was measured to be 280 nm, which is different than that of most peptides (214 nm). This feature can be potentially used as another handle for retrieving peptides carrying cross-links. Since the spacer chain length of the cross-linker in a fully extended confirmation was calculated to be about 43 Å, it was of concern initially that the length of our cross-linker might be too big to effectively cross-link interacting proteins. However, it has been reported that the realistic lengths of many commercial cross-linkers are significantly shorter than they are cited.[28] These results emphasized that the cross-linker is not rigid in solution. The bonds in cross-linker are flexible and can be folded in different confirmation to give a broad range of distances between the two reactive groups, which was observed in our studies as well. This feature is preferred for general protein-protein interaction studies. For the inter-cross-link identified between the N-terminus of S-peptide and Lys 17 of S-protein,

the distance was estimated to be 14 Å from X-ray structure. In the intra-cross-linked S-peptide 1-19, the cross-link was formed between the α-amine of N-terminus and the ε-amine of Lys 1 instead of Lys 7. The maximum distance of the α-amine of N-terminus and the ε-amine of Lys 1 was calculated to be 9.1 Å. These observations further suggest that the cross-linker is fairly flexible in solution to react with the most accessible primary amines in proximity.

## Conclusions

The new mass spectrometry identifiable cross-linking strategy has been developed and validated with our first-generation PIR cross-linker. The new cross-linker was successfully synthesized and applied to cross-link the RNase S complex. The presence of two RINK groups and a lysine residue in the spacer chain did not hinder the cross-linking reactivity of NHS groups. Low energy MS/MS of cross-linked peptides not only readily released the distinctive reporter ions allowing rapid screening of cross-linked peptides in complex mixtures but also preserved intact peptide chains which allowed further fragmentation by another stage of MS/MS. The cross-linker was synthesized based on a peptide scaffold which makes the synthesis flexible and versatile for a variety of modifications. Our next-generation cross-linkers will incorporate biotin or other affinity groups to enrich cross-linked peptides from a more complex system such as a whole proteome. Our development of new PIR cross-linkers with specific affinity and cleavage properties will ultimately facilitate profiling of proteome-wide protein-protein interactions.

**Acknowledgments**

This research wassupported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924; NIH-NCRR, Grant No. 1 S10 RR017805-01; and The Murdock Charitable Trust.

**References**

(1) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180-183.

(2) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.;

Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141-147.

(3) Blagoev, B.; Kratchmarova, I.; Ong, S. E.; Nielsen, M.; Foster, L. J.; Mann, M. *Nat Biotechnol* **2003**, *21*, 315-318.

(4) Drewes, G.; Bouwmeester, T. *Curr Opin Cell Biol* **2003**, *15*, 199-205.

(5) Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W. *J Am Chem Soc* **2003**, *125*, 2416-2425.

(6) Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. *Anal Chem* **2000**, *72*, 267-275.

(7) Lanman, J.; Lam, T. T.; Barnes, S.; Sakalian, M.; Emmett, M. R.; Marshall, A. G.; Prevelige, P. E., Jr. *J Mol Biol* **2003**, *325*, 759-772.

(8) Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. *Biochemistry* **2004**, *43*, 4703-4715.

(9) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc Natl Acad Sci U S A* **2000**, *97*, 5802-5806.

(10) Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J Mol Biol* **2003**, *331*, 303-313.

(11) Sinz, A. *J Mass Spectrom* **2003**, *38*, 1225-1237.

(12) Pierce Chemicals, Double-Agents Cross-Linking Guide **1999**.

(13) Swaim, C. L.; Smith, J. B.; Smith, D. L. *J Am Soc Mass Spectrom* **2004**, *15*, 736-749.

(14) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J Am Soc Mass Spectrom* **2003**, *14*, 834-850.

(15) Itoh, Y.; Cai, K.; Khorana, H. G. *Proc Natl Acad Sci U S A* **2001**, *98*, 4883-4887.

(16) Alley, S. C.; Trakselis, M. A.; Mayer, M. U.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J. *J Biol Chem* **2001**, *276*, 39340-39349.

(17) Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. *Protein Sci* **2000**, *9*, 1503-1518.

(18) Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G.; De Koster, C. G.; Grivell, L. A.; Van Der Spek, H.; Muijsers, A. O. *Protein Sci* **2002**, *11*, 2471-2478.

(19) Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal Chem* **2001**, *73*, 1927-1934.

(20) Pearson, K. M.; Pannell, L. K.; Fales, H. M. *Rapid Commun Mass Spectrom* **2002**, *16*, 149-159.

(21) Taverner, T.; Hall, N. E.; O'Hair, R. A.; Simpson, R. J. *J Biol Chem* **2002**, *277*, 46487-46492.

(22) Back, J. W.; Notenboom, V.; de Koning, L. J.; Muijsers, A. O.; Sixma, T. K.; de Koster, C. G.; de Jong, L. *Anal Chem* **2002**, *74*, 4417-4422.

(23) Back, J. W.; Hartog, A. F.; Dekker, H. L.; Muijsers, A. O.; de Koning, L. J.; de Jong, L. *J Am Soc Mass Spectrom* **2001**, *12*, 222-227.

(24) Rink, H. *Tetrahedron Lett.* **1987**, *28*, 3787-3790.

(25) Kirchner, M.; Li, Y.; Deng, H., 52nd ASMS Conference on Mass Spectrometry and Allied Topics, Nashville, TN **2004**.

(26) Wyckoff, H. W.; Hardman, K. D.; Allewell, N. M.; Inagami, T.; Tsernoglou, D.; Johnson, L. N.; Richards, F. M. *J Biol Chem* **1967**, *242*, 3749-3753.

(27) Taylor, H. C.; Komoriya, A.; Chaiken, I. M. *Proc Natl Acad Sci U S A* **1985**, *82*, 6423-6426.

(28) Green, N. S.; Reisler, E.; Houk, K. N. *Protein Sci* **2001**, *10*, 1293-1304.

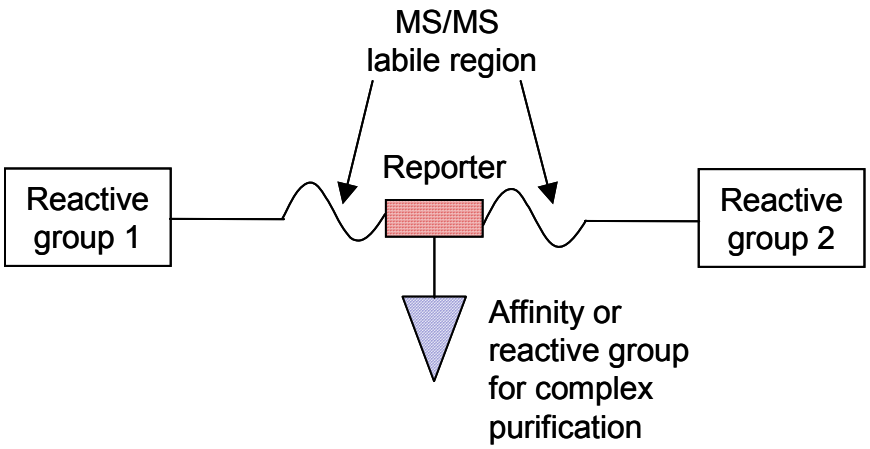



**Figure 1.** Conceptual modular design of novel cross-linkers

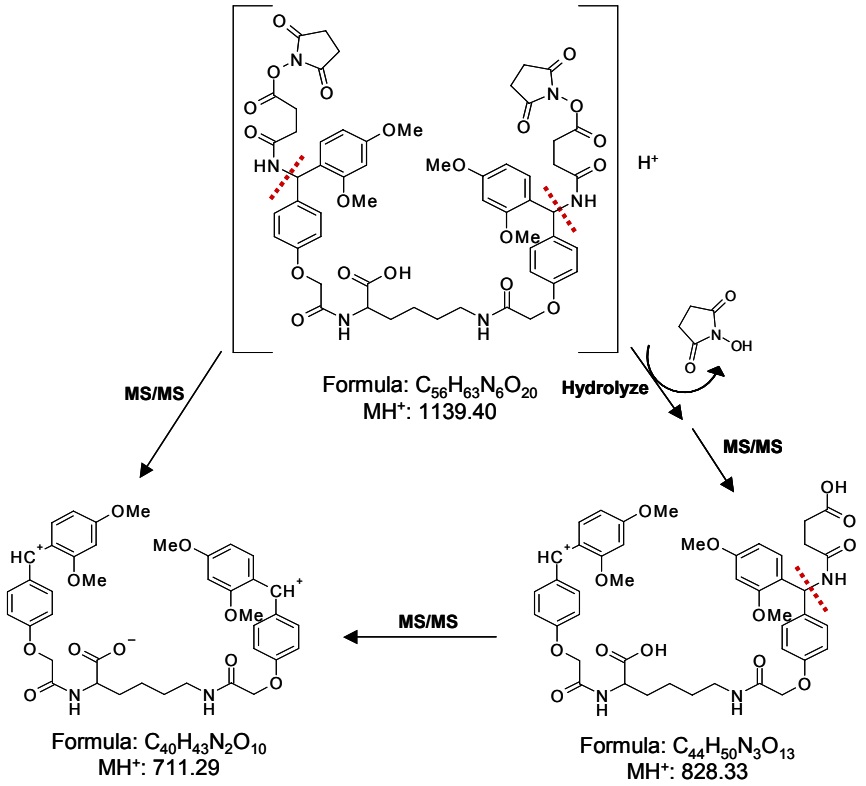



**Figure 2.** Cross-linker structure and proposed reaction scheme. Two labile bonds preferably cleaved by low energy MS/MS are indicated by dashed lines. Cleavage of both labile bonds gives a reporter ion at *m/z* 711. If one end of the cross-linker is hydrolyzed and then the labile bond on the other end is cleaved, a reporter ion at *m/z* 828 is generated.

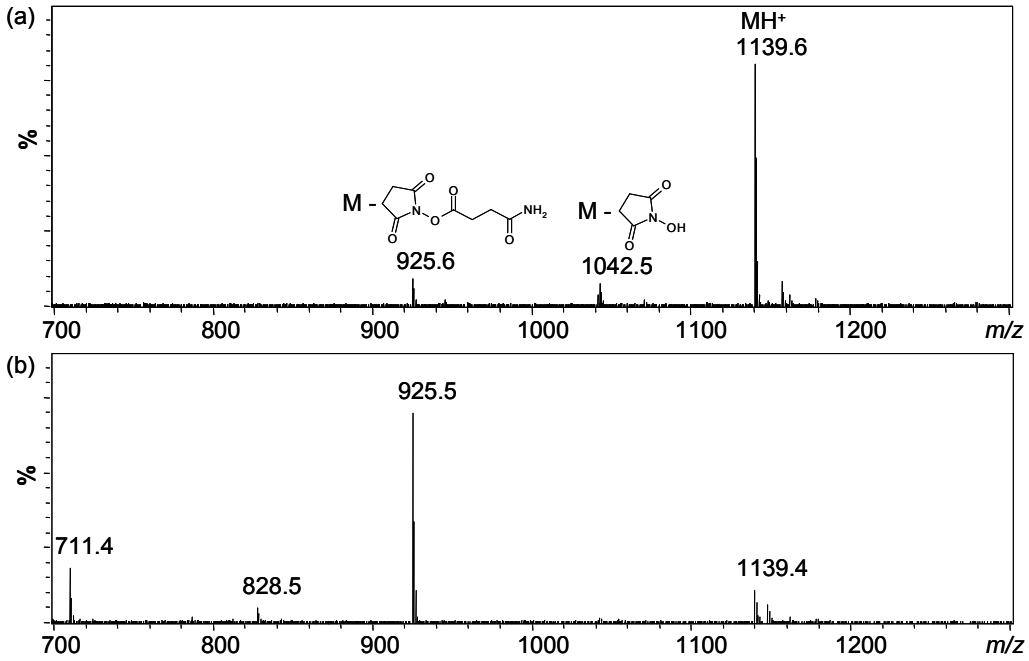



**Figure 3.** ESI mass spectra of the cross-linker. (a) The spectrum was acquired using low capillary exit voltage at 110 v. Most cross-linker stays intact as shown at *m/z* 1139.6. (b) The spectrum was acquired using high capillary exit voltage at 200 v. The cross-linker is fragmented by capillary exit and skimmer dissociation. Reporter ions at *m/z* 828 and 711 are readily generated from the preferred cleavage of the labile bonds.

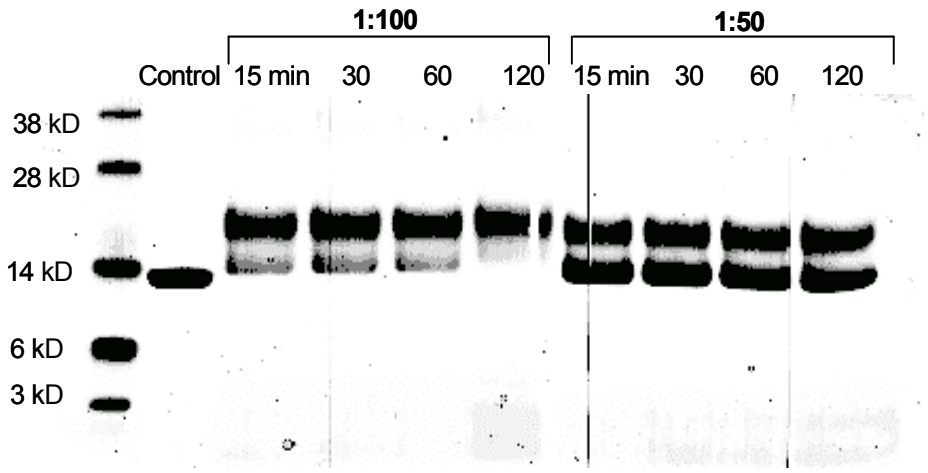



**Figure 4.** SDS-PAGE of cross-linking reaction mixtures.

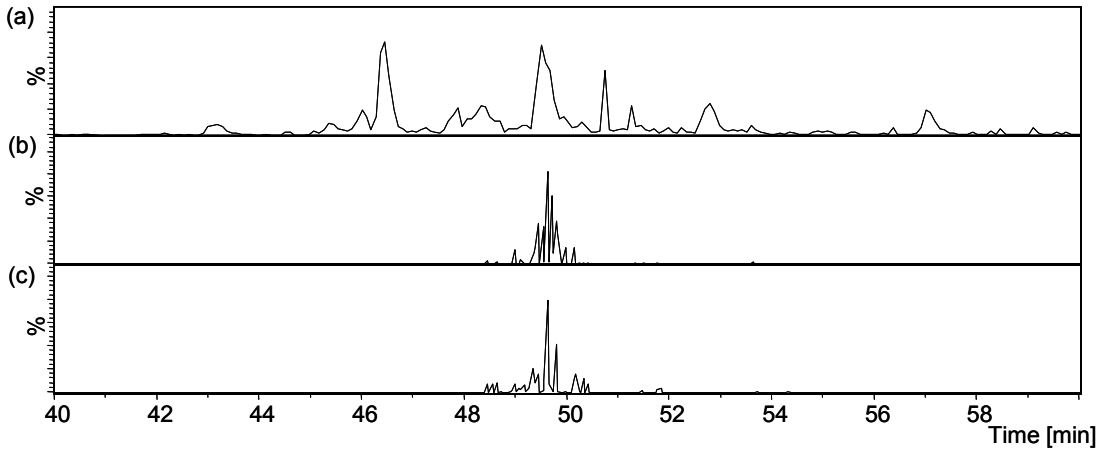



**Figure 5.** Nano-LC/MS/MS of tryptic digest of cross-linked RNase S complex. (a) Base peak chromatogram. Data were filtered to include MS scan only. (b) EIC of ions at *m/z* 828.3. Data were filtered to include MS/MS scan only. (c) EIC of ions at *m/z* 711.3. Data were filtered to include only MS/MS scan.

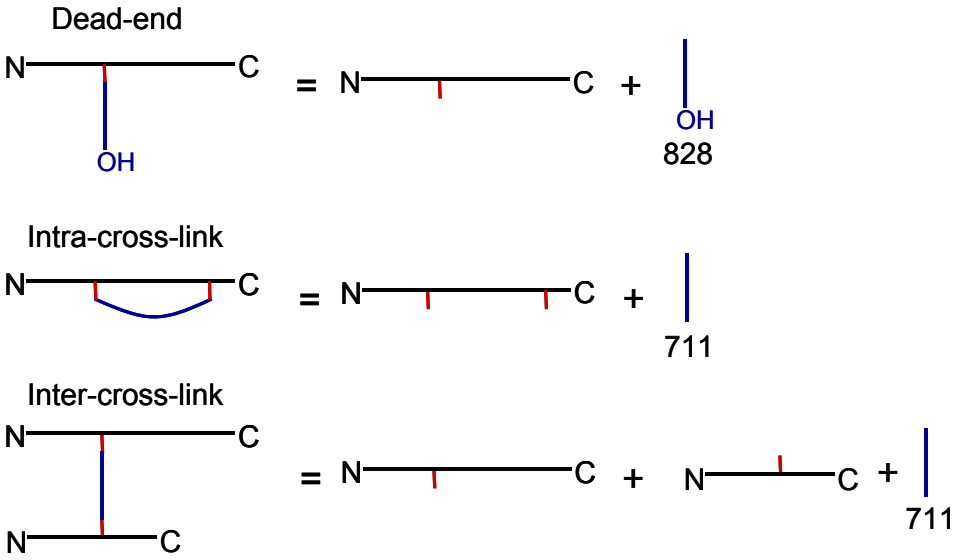



**Figure 6.** The specific fragmentation pattern distinguishes dead-end, intra-, and inter-cross-linked peptides.

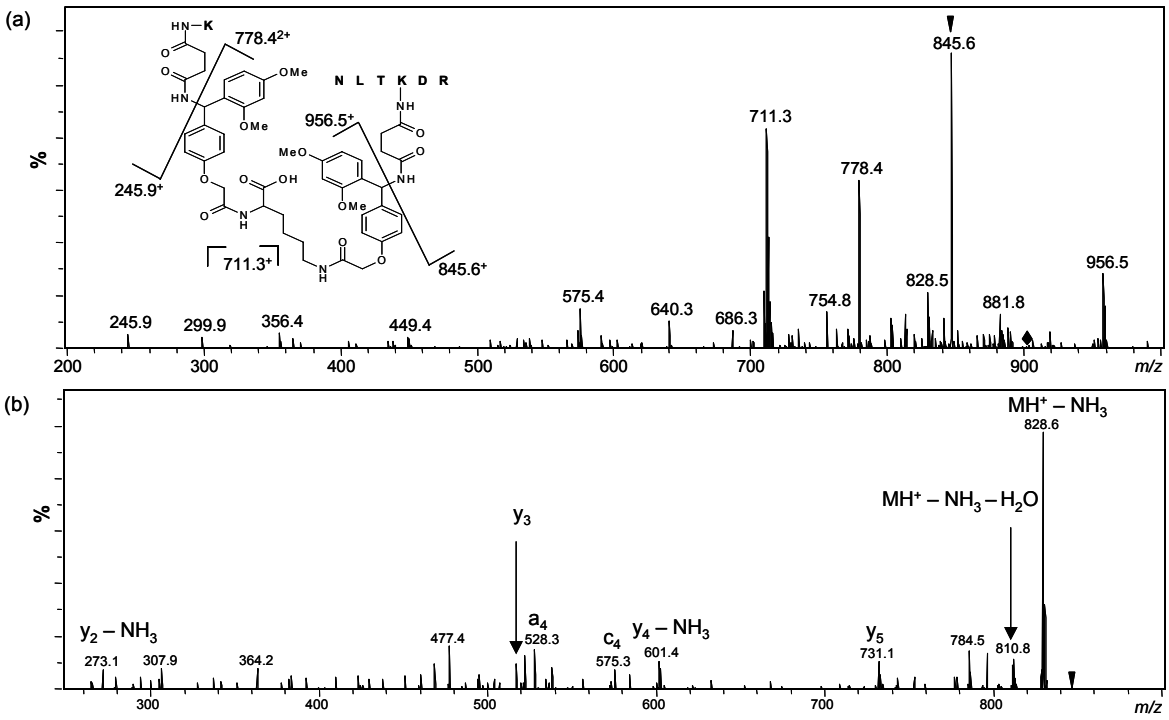
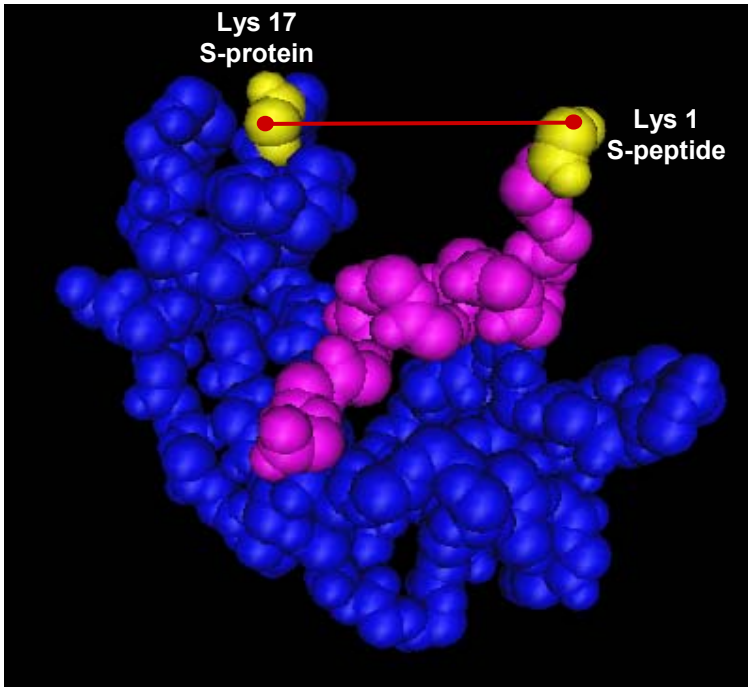



**Figure 7.** Mass spectra of an inter-cross-linked peptides from RNase S complex. (a) MS/MS spectrum of the ions at *m/z* 900.9$^{2+}$ which is labeled with a diamond. The structure of the spacer chain in the cross-linker is fully drawn. Most intense peaks were generated from favorable cleavage of labile bonds as indicated in figure. (b) MS$^3$ spectrum of ions at *m/z* 845.6 which is labeled with a triangle.




S-peptide $^{01}$KETAAAKFERQHMDSSTSAA

S-protein $^{01}$SSSNYCNQMMKSRNLTKDRCKPVNTFVHES
$^{31}$LADVQAVCSQKNVACKNGQTNCYQSYSTMS
$^{61}$ITDCRETGSSKYPNCAYKTTQANKHIIVAC
$^{91}$EGNPYVPVHFDASV

**Figure 8.** X-ray structure and amino acid sequences of RNase S complex. The S-protein domain is marked in blue and the S-peptide domain is marked in pink. The observed cross-linked residues Lys1 of S-peptide and Lys17 of S-protein are indicated in yellow. Cross-link is indicated by red line.

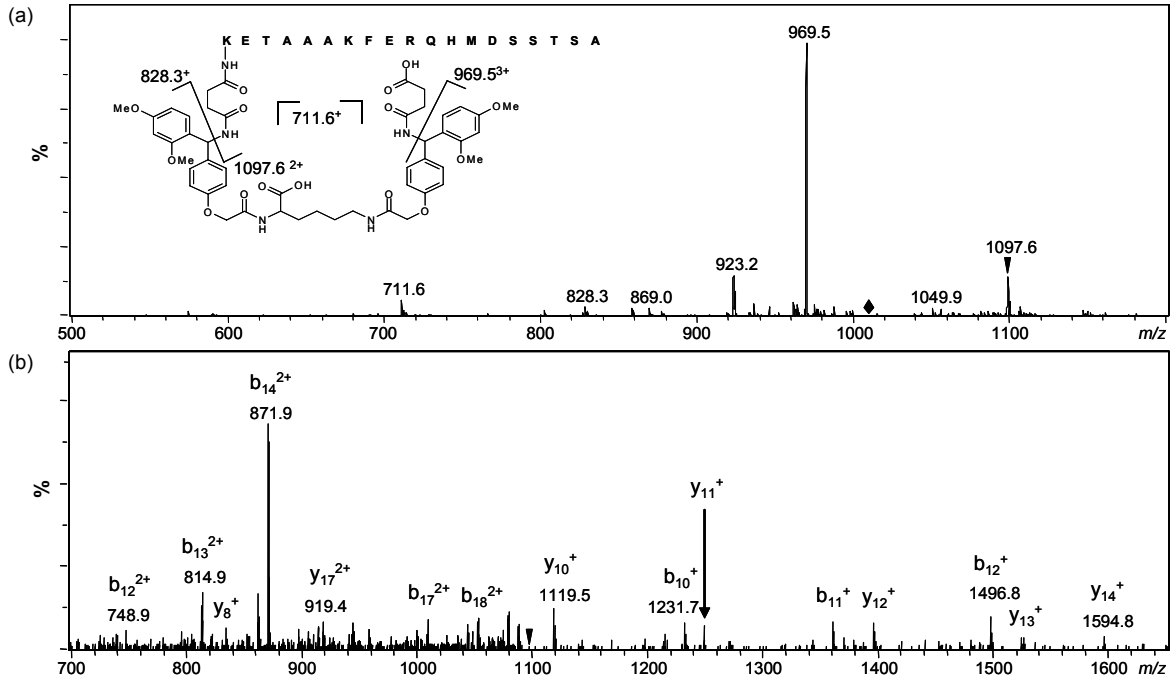



**Figure 9.** Mass spectra of a dead-end modified peptide from RNase S complex. (a) MS/MS spectrum of the ions at *m/z* $1007.3^{3+}$ which is labeled with a diamond. The structure of the spacer chain in the cross-linker is fully drawn. Most intense peaks were generated from favorable cleavage of labile bonds as indicated in figure. (b) $MS^3$ spectrum of ions at *m/z* $1097.6^{2+}$ which is labeled with a triangle.

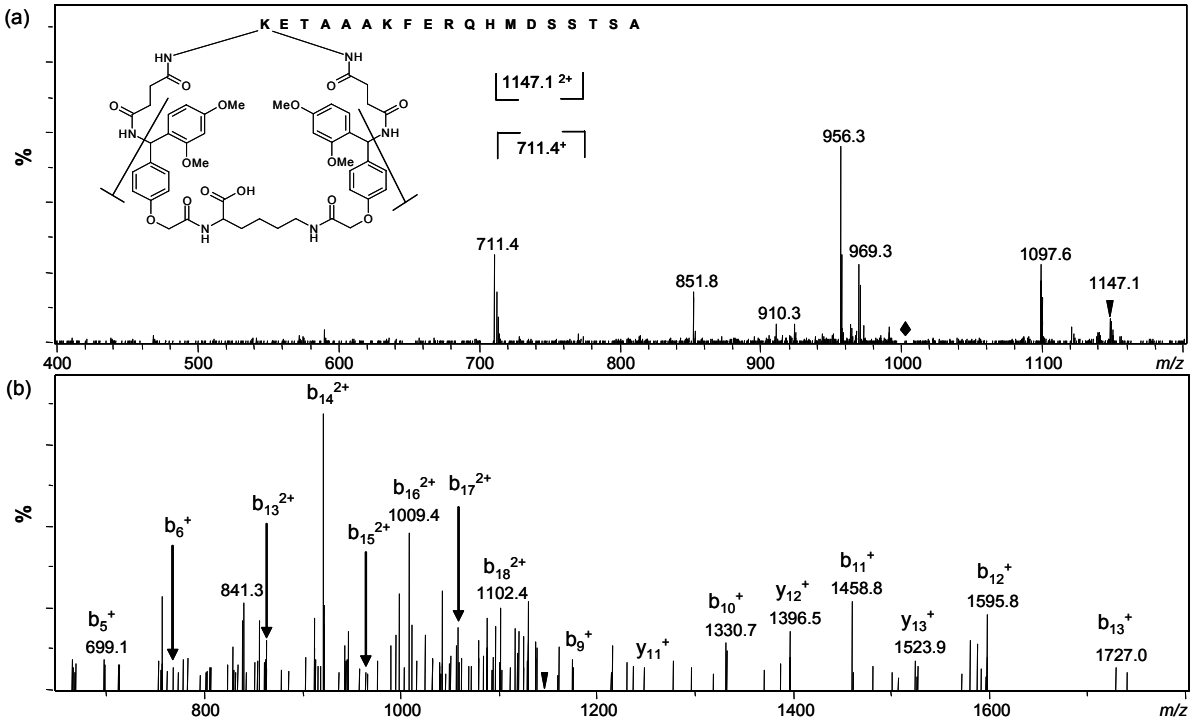



**Figure 10.** Mass spectra of an intra-cross-linked peptide from RNase S complex. (a) MS/MS spectrum of the ions at *m/z* $1001.3^{3+}$ which is labeled with a diamond. The structure of the spacer chain in the cross-linker is fully drawn. Most intense peaks were generated from favorable cleavage of labile bonds as indicated in figure. (b) $MS^3$ spectrum of ions at *m/z* $1147.1^{2+}$ which is labeled with a triangle.

## CHAPTER 3

## Profiling the Membrane Proteome of *Shewanella oneidensis* MR-1 with New Affinity Labeling Probes


### Abstract

The membrane proteome plays a critical role in electron transport processes in *Shewanella oneidensis* MR-1, a bacterial organism that has great potential for bioremediation. Biotinylation of intact cells with subsequent affinity-enrichments has become a useful tool for characterization of the membrane proteome. As opposed to these commonly used, water-soluble commercial reagents, we here introduced a family of hydrophobic, cell-permeable affinity probes for extensive labeling and detection of membrane proteins. When applied to *S. oneidensis* cells, all three new chemical probes allowed identification of a substantial proportion of membrane proteins from total cell lysate without the need for specific membrane isolation. From a total of 384 unique proteins identified, approximately 43% are cell envelop proteins that include outer membrane, periplasmic, and inner membrane proteins. This report demonstrated the first extensive study of membrane proteome with biotinylation of intact cells and revealed many proteins that are involved in metal reduction process in *S. oneidensis*. As a general labeling method, all chemical probes we introduced in this study can be extended to other organisms or cell types and will help expedite the characterization of membrane proteome.

## Introduction

Membranes provide important physical interface between cells and their environment and play crucial roles in fundamental biological processes, such as protein export, electron transmission, signal relay, and cell-cell communications. Proteins spanning the membrane lipid bilayers and those associated with membranes are the essential components regulating these biological functions. Some estimates suggest that approximately 20-30% of all open reading frames encode for membrane proteins.[1,2] And importantly, more than 50% of current targets for therapeutic intervention and drug discovery target are membrane proteins because of their accessibility.[3,4]

The organism of interest for this study is *Shewanella oneidensis* MR-1,[5] a gram-negative facultatively anaerobic bacterium. *S. oneidensis* can use a variety of terminal electron acceptors, including both soluble and insoluble, organic and inorganic compounds, for anaerobic respiration.[6,7] Many of these compounds, especially those with metals such as uranium and chromium, are toxic to humans and other organisms when concentrated in environment. Thus, *S. oneidensis* has great potential for bioremediation because of its ability to reduce and precipitate these toxic compounds from solution and prevent them from spreading. Like other gram-negative bacteria, the cell envelop of *S. oneidensis* includes 3 layers: outer membrane, periplasmic space (peptidoglycan), and inner membrane (plasma membrane). The versatile respiratory capabilities of *S. oneidensis* are resultant from its multi-component branched electron transport system, which includes primarily membrane proteins associated with all three layers of the cell envelop. In particular, those cell surface exposed proteins have direct roles in the metal

reducing activities of this organism. For example, *c*-type cytochromes comprise one of the most important multi-component complexes involved in electron transport in *S. oneidensis*. 80% of the membrane-bound *c*-type heme is localized to the outer membrane[8], indicating the anaerobic respiratory activities of *S. oneidensis* are facilitated by direct contact between surface proteins and extracellular electron acceptors. Therefore, defining these membrane proteins, especially surface-exposed outer membrane proteins, are crucial for better understanding the versatile electron transport mechanisms of this organism.

Although membrane proteins are of considerable importance for fundamental biology, biomedical applications, and bioremediation efforts, the hydrophobic nature and relatively low abundance of most membrane proteins have imposed extreme challenges for proteomic analysis to determine their identity and function. Numerous efforts have targeted these inherent difficulties with membrane proteins. These methods include either enrichment of membrane proteins by subcellular fractionation, e.g., sucrose density centrifugation[9] and differential centrifugation,[10] or increasing solubility of membrane proteins with inclusion of organic acid,[11] organic solvent,[12] or low level detergent[13] that is compatible with subsequent proteolysis and mass spectrometric analysis. 2D SDS-PAGE methods, though effective in separation of complex protein mixtures, have been problematic in analyzing membrane proteins since many are prone to precipitation at their isoelectric point.[14] Replacement of the first dimensional IEF with anion exchange chromatography[15] or with acidic discontinuous PAGE by using cationic detergent (BAC)[16] has shown significant improvement in membrane protein identifications. LC/MS-based shotgun methods,[17] as alternative to gel-based methods, are becoming the

technique of choice for high-throughput and large-scale proteomics analysis. The Yates group demonstrated the first proteome-wide application of the MudPIT method for profiling proteins from isolated yeast membranes.[11] Later, they reported an improved method, using high pH and non-specific proteolysis, for efficient profiling membrane proteins from brain homogenate.[18] Although the large-scale shotgun approaches can effectively identify abundant proteins, low-abundance proteins are often under-represented due to dynamic range limitations.

Alternatively, selective enrichment of membrane proteins by intact cell labeling and subsequent affinity purification can further improve the throughput of detection of low-abundance membrane proteins. With this strategy, many applications have recently shown improved coverage and selectivity in profiling membrane proteome.[19-28] Though it was reported earlier that both water-soluble and insoluble forms of chemical probes label outer membrane, periplasmic, and inner membrane proteins in *E. coli* cells,[29] most recent applications exploited water-soluble forms of chemical probes that are commercially available, such as sulfo-NHS-LC-biotin (primary amine reactive) or iodoacetyl-PEG-biotin (sulfhydryl reactive). Conceptually at least, water-soluble chemical reagent are cell-impermeable and thus, preferentially label cell surface proteins. However, in nearly all the membrane protein profiling studies published to date, a substantial proportion of cytosolic proteins were inevitably observed presumably due to contamination.

In contrast to these previous proteomic affinity labeling methods, we report here our results using novel hydrophobic cell-permeable chemical probes, which were designed and synthesized in house on a modular scaffold, for profiling membrane

proteome of *S. oneidensis*. Our data demonstrated that these new affinity probes allow concurrent labeling of both membrane and soluble proteins. Importantly, a substantial proportion of membrane proteins were identified without the need for tedious membrane fraction preparations. This report presents the first application of affinity labeling method for extensive profiling cell surface and membrane proteins in *S. oneidensis* bacterial cells.

## Experimental Procedures

**Materials.** All chemicals were purchased from Sigma (St. Louis, MO) unless otherwise noted. Fmoc-protected amino acids and HMPB-MBHA resins used for synthesis of affinity chemical probe were purchased from Novabiochem (San Diego, CA). Water used for preparing buffer and solution was 18-MΩ deionized water prepared with a Barnstead Nanopure Water System.

**Synthesis of affinity chemical probes.** Three affinity chemical probes PIR-I, PIR-II, and PIR-III (Figure 1) are synthesized using a 431A Peptide Synthesizer (Applied Biosystem, Foster City, CA, USA) with solid phase peptide synthesis chemistry. Glycine was coupled to HMPB-MBHA (4-hydroxymethyl-3-methoxyphenoxybutyric acid) resin using the standard symmetric anhydride method. For PIR-I, the biotin group was introduced in the form of N-α-Fmoc-N-ε-biotinyl-L-lysine using standard coupling reactions. For PIR-II and PIR-III, the biotin group and PEG linkers were added in the form of Fmoc-Glu (biotinyl-PEG)-OH. For PIR-III, one photo-cleavable group, Fmoc-

aminoethyl photo linker was introduced in between lysine and Glu(biotinyl-PEG)-OH with same solid phase chemistry. The second lysine in the form of N-α,ε-di-Fmoc-L-lysine was coupled to the biotinylated lysine (PIR-I) or the Fmoc-Glu (biotinyl-PEG)-OH (PIR-II and PIR-III) and was then used as the branch point for additional symmetric incorporations. The RINK groups, succinic acids, and N-hydroxysuccinimides (NHS) were then coupled using the same standard activation, coupling and deprotection procedures. It was found that reduced loading of the resin to about one third is necessary to get efficient coupling of the last step, since the biotin group seems to limit the coupling of the Fmoc-lys-Fmoc. The final product was cleaved using either 0.5 % or 1.0 % TFA in chloroform and then neutralized with pyridine. The chloroform and TFA pyridine salts were removed under vacuum until constant weight noted.

**Sample preparation.** *S. oneidensis* MR-1 cells, ATCC 700550, were obtained from American Type Culture Collection (Manassas, VA) and maintained in Luria Bertani (LB) broth. 200 mL sterile LB broth was inoculated with 4 mL of overnight culture in 1 L flask and stirred with a rotary shaker at 100 rpm at room temperature. Cells were harvested at mid-log phase ($OD_{600}$ = 1.2) by centrifuging at 3,200 rpm for 20 min. Pelleted cells were washed three times in 40 mL ice-chilled PBS buffer (150 mM sodium phosphate, 100 mM NaCl, pH 7.5). After final wash, the cells were resuspended in 1 mL PBS and the PIR compound was added to cell suspension to a final concentration of 1 mM. The reactions were carried out at room temperature with shaking for 5 min and then terminated with quenching by 50 μL 1 M Tris buffer (pH 7.5). The labeled cells were washed with PBS buffer 5 times to remove excessive, unreacted PIR compounds and

secreted proteins. The cells were then lysed in 1 mL PBS/2% NP-40 (v/v) by sonication for 2 min. The lysed cells were centrifuged at 15,000 rpm at 4$^{o}$C for 40 min. The pellets were discarded and supernatants were collected in a clean tube. The labeled cell lysates were passed through a Dextran de-salting column (Pierce, Rockford, IL) to further remove unreacted PIR compounds and salts. The labeled proteins were then captured and separated from whole cell lysate by an immobilized monomeric avidin column (Pierce, Rockford, IL) in a 1 mL volume. After incubation of labeled cell lysate in avidin column for 30 min, the column was washed extensively with 1 mL buffer A (150 mM NaCl, 10 mM Tris, 2 mM EDTA, 0.1% NP-40, and 0.1% Triton X-100, pH 7.5) 4 times, 1 mL buffer B (500 mM NaCl, 10 mM Tris, 2 mM EDTA, 0.1% NP-40, and 0.1% Triton X-100, pH 7.5) 3 times, and 1 mL buffer C (10mM Tris, pH 7.5) 3 times. The labeled proteins were then eluted with 200 μL of elution buffer (8 M urea, 2% SDS, and 5 mM d-biotin). The eluent was TCA precipitated to remove salts and detergent. The washed pellet was resuspended in a 100 μL digestion buffer (100 mM $NH_4HCO_3$, pH 7.8, and 8 M urea). Proteins are reduced with DTT at final concentration of 5 mM and alkylated with iodoacetamide at final concentration of 25 mM. Then the solution was diluted 4 times with 100 mM $NH_4HCO_3$ buffer and 20 μg trypsin (Promega, Madison, WI) was added for overnight digestion at 37$^{o}$C. 100 μL of tryptic digest was used for C18 ZipTip® (Millipore, Billerica, MA) to remove salts. The concentrated peptides were finally resuspended in 20 μL 0.1% formic acid for 2D-LC/MS/MS.

**SDS-PAGE and Western blot analysis.** SDS-PAGE and Western blot analysis were used to optimize labeling reaction conditions and biotin-avidin purification conditions.

For SDS-PAGE analysis, the same amount of cell lysate prepared from the equal amount of cells with and without PIR compound labeling after affinity purification were separated by 8% SDS-PAGE and visualized with Coomassie Blue R250 (Bio-Rad, Hercules, CA, USA). For Western blot analysis, proteins were separated by 8% SDS-PAGE gel and then transferred to a nitrocellulose membrane (Whatman, Sanford, ME) with a Trans-Blot® semidry transfer cell (BioRad, Hercules, CA). After transferring, the membrane was blocked with 5% nonfat milk in TBS overnight at 4°C followed by incubation with the primary antibody anti-biotin (Sigma, St. Louis, MO) at 1:5000 dilution for an hour at room temperature. And finally the membrane was probed by HRP-conjugated secondary antibody and chemiluminescence peroxidase substrate (Sigma, St. Louis, MO).

**Electron microcopy (EM) cell imaging.** *S. oneidensis* MR-1 cells in PBS suspension with or without compound PIR labeling were loaded onto the carbon-coated nickel grids for immunogold EM detection of surface proteins labeling. Fixation buffer (2% paraformaldehyde, 1.25% glutaraldehyde, 0.05 M phosphate buffer, pH 7.2) was used to incubate with sample grids for 15 min with subsequent PBS washing 2 times. The residual aldehyde was then quenched by incubation with 0.02 M glycine in PBS for 5 min. Anti-biotin monoclonal antibody (Sigma, St. Louis, MO) was added in PBS-BSA (20 mM phosphate, 500 mM NaCl, 0.5% BSA, 0.1% Gelatin, 0.05% Tween20, pH 7.4) at 1:25 ratio for primary antibody incubation for 10 min with occasional shaking at room temperature. After washing with PBS-BSA 3 times, 10 nm colloidal gold particle-conjugated with goat-anti-mouse IgGs (Sigma, St. Louis, MO) at 1:50 ratio in PBS-BSA

was added for secondary antibody incubation for 30 min. Cells were washed 3 times with PBS-BSA before the final fixation of cells and antibodies using 1% glutaraldehyde in PBS for 15 min. The fixatives were removed by buffer (20 mM phosphate, 150 mM NaCl, 4% BSA, 2 mM $NaN_3$, pH 7.4). The sample grids were examined by EM using JEOL 1200 EX transmission electron microscope after staining with uranyl acetate for contrast enhancement. To ascertain membrane permeability of the PIR compound, the immunogold EM assay was also carried out using thin sections of *S. oneidensis* MR-1 cells. The labeled and unlabeled cells were treated with a fixation buffer (2% paraformaldehyde, 1.25% glutaraldehyde, 0.05 M phosphate buffer, pH 7.2) for an hour. Sample dehydration was performed in a series of graded ethanol treatment steps (10%, 25%, 50%, 75%, 90%, and 100%) for consecutive overnight periods. The dehydrated samples were embedded in a LR White resin for infiltration and polymerization. The thin sections were generated with a diamond knife mounted in an ultracut microtome for loading onto carbon-coated nickel grids. First, the grids were floated on a drop of water for 10 min. PBS-BSA was then incubated with sample grids for 5 min to block non-specific binding. A solution of 1:25 diluted anti-biotin monoclonal antibody was added in PBS-BSA for an hour incubation with section grids. After rinsing with PBS-BSA 3 times, 10 nm colloidal gold particle labeled goat-anti-mouse IgGs at 1:50 dilution ratio were incubated with sample grids for 30 min. Sections were then washed with PBS 3 times and post-fixed by 1% glutaraldehyde in PBS for 3 min. Deionized water droplets were used to rinse the section grids for 10 min before EM examination. Uranyl acetate was also used to increase the contrast of section images. Negative controls include the

unlabeled cells and the labeled cells without primary antibody incubation. JEOL 1200 EX transmission electron microscope was utilized to perform EM imaging.

**2D-LC/MS/MS**. Data-dependent nano 2D-LC/MS/MS was performed using ion trap mass spectrometer (Esquire HCT, Bruker Daltonics, Billerica, MA) equipped with a nano-ESI source and nano-HPLC systems (Ultimate, Dionex, Sunnyvale, CA). For 2D-LC/MS/MS, a strong cation exchange (SCX) cartridge (500 μm × 15 mm, 5 μm, Dionex) was used for the first dimension separation and a nano C18 reversed phase column (C18 PepMap, 75 μm × 150 mm, 3 μm, 100 Å, Dionex) was used for the second dimension separation. 10 μL salt plug (10, 25, 50, 100, 250, 500, and 1000 mM ammonium acetate in 0.1% formic acid and 2% acetonitrile) was used for each salt step elution of the SCX cartridge. Samples and salt plugs were injected to the SCX cartridge by autosampler and the eluent from SCX cartridge was trapped in a micro pre-column (C18 PepMap, 300 μm × 1 mm, 5 μm, Dionex). The pre-column was washed with solvent A (0.1% formic acid, 2% acetonitrile in water) at a flow rate of 50 μL/min for 3 min and then peptides were eluted at a flow rate of 200 nL/min to the reversed phase nano-column and separated using the following gradient: 0% B for 0-3 min, 0-15% B for 3-15 min, 15-25% B for 15-60 min, 25-40% B for 60-90 min, 95% B for 90-105 min, and 0% B for 105-120 min. Solvent B was 0.1% formic acid in 95% acetonitrile.

**Data analysis.** Bruker Daltonics *DataAnalysis* software (version 3.1) and *BioTools* software (version 2.2) were used for data analysis and protein identification. A data processing script provided by Bruker Daltonics was used to generate one .mgf file

compiled from the entire 2D-LC/MS/MS run. The .mgf file was then used to search database using the Mascot program (version 2.1.0, MatrixScience Ltd, London) licensed in house. The database containing the complete genome of *S. oneidensis* MR-1 (4,854 ORFs) was downloaded in FASTA file format from The Institute for Genomic Research (TIGR) (http://www.tigr.org/). Search parameters were set as follows: enzyme, trypsin; allowed miss cleavage up to 3; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M); peptide tolerance, 3 Da; and MS/MS tolerance, 0.8 Da. Auto-hits option was selected to allow reporting all the protein hits with the probability-based MOWSE scores that exceeded their thresholds ($p < 0.05$), indicating significance at the 95% confidence level.[30] The output of search result was further filtered using more stringent MudPIT scoring and ions score cut-off of 0.1. Those proteins identified with more than 2 peptides which have minimal peptide ion score > 30 were accepted as true identifications without additional manual spectral inspection. For those proteins identified with a single peptide, if the expect value (E) is less than 0.01 (99% confidence level), the protein hit is accepted as true identification without further manual inspection; if E is between 0.01 and 0.05, the protein hit was accepted only when > 3 consecutive y or b ions were observed in the MS/MS spectrum; Any other single peptide hits with E > 0.05 were discarded.

## Results

### Features of new affinity chemical probes

We previously reported a new class of chemical cross-linkers, which we call protein interaction reporters (PIRs),[31] for study of protein-protein interactions. The name PIR was retained for the current affinity chemical probes, since they are the same family of compounds designed and synthesized on the same modular scaffold using solid-phase Fmoc peptide synthesis chemistry. The structure of the PIR compound can be readily modified with incorporation of various building blocks. For current goal of intact cell labeling and affinity purifications, two N-hydroxysuccinimide (NHS) esters as reactive groups and a biotin group as affinity tag were included in the structure (Figure 1). Aside from these common features that can be used as affinity labeling reagents, these compounds include additional properties such as low-energy mass spectrometry cleavable bonds (highlighted by the red dashed lines in Figure 1) and cross-linking capabilities provided by the double NHS groups for protein-protein interaction studies. A hydrophilic side chain PEG group was incorporated in compound PIR-II and PIR-III to provide better solubility. PIR-III includes an additional photo-cleavable group which could be used for elution of captured biotinylated proteins from avidin beads by breaking the photo-cleavable bond. These additional features for compound PIR-I, II, and III were evaluated and reported elsewhere.(manuscript submitted to Analytical Chemistry) In the present study, our efforts were focused on exploring these compounds as affinity labeling probes for profiling cell envelop proteins in *Shewanella oneidensis* MR-1. In general all three compounds are hydrophobic because of multi-aromatic rings and NHS ester groups. PIR-II and PIR-III have relatively better solubility than PIR-I due to incorporation of a hydrophilic side chain. In addition, all three new probes are relatively large as compared to the commercial water-soluble labeling reagent, sulfo-NHS-LC-biotin (MW 556.59), a

most commonly used chemical probe for characterizing membrane proteins. Despite the fact that our compounds possess features very different from the commercial, water-soluble labeling reagents, the present study demonstrated all three compounds can effectively label and enrich surface and membrane proteins. PIR-I was primarily used for most studies. The results from initial on-cell applications of PIR-II and III were also presented.

**Optimizing cell labeling with PIR-I**

SDS-PAGE and anti-biotin Western blot analysis were used for optimization of PIR-I labeling conditions and affinity purification conditions, which included compound concentration, pH of reaction buffer, reaction time, washing buffer, number of washes, lysis buffer, elution buffer used for dissociation of biotinylated proteins from the monomeric avidin column, etc. Based on observations of labeled protein intensities on SDS-PAGE and Western blot images as compared to control, we obtained the optimal conditions as follows: 1 mM PIR-I reaction concentration, 5-min reaction time at room temperature, reaction buffer using PBS buffer (pH 7.2), 5 washes of labeled cells in PBS buffer prior to cell lysis, lysis buffer including 2% NP-40 in PBS, and elution buffer for affinity purification using 8 M urea/2% SDS/5 mM d-biotin. Figure 2 shows selected results from optimization experiments. Many proteins were consistently observed for PIR-I labeled samples and not observed in parallel control samples using both SDS-PAGE (Figure 2a and 2b) and Western blot analysis (Figure 2c). Three major bands observed in the control lane (highlighted in white box in Figure 2b) were identified to be acetyl-CoA carboxylase multienzyme complex (SO0840) (168 kD), acetyl-CoA

carboxylase (SO1894) (76 kD), and translation elongation factor Tu (SO0217) (44 kD) with in-gel digestion and LC/MS/MS analysis. The first two proteins are endogenously biotinylated proteins which can serve as good internal control for affinity purification experiments. Protein translation elongation factor is one of the most abundant proteins in *S. oneidensis* cells, which was observed as contamination by others as well.[32] More proteins were labeled with the use of 1 mM PIR-I than 0.1 mM PIR-I (Figure 2a). Further increasing PIR-I concentration did not show further improvement. Figure 2c shows the Western blot results of labeled cell lysate (1 mM PIR-I was used) at two different reaction times versus control cell lysates, which indicates that the labeling reaction approached completion in 5 min and further extending reaction time to 30 min did not produce more labeled proteins. We also characterized the rates of hydrolysis of the PIR-I compound at pH 3.0 and pH 7.2 and found the half-life was 8-10 hr and 10 min, respectively. Thus, we further performed labeling experiments by adding PIR-I stock solution in DMSO multiple times to the cell suspension at another 5-min reaction time for each addition. However, multiple addition experiments did not generate significantly more labeled proteins (Figure 2b), indicating the exposed sites that can be labeled by PIR-I are close to saturation from a single reaction period. In addition, we performed anti-biotin Western blot analysis to investigate if PIR-labeled proteins were present in the PBS labeling buffer and in subsequent washing buffers after on-cell labeling. Western blot analysis of the washing solutions showed that significant levels of anti-biotin reactivity were observed in the first couple of washes after labeling. This could be the result of cell lysis during labeling or from active protein secretion by *S. oneidensis* cells. However, by the 3rd wash most reactivity on the Western blot was eliminated (Figure

2d). Thus, at least 5 washing steps were taken prior to cell lysis to reduce this contamination.

**Cell permeability of PIR-I**

It is believed that water-soluble labeling reagents are cell-impermeable, and therefore should react primarily with cell surface proteins. However, in these studies which employed water-soluble labeling probes, a significant proportion of cytosolic proteins (50% in average) were often observed along with membrane proteins regardless of subcellular membrane enrichments and stringent washes. It was thought these cytosolic proteins could come from contamination[22] or non-viable cells.[28] We do not eliminate these possibilities for our case although we washed labeled cells extensively prior to cell lysis (Figure 2d). However, our PIR compound is cell membrane-permeable due to its hydrophobic nature and the identification of some cytosolic proteins is the result of labeling inside the cells. To verify this, we carried out electron microscopy (EM) imaging analysis of labeled cells. These experiments included both whole cell EM imaging and thin section EM imaging to determine if proteins on the cell membranes and in the cytoplasm were labeled with PIRs. Figure 3 illustrates EM images generated from *S. oneidensis* cells that were reacted with PIR and from control cells that were not treated with PIR. Anti-biotin primary antibodies and immunogold nanoparticle-conjugated secondary antibodies were used to determine if proteins were effectively labeled with PIR. Only those proteins that are PIR-labeled or endogenously biotinylated are expected to react with the anti-biotin antibody and immunogold nanoparticles to result in detected signals (dark dots in the images). Since the thin sections are much smaller than the

average cell dimensions, these experiments allow the interrogation of the inside of the cells. As shown in Figure 3, the control cells exhibited very little reactivity with the immunogold nanoparticles, while the PIR-treated cells showed significant levels of reactivity on both exposed cell surfaces and thin-sections. An additional control experiment performed with PIR-labeled cells that involved the same immunogold nanoparticle treatment, but excluded the anti-biotin antibody treatment, resulted in no observed reactivity. Therefore, these data indicate that PIR-I is able to penetrate *S. oneidensis* cell membranes and label both cell surface proteins and cytosolic proteins in intact cells.

**Protein identification by 2D-LC/MS/MS**

After cell labeling using PIR-I with optimized conditions, tryptic proteolysis was performed with the enriched, biotinylated proteins from unfractionated total cell lysate. The tryptic digest was subsequently analyzed by automated 2D-LC/MS/MS with a total run time of ~20 h. The peak lists from each salt step elution were extracted at the threshold of 30,000 counts. Then the peak lists from all salt steps in a single 2D-LC/MS/MS analysis were compiled to give rise to final composite peak lists in a single .mgf text file, which was used for Mascot search. All proteins with probability-based MOWSE scores that exceeded their thresholds ($p < 0.05$) were automatically reported in the search result. However, we further filtered the search result with more stringent MudPIT scoring and using ions score cut-off of 0.1, which will remove all the peptides with expect value > 0.1. Furthermore, those protein hits that were identified by a single peptide with expect value > 0.05 were removed. In the case of single peptide

identification with expect value > 0.01, its MS/MS spectrum was manually checked to ensure that at least 3 consecutive y- or b- ions were observed. Thus, a total number of 341 proteins were identified with PIR-I labeling experiments from three independent sample preparations (three different batches of cell culture). Initial attempts of labeling cells with PIR-II and PIR-III compounds using the conditions optimized for PIR-I also resulted in significant amount of proteins identified from a single sample preparation, which were 187 and 163, respectively. Overall, 384 unique proteins were identified from all three compound labeling experiments. These results are summarized according to protein identification guidelines suggested by Bradshaw et al.[33]

**Characterizing the identified proteins**

We first analyzed all 384 identified proteins based on their subcellular locations. PSORTb v.2.0[34] (http://www.psort.org/psortb/index.html) and CELLO v.2.5[35] (http://cello.life.nctu.edu.tw/) programs were used to help classify protein subcellular locations. The pie chart representation of the distribution of all 384 identified proteins in *S. oneidensis* cells is shown in Figure 4. Among the total identified proteins, 165 (43%) proteins are the constituents of cell envelop proteins including: 60 (16%) outer membrane proteins, 35 (9%) periplasmic proteins, 65 (17%) inner membrane proteins, and 6 (2%) extracellular proteins. Of the cell envelop proteins, 58 (35%) proteins are predicted to be integral membrane proteins with transmembrane domains ranging from 1 to 15 by using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).[36] Cytochrome *c*-type biogenesis protein CcmF (ccmF-1) (SO0266) has 15 TMDs and it was identified with 2 unique peptides. It should be noted that TMHMM program can only predict

integral proteins that have α-helices. However in gram-negative bacteria, the integral inner membrane proteins primarily consist of α-helices while integral outer membrane proteins generally have multiple β-strands which form β-barrels.[37,38] Many proteins identified from our PIR labeling experiments are known to have transmembrane β-barrels such as OmpA family protein[39] (SO3545) and outer membrane porin proteins[40] (SO1821) and (SO3896). Therefore, the total number of identified integral membrane proteins is underrepresented. In summary, a select subset of cell envelop proteins is presented in Table 1. The full list of characterization of all protein identified is provided in Supplementary Table 1. Other proteins include 41% cytosolic proteins, 4% membrane-associated proteins, and 12% unclassified proteins. Out of all the identified cytosolic proteins, some may be due to contamination of abundant proteins as also reported by others, but a majority of them are the result of PIR labeling as shown in Figure 2d.

We further grouped the proteins identified in this study in their corresponding functional categories using the "clusters of orthologous groups of proteins" (COG) classification systems (http://www.ncbi.nlm.nih.gov/sutils/coxik.cgi?gi=261).[41,42] The pie chart diagram of COG classification of all 384 proteins is shown in Figure 5. We have identified proteins in all functional categories including information storage and processing, cellular processes, and metabolism. The identified proteins involved in translation (J) are in the largest group which includes 61 proteins, corresponding to 16% of the total identified proteins. The second largest group was energy production and conversion (C) consisting of 54 proteins (14% of total identified proteins). The proportions of identified proteins in group J and C both significantly exceed their corresponding proportions in entire genome of *S. oneidensis*, which are 5% and 4%,

respectively. This could be due to high abundance of proteins in these two groups. However, approximately three-fold enrichment of these two groups of proteins also suggests their high accessibility to our labeling probes. Another major category of proteins is cell wall/membrane biogenesis (M) including approximately 7% of total identified proteins (27 proteins), which is approximately twice as high as its corresponding percentage in the whole genome (~4%). Cell wall and biogenesis proteins are primarily composed of cell envelop proteins, thus these appear highly accessible to labeling reagents.

**Discussion**

Biotinylation of exposed residues in proteins has become a useful tool for enriching and profiling low-abundance membrane proteins. Water-soluble reagents are commonly employed in most applications since presumably these reagents are cell-impermeable and thus, only react with exposed regions of proteins on cell surface. However, in addition to successful identification of large proportion of cell surface proteins with water-soluble reagents, significant amount of cytosolic proteins are also observed in spite of subcellular membrane enrichment and stringent washes. In this report, we present the first application of hydrophobic labeling probes for extensive profiling membrane proteins. All three biotin-tagged probes used in this research allow efficient labeling and enrichment of a large fraction of membrane proteins with minimal need for sample preparation. In the present study, unfractionated total cell lysate was directly subjected to affinity purification to enrich biotinylated proteins after intact cell labeling. In summary,

we identified 384 unique proteins in *S. oneidensis* with the use of three PIR compounds. Of these, approximately 43% of proteins are cell envelop proteins located in all three layers of the cells. In particular, out of 166 cell envelop proteins identified, approximately 40% are extracellular and outer membrane proteins. Since all three PIR compounds are highly hydrophobic, we speculate after cell penetration, these compounds probably stay close to the lipid bilayers of the membranes, or sub-organelle membranes rather than the hydrophilic environment of the cytoplasm. Thus it is very likely that these compounds preferentially react with membrane proteins or proteins associated with membranes. Another point worth mentioning is that the three PIR probes used in this study do have slightly different structures and hydrophobicities, which could lead to different labeling selectivity. However in the current study the labeling differences of these three compounds are not compared and discussed since the labeling conditions are not optimized for PIR-II and PIR-III and variations could occur from different batches of cell cultures as observed by us and others.[25,28]

It is believed that cell envelop proteins, especially the cell surface proteins, play a critical role in electron transfer and metal reduction in *S. oneidensis*. Recently, Giometti et al[43] and Ruebush et al[44] reported the results from 2D-PAGE analysis of isolated membrane fractions in *S. oneidensis*. Nearly all the membrane proteins observed in their studies were repeatedly observed in our study as well. For example, among the 21 membrane proteins reported in Ruebush's study[44], we observed 18 of them such as outer membrane precursor MtrB (SO1776), TonB-dependent receptor (SO2907), OmpA family protein (SO3545), and peptidase M13 family (SO0429). In addition, all three membrane proteins in the membrane fraction reported in Giometti's study,[43] i.e., periplasmic nitrate

reductase (napA) (SO0848), periplasmic glucan biosynthesis protein (SO1051), and fumarate reductase (SO0790), were also observed in our study with high Mascot scores. In addition, we have identified a large number of membrane proteins that are ascribed to energy production and conversion category by COG, such as cytochrome *c* oxidase (SO2361 and SO 2363), ATP synthase (SO4747, SO4748, and SO4749), NADH: ubiquinone oxidoreductase (SO0904, SO1103, and SO1108), ABC transporter (ATP-binding protein) (SO3779 and SO3780), etc. Other proteins we observed such as outer membrane cytochromes OmcA (SO1779) and OmcB (SO1778) have been reported to be exposed on the outer membrane surfaces to allow direct electron transfer to extracellular electron acceptors;[45] and outer membrane protein MtrB (SO1776) is required for proper localization of OmcA and OmcB to the outer membrane of the cells.[46] Very recently Shi et al reported that OmcA and OmcB form a complex to facilitate electron transport in *S. oneidensis*.[47] In our experiments using SDS-PAGE separation of affinity purified, PIR labeled cell lysate, we repeatedly observed high MW (> 250 kD) bands as highlighted in the white box in Figure 2b. In-gel digestion and LC/MS/MS analysis identified both OmcA (MW 80 kD) and OmcB (MW 73 kD) proteins in this band. Since our PIR probes do possess cross-linking capabilities, observation of both OmcA and OmcB in the high MW band indicates cross-linking of these two interacting proteins as well as other proteins in the complex. However, further experiments need to be performed to validate this result. Taken together, biotinylation of intact cells with our compounds showed great promise for efficient labeling cell surface proteins and membrane proteins. To our knowledge, this report represents the most extensive profiling of the membrane proteome of *S. oneidensis* cells.

In addition to membrane protein labeling, the cell-permeable nature of these PIR probes allows concurrent labeling proteins in the cytoplasm as verified by immuno-gold EM experiments (Figure 3d). In our study, approximately 41% of the identified proteins are cytosolic proteins. Some of these cytosolic proteins are probably the results of contamination, as also observed by others. However we believe a majority of these cytosolic proteins are the consequences of labeling reactions that occurred after PIR compounds penetrating to the cells. As mentioned above, we speculate that after penetration, most PIR compounds should stay close to the inner surface of the inner membranes or sub-organelle membranes such as endoplasmic reticulum membranes instead of hydrophilic environment of cytoplasm. Thus, labeling soluble proteins is most likely to occur when cytosolic proteins are in proximity to membranes or when they interact with membranes. For example, we have identified ~63% (34 identified, 54 total in entire genome) of the ribosomal proteins in *S. oneidensis*. It is known that ribosome interacts with membrane-bound receptors to facilitate translocation of membrane proteins and secetory proteins.[48]

In summary, we present the first demonstration of extensive analysis of membrane proteins using water-insoluble labeling probes. With application of these hydrophobic, novel chemical probes to *S. oneidensis* cell systems, we identified large proportion of cell envelop proteins with minimal sample preparation requirement. These probes can be extendable to other types of cells and organisms, or a particular subgroup of proteins of interest such as inner membrane proteins and outer membrane proteins, with the aid of subcellular fractionation. Moreover, these chemical probes can be

modified to include stable isotope for quantitative measurement of changes occurring on cell surfaces and membranes with response to specific perturbations.

**Acknowledgements**

We thank Nikola Tolic (PNNL) for support of data sorting and formatting. This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

**References**

(1) Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **1998***, 7*, 1029-1038.

(2) Stevens, T. J.; Arkin, I. T. Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **2000***, 39*, 417-420.

(3) Reiss, T. Drug discovery of the future: the implications of the human genome project. *Trends Biotechnol* **2001***, 19*, 496-499.

(4) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat Rev Drug Discov* **2002***, 1*, 727-730.

(5) Venkateswaran, K.; Moser, D. P.; Dollhopf, M. E.; Lies, D. P.; Saffarini, D. A. et al. Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov. *Int J Syst Bacteriol* **1999***, 49 Pt 2*, 705-724.

(6) Beliaev, A. S.; Thompson, D. K.; Khare, T.; Lim, H.; Brandt, C. C. et al. Gene and protein expression profiles of Shewanella oneidensis during anaerobic growth with different electron acceptors. *Omics* **2002**, *6*, 39-60.

(7) Heidelberg, J. F.; Paulsen, I. T.; Nelson, K. E.; Gaidos, E. J.; Nelson, W. C. et al. Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. *Nat Biotechnol* **2002**, *20*, 1118-1123.

(8) Myers, C. R.; Myers, J. M. Localization of cytochromes to the outer membrane of anaerobically grown Shewanella putrefaciens MR-1. *J Bacteriol* **1992**, *174*, 3429-3438.

(9) Blonder, J.; Terunuma, A.; Conrads, T. P.; Chan, K. C.; Yee, C. et al. A proteomic characterization of the plasma membrane of human epidermis by high-throughput mass spectrometry. *J Invest Dermatol* **2004**, *123*, 691-699.

(10) Simpson, R. J.; Connolly, L. M.; Eddes, J. S.; Pereira, J. J.; Moritz, R. L. et al. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. *Electrophoresis* **2000**, *21*, 1707-1732.

(11) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19*, 242-247.

(12) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D. et al. Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2002**, *1*, 351-360.

(13) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **2001***, 19*, 946-951.

(14) Bunai, K.; Yamane, K. Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005***, 815*, 227-236.

(15) Schluesener, D.; Fischer, F.; Kruip, J.; Rogner, M.; Poetsch, A. Mapping the membrane proteome of Corynebacterium glutamicum. *Proteomics* **2005***, 5*, 1317-1330.

(16) Zahedi, R. P.; Meisinger, C.; Sickmann, A. Two-dimensional benzyldimethyl-n-hexadecylammonium chloride/SDS-PAGE for membrane proteomics. *Proteomics* **2005***, 5*, 3581-3588.

(17) MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M. et al. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **2002***, 99*, 7900-7905.

(18) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003***, 21*, 532-538.

(19) Sabarth, N.; Lamer, S.; Zimny-Arndt, U.; Jungblut, P. R.; Meyer, T. F. et al. Identification of surface proteins of Helicobacter pylori by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. *J Biol Chem* **2002***, 277*, 27896-27902.

(20) Shin, B. K.; Wang, H.; Yim, A. M.; Le Naour, F.; Brichory, F. et al. Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J Biol Chem* **2003**, *278*, 7607-7616.

(21) Jang, J. H.; Hanash, S. Profiling of the cell surface proteome. *Proteomics* **2003**, *3*, 1947-1954.

(22) Zhang, W.; Zhou, G.; Zhao, Y.; White, M. A. Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* **2003**, *24*, 2855-2863.

(23) Zhao, Y.; Zhang, W.; White, M. A. Capillary high-performance liquid chromatography/mass spectrometric analysis of proteins from affinity-purified plasma membrane. *Anal Chem* **2003**, *75*, 3751-3757.

(24) Goshe, M. B.; Blonder, J.; Smith, R. D. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J Proteome Res* **2003**, *2*, 153-161.

(25) Chen, W. N.; Yu, L. R.; Strittmatter, E. F.; Thrall, B. D.; Camp, D. G., 2nd et al. Detection of in situ labeled cell surface proteins by mass spectrometry: application to the membrane subproteome of human mammary epithelial cells. *Proteomics* **2003**, *3*, 1647-1651.

(26) Gauthier, D. J.; Gibbs, B. F.; Rabah, N.; Lazure, C. Utilization of a new biotinylation reagent in the development of a nondiscriminatory investigative approach for the study of cell surface proteins. *Proteomics* **2004**, *4*, 3783-3790.

(27) Zhao, Y.; Zhang, W.; Kho, Y. Proteomic analysis of integral plasma membrane proteins. *Anal Chem* **2004**, *76*, 1817-1823.

(28) Nunomura, K.; Nagano, K.; Itagaki, C.; Taoka, M.; Okamura, N. et al. Cell surface labeling and mass spectrometry reveal diversity of cell surface markers and signaling molecules expressed in undifferentiated mouse embryonic stem cells. *Mol Cell Proteomics* **2005**, *4*, 1968-1976.

(29) Bradburne, J. A.; Godfrey, P.; Choi, J. H.; Mathis, J. N. In vivo labeling of Escherichia coli cell envelope proteins with N-hydroxysuccinimide esters of biotin. *Appl Environ Microbiol* **1993**, *59*, 663-668.

(30) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.

(31) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem* **2005**, *77*, 311-318.

(32) Mayer, M. U.; Shi, L.; Squier, T. C. One-step, non-denaturing isolation of an RNA polymerase enzyme complex using an improved multi-use affinity probe resin. *Molecular BioSystems* **2005**, *1*, 53-56.

(33) Bradshaw, R. A.; Burlingame, A. L.; Carr, S.; Aebersold, R. Protein Identification: The Good, the Bad, and the Ugly. *Mol Cell Proteomics* **2005**, *4*, 1221-1222.

(34) Gardy, J. L.; Laird, M. R.; Chen, F.; Rey, S.; Walsh, C. J. et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **2005**, *21*, 617-623.

(35) Yu, C. S.; Lin, C. J.; Hwang, J. K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* **2004**, *13*, 1402-1406.

(36) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001**, *305*, 567-580.

(37) Schulz, G. E. [beta]-Barrel membrane proteins. *Current Opinion in Structural Biology* **2000**, *10*, 443-447.

(38) Tamm, L. K.; Arora, A.; Kleinschmidt, J. H. Structure and Assembly of beta -Barrel Membrane Proteins. *J. Biol. Chem.* **2001**, *276*, 32399-32402.

(39) Pautsch, A.; Schulz, G. E. High-resolution structure of the OmpA membrane domain. *Journal of Molecular Biology* **2000**, *298*, 273-282.

(40) Saier, M. H., Jr. A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 354-411.

(41) Tatusov, R. L.; Koonin, E. V.; Lipman, D. J. A genomic perspective on protein families. *Science* **1997**, *278*, 631-637.

(42) Tatusov, R.; Fedorova, N.; Jackson, J.; Jacobs, A.; Kiryutin, B. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **2003**, *4*, 41.

(43) Giometti, C. S.; Khare, T.; Tollaksen, S. L.; Tsapin, A.; Zhu, W. et al. Analysis of the Shewanella oneidensis proteome by two-dimensional gel electrophoresis under nondenaturing conditions. *Proteomics* **2003**, *3*, 777-785.

(44) Ruebush, S. S.; Brantley, S. L.; Tien, M. Reduction of soluble and insoluble iron forms by membrane fractions of Shewanella oneidensis grown under aerobic and anaerobic conditions. *Appl Environ Microbiol* **2006**, *72*, 2925-2935.

(45) Meyer, T. E.; Tsapin, A. I.; Vandenberghe, I.; de Smet, L.; Frishman, D. et al. Identification of 42 possible cytochrome C genes in the Shewanella oneidensis genome and characterization of six soluble cytochromes. *Omics* **2004**, *8*, 57-77.

(46) Myers, C. R.; Myers, J. M. MtrB is required for proper incorporation of the cytochromes OmcA and OmcB into the outer membrane of Shewanella putrefaciens MR-1. *Appl Environ Microbiol* **2002**, *68*, 5585-5594.

(47) Shi, L.; Chen, B.; Wang, Z.; Elias, D. A.; Mayer, M. U. et al. Isolation of high-affinity functional protein complex between OmcA and MtrC: two outer membrane decaheme c-type cytochromes of Shewanella oneidensis MR-1. *Journal of Bacteriology* **2006**, *188*, 4705-4714.

(48) Luirink, J.; Sinning, I. SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta (BBA) Molecular Cell Research Protein Export/Secretion in Bacteria* **2004**, *1694*, 17-35.

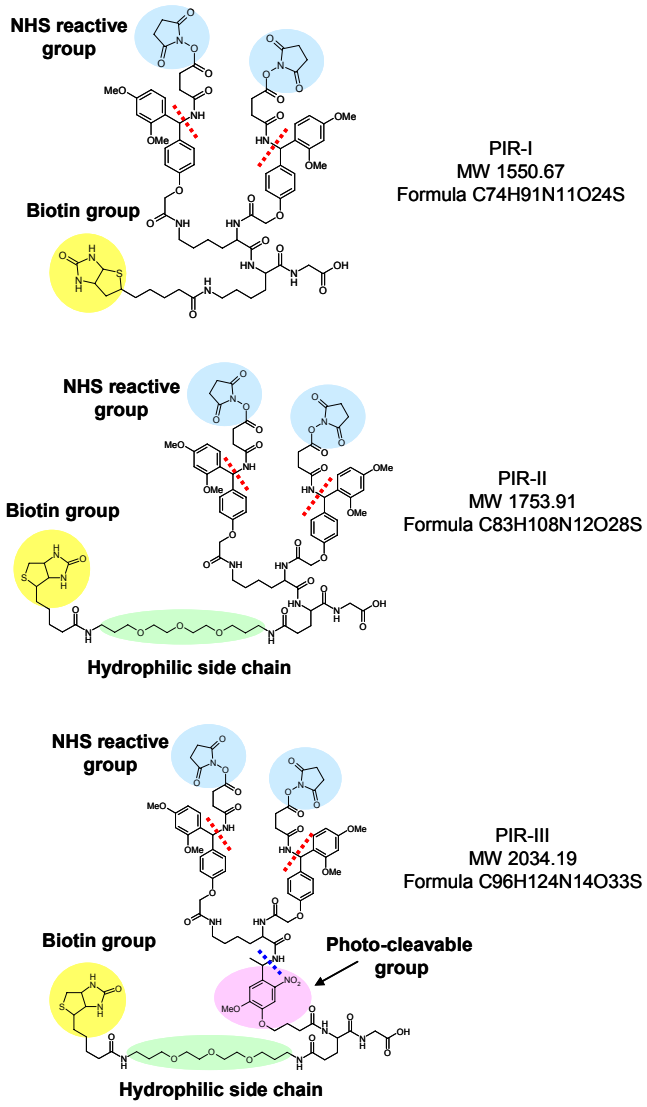



**Figure 1.** Structure of three new affinity chemical probes, named PIR-I, PIR-II, and PIR-III. Each compound includes a biotin group as affinity tag (highlighted in yellow shade) and two reactive NHS groups (in blue shade). Additional features of these compounds are also highlighted in the figure, such as low-energy mass spectrometry cleavable bonds (red dashed lines), a PEG hydrophilic side chain in PIR-II and PIR-III (in green shade), and a photo-cleavable group (in pink shade, cleavable bond as specified by blue dashed lines) for PIR-III.

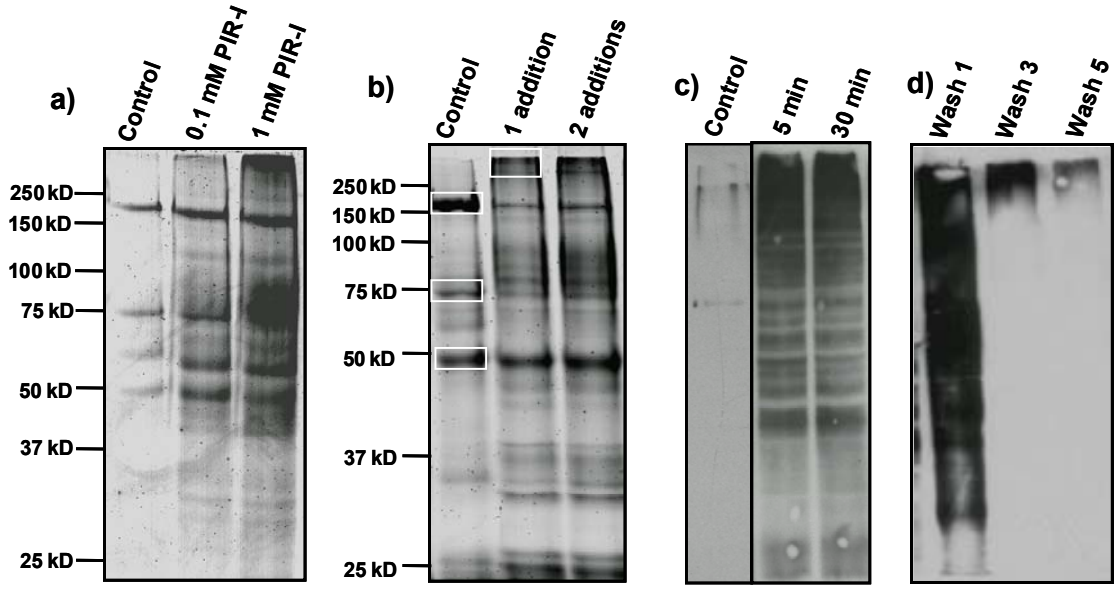



**Figure 2.** Optimization of cell labeling experiments with PIR-I. a) SDS-PAGE of avidin-enriched proteins from control cell lysates (without PIR-I labeling) and labeled cell lysate using 0.1 mM and 1 mM PIR-I. b) SDS-PAGE of avidin-enriched proteins from control cell lysates and labeled cell lysates prepared by adding 10 μL, 100 mM PIR-I to 1 mL cell suspension once and twice. c) Anti-biotin Western blot image of control cell lysate and PIR-I labeled cell lysate at two reaction times: 5 min and 30 min. d) Anti-biotin Western blot image of washing solutions after PIR-I labeling of cells.

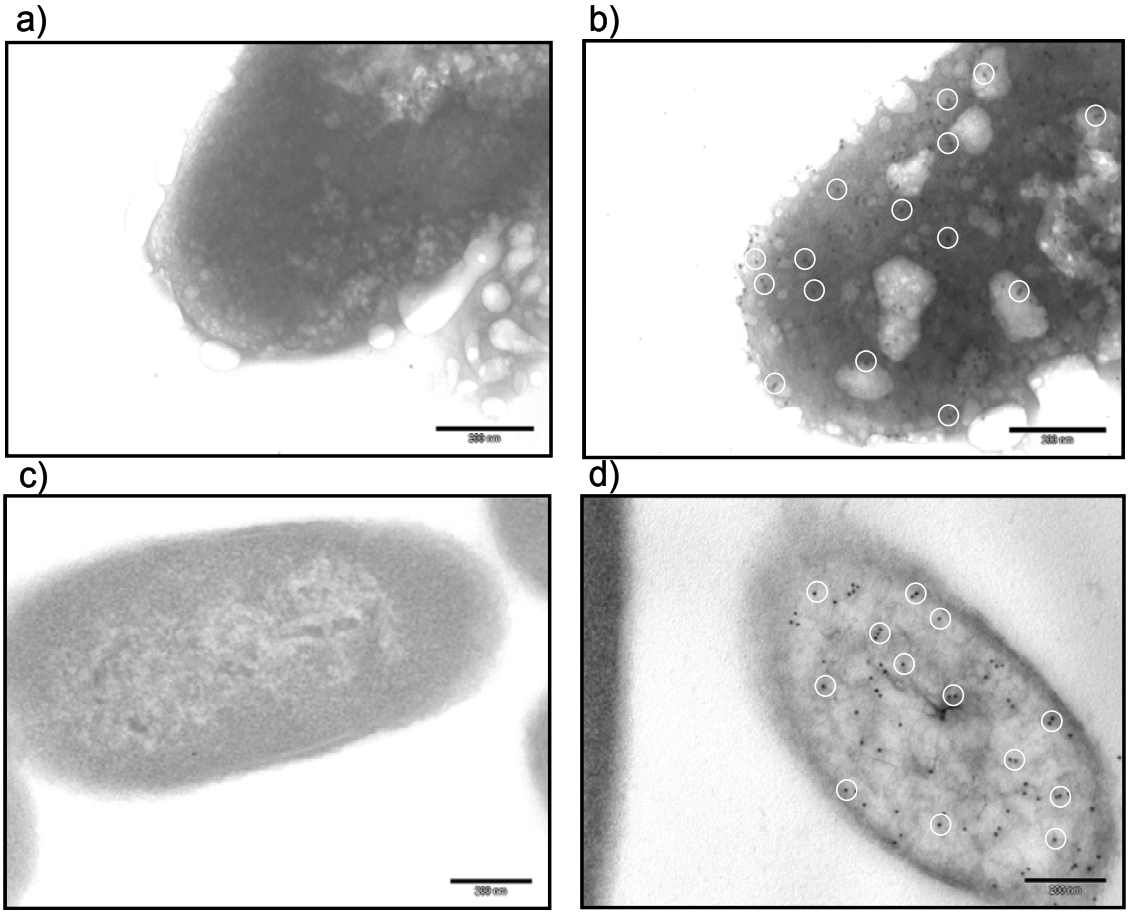



**Figure 3.** EM immunogold detection of PIR-labeled proteins on S. oneidensis cells. a) control cell surface; b) PIR-labeled cell surface; c) control cell thin section; and d) PIR-labeled cell thin section. The bar scale in each image is 200 nm. Dark dots indicate the detection of nanoparticles. To help better visualization, white circles were used to highlight several detected nanoparticles.

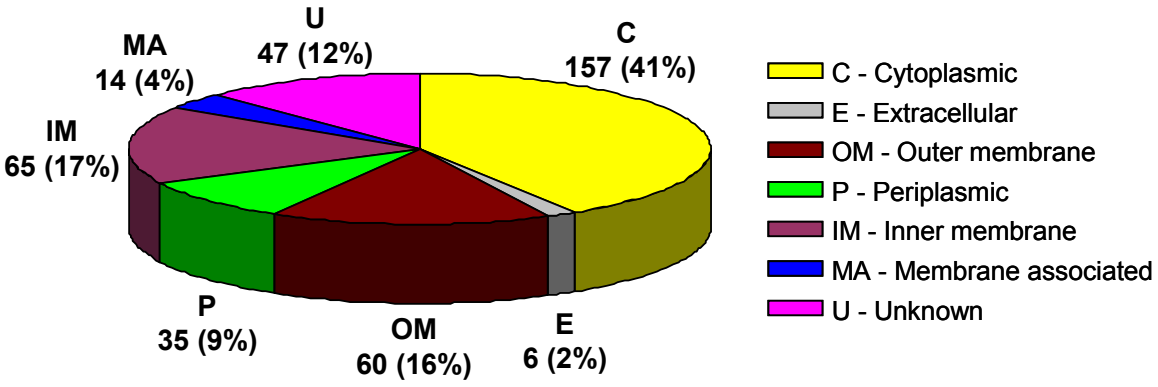



**Figure 4.** Pie diagram representation of the subcellular locations of 384 identified proteins.

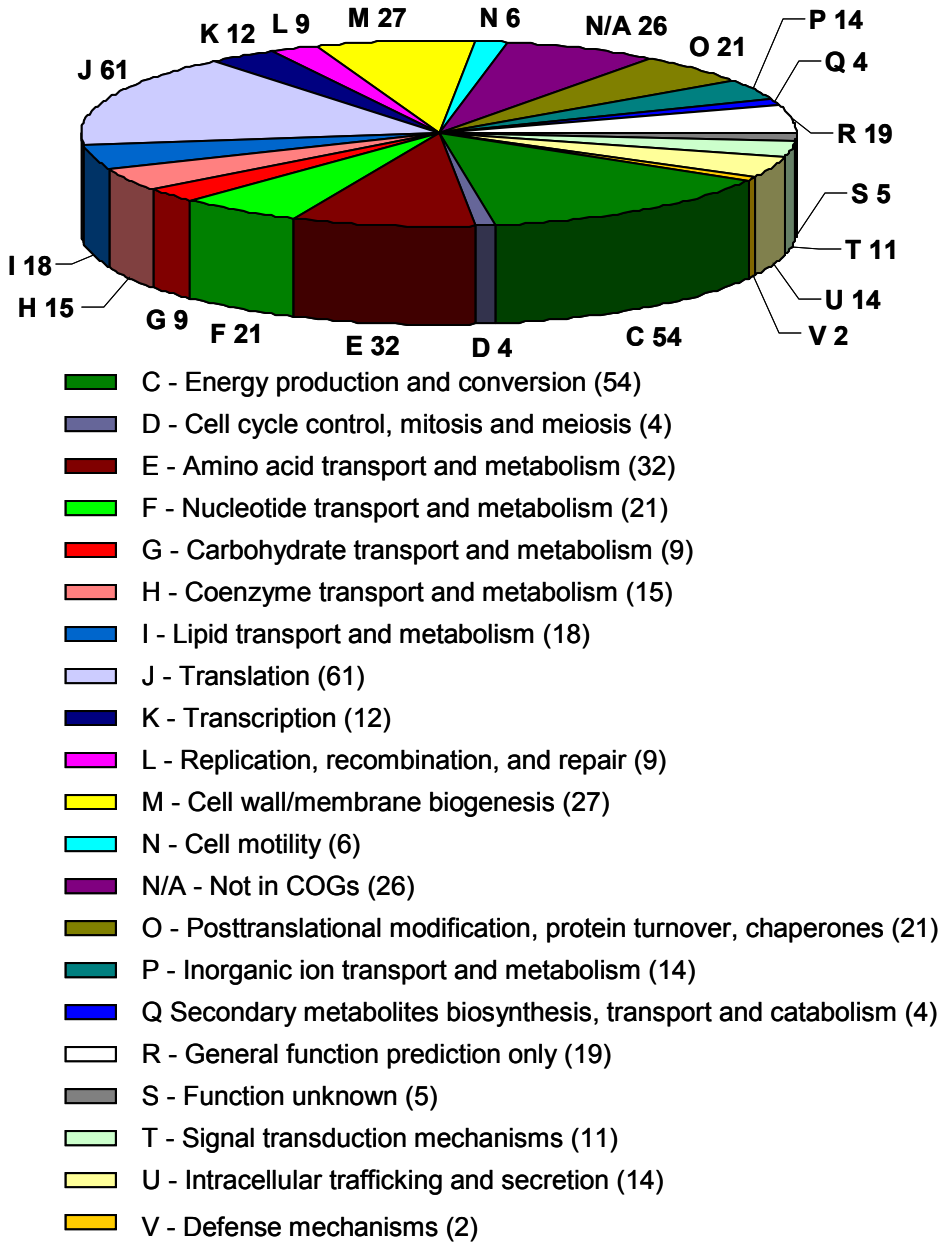



**Figure 5.** Pie diagram representation of 384 identified proteins, classified by COG functional category.

**Table 1.** Select subset of the identified cell envelope proteins in *S. oneidensis*.

| Accession[a] | Protein Description | MW (Da) | Score[b] | Loc[c] | TMDs[d] |
|---|---|---|---|---|---|
| SO0004 | inner membrane protein, 60 kDa | 60469 | 59 | IM | 5 |
| SO0108 | conserved hypothetical protein | 44635 | 60 | IM | 10 |
| SO0162 | phosphoenolpyruvate carboxykinase (ATP) (pckA) | 56190 | 103 | P | 0 |
| SO0251 | preprotein translocase, SecY subunit (secY) | 48709 | 178 | IM | 10 |
| SO0265 | conserved hypothetical protein | 45291 | 60 | IM | 2 |
| SO0266 | cytochrome c-type biogenesis protein CcmF (ccmF-1) | 72323 | 53 | IM | 15 |
| SO0347 | acyltransferase family protein | 34624 | 37 | IM | 1 |
| SO0404 | hypothetical protein | 129383 | 713 | OM | 0 |
| SO0405 | transcription termination factor Rho (rho) | 47189 | 191 | IM | 0 |
| SO0424 | pyruvate dehydrogenase complex, E1 component (aceE) | 99781 | 320 | IM | 0 |
| SO0426 | pyruvate dehydrogenase complex, E3 component (lpdA) | 50733 | 712 | IM | 0 |
| SO0578 | hypothetical protein | 88218 | 36 | OM | 1 |
| SO0605 | hflK protein (hflK) | 42627 | 272 | OM | 1 |
| SO0606 | hflC protein (hflC) | 33184 | 166 | OM | 1 |
| SO0609 | ubiquinol-cytochrome c reductase, cytochrome b (petB) | 45979 | 38 | IM | 10 |
| SO0610 | ubiquinol-cytochrome c reductase, cytochrome c1 (petC) | 26071 | 35 | OM | 1 |
| SO0719 | TonB-dependent receptor, putative | 84248 | 298 | OM | 0 |
| SO0815 | TonB-dependent receptor C-terminal domain protein | 71571 | 299 | OM | 0 |
| SO0904 | NADH:ubiquinone oxidoreductase, Na translocating, γ subunit (nqrC-1) | 29153 | 35 | IM | 1 |
| SO0970 | fumarate reductase flavoprotein subunit precursor | 62865 | 206 | P | 0 |
| SO1016 | NADH dehydrogenase I, G subunit (nuoG) | 101099 | 159 | P | 0 |
| SO1065 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA (fkpA) | 27741 | 203 | OM | 0 |
| SO1066 | extracellular nuclease | 93863 | 61 | OM | 1 |
| SO1103 | NADH:ubiquinone oxidoreductase, Na translocating, α subunit (nqrA-2) | 47999 | 506 | IM | 0 |
| SO1108 | NADH:ubiquinone oxidoreductase, Na translocating, β subunit (nqrF-2) | 46573 | 77 | IM | 1 |
| SO1197 | cell division protein FtsH (ftsH) | 71353 | 293 | IM | 2 |
| SO1201 | preprotein translocase, SecG subunit (secG) | 11554 | 100 | IM | 2 |
| SO1295 | major outer membrane lipoprotein, putative | 9235 | 122 | OM | 0 |
| SO1424 | hypothetical protein | 83796 | 115 | OM | 0 |
| SO1429 | anaerobic dimethyl sulfoxide reductase, A subunit (dmaA-1) | 91916 | 1045 | P | 0 |
| SO1482 | TonB-dependent receptor, putative | 90371 | 137 | OM | 0 |
| SO1580 | TonB-dependent heme receptor | 82653 | 52 | OM | 0 |
| SO1642 | lipid A disaccharide synthase (lpxB) | 42832 | 47 | IM | 1 |
| SO1682 | 3-hydroxyisobutyrate dehydrogenase (mmsB) | 31085 | 288 | P | 0 |
| SO1683 | 3-oxoacyl-(acyl-carrier-protein) reductase, putative | 26564 | 41 | IM | 1 |
| SO1776 | outer membrane protein precursor MtrB (mtrB) | 77743 | 694 | OM | 0 |
| SO1778 | decaheme cytochrome c (omcB) | 72503 | 433 | OM | 0 |
| SO1779 | decaheme cytochrome c (omcA) | 79992 | 4484 | OM | 0 |
| SO1798 | peptidyl-prolyl cis-trans isomerase D (ppiD) | 67456 | 63 | P | 1 |
| SO1825 | MotA/TolQ/ExbB proton channel family protein | 48740 | 591 | IM | 3 |
| SO1826 | TonB system transport protein ExbB2 (exbB2) | 18104 | 177 | IM | 3 |
| SO1827 | TonB system transport protein ExbD2 (exbD2) | 14602 | 38 | IM | 1 |
| SO1829 | TPR domain protein | 46578 | 68 | OM | 1 |
| SO1896 | 3-methylcrotonyl CoA carboxylase, beta subunit (pccB-1) | 58462 | 326 | IM | 0 |
| SO1928 | succinate dehydrogenase, flavoprotein subunit (sdhA) | 64872 | 131 | IM | 0 |
| SO1930 | 2-oxoglutarate dehydrogenase, E1 component (sucA) | 105483 | 324 | IM | 0 |
| SO1931 | 2-oxoglutarate dehydrogenase, E2 component (sucB) | 42804 | 297 | IM | 0 |
| SO2001 | 5-nucleotidase (ushA) | 61353 | 422 | P | 0 |
| SO2361 | cytochrome c oxidase, cbb3-type, subunit III (ccoP) | 35338 | 72 | OM | 2 |
| SO2363 | cytochrome c oxidase, cbb3-type, subunit II (ccoO) | 23530 | 150 | OM | 1 |
| SO2397 | oxidoreductase, short-chain dehydrogenase/reductase family | 28477 | 165 | IM | 1 |
| SO2427 | TonB-dependent receptor, putative | 92130 | 63 | OM | 1 |
| SO2469 | conserved hypothetical protein | 97928 | 55 | OM | 1 |

| | | | | | |
|---|---|---|---|---|---|
| SO2492 | oxidoreductase, acyl-CoA dehydrogenase family | 83772 | 154 | IM | 1 |
| SO2593 | conserved hypothetical protein | 183017 | 2401 | OM | 0 |
| SO2629 | isocitrate dehydrogenase, NADP-dependent (icd) | 80679 | 180 | P | 0 |
| SO2796 | conserved hypothetical protein | 117008 | 56 | OM | 1 |
| SO2881 | superoxide dismutase, Fe (sodB) | 21539 | 152 | E | 0 |
| SO2886 | Na+/H+ antiporter (nhaB) | 58506 | 33 | IM | 9 |
| SO2907 | TonB-dependent receptor domain protein | 95930 | 236 | OM | 1 |
| SO2912 | formate acetyltransferase (pflB) | 85205 | 289 | IM | 0 |
| SO2916 | phosphate acetyltransferase (pta) | 78203 | 102 | IM | 1 |
| SO3111 | protein-export membrane protein SecD (secD-2) | 66508 | 251 | IM | 6 |
| SO3157 | lipoprotein, putative | 26308 | 161 | OM | 0 |
| SO3190 | polysaccharide biosynthesis protein | 47603 | 103 | IM | 1 |
| SO3193 | polysaccharide biosynthesis protein | 100313 | 68 | OM | 1 |
| SO3286 | cytochrome d ubiquinol oxidase, subunit I (cydA) | 58135 | 402 | IM | 9 |
| SO3340 | conserved hypothetical protein | 29984 | 149 | IM | 3 |
| SO3441 | CTP synthase (pyrG) | 60453 | 92 | IM | 1 |
| SO3545 | OmpA family protein | 40271 | 159 | OM | 0 |
| SO3669 | heme transport protein (hugA) | 76438 | 177 | OM | 0 |
| SO3690 | ABC transporter, permease protein | 44303 | 34 | IM | 4 |
| SO3706 | NupC family protein | 44460 | 72 | IM | 10 |
| SO3772 | conserved hypothetical protein | 20142 | 76 | OM | 1 |
| SO3779 | ABC transporter, ATP-binding protein CydC (cydC) | 64719 | 49 | IM | 6 |
| SO3780 | ABC transporter, ATP-binding protein CydD (cydD) | 70547 | 58 | IM | 6 |
| SO3844 | peptidase, M13 family | 77272 | 83 | P | 0 |
| SO3870 | disulfide bond formation protein | 24278 | 34 | IM | 4 |
| SO3896 | outer membrane porin, putative | 39874 | 509 | OM | 0 |
| SO3906 | conserved hypothetical protein | 35988 | 39 | OM | 1 |
| SO3942 | serine protease, HtrA/DegQ/DegS family | 46522 | 329 | P | 1 |
| SO4012 | hypothetical protein | 10465 | 34 | IM | 1 |
| SO4077 | TonB-dependent receptor, putative | 78344 | 98 | OM | 0 |
| SO4105 | MSHA pilin protein MshA (mshA) | 17759 | 67 | OM | 1 |
| SO4202 | Sec-independent protein translocase protein TatA (tatA) | 9333 | 61 | E | 1 |
| SO4211 | preprotein translocase, SecA subunit (secA) | 102787 | 858 | IM | 0 |
| SO4215 | cell division protein FtsZ (ftsZ) | 40910 | 133 | P | 0 |
| SO4473 | outer membrane protein, putative | 20074 | 104 | OM | 0 |
| SO4513 | formate dehydrogenase, alpha subunit | 106403 | 122 | P | 0 |
| SO4557 | methyl-accepting chemotaxis protein | 67264 | 157 | IM | 2 |
| SO4602 | glycerol-3-phosphate acyltransferase (plsB) | 91846 | 133 | IM | 0 |
| SO4719 | conserved hypothetical protein | 29348 | 131 | P | 0 |
| SO4747 | ATP synthase F1, beta subunit (atpD) | 49856 | 2966 | OM | 0 |
| SO4748 | ATP synthase F1, gamma subunit (atpG) | 31609 | 382 | OM | 0 |
| SO4749 | ATP synthase F1, alpha subunit (atpA) | 55228 | 1555 | OM | 0 |
| SOA0048 | prolyl oligopeptidase family protein | 72242 | 72 | P | 1 |
| SOA0099 | conserved hypothetical protein | 63665 | 164 | P | 0 |
| SOA0106 | methyl-accepting chemotaxis protein | 59803 | 49 | IM | 2 |
| SOA0161 | zinc-binding dehydrogenase | 40772 | 48 | IM | 1 |

[a] Accession no of proteins according to TIGR annotation (http://www.tigr.org/). [b] Mascot protein score. [c] Subcellular location. E denotes extracellular; IM denotes inner membrane; OM denotes outer membrane; and P denotes periplasmic. [d] Predicted number of transmembrane helices according to TMHMM (http://www.cbs.dtu.dk/services/TMHMM/).

**Supplementary Table 1.** Characterizing all identified proteins in *S. oneidensis*.

| Accession[a] | Protein Description | MW (Da) | Score[b] | Loc[c] | COG[d] | TMDs[e] |
|---|---|---|---|---|---|---|
| SO0004 | inner membrane protein, 60 kDa | 60469 | 59 | IM | U | 5 |
| SO0011 | DNA gyrase, B subunit (gyrB) | 90216 | 162 | C | L | 0 |
| SO0014 | glycyl-tRNA synthetase, beta subunit (glyS) | 74946 | 138 | U | J | 0 |
| SO0020 | fatty oxidation complex, beta subunit (fadA) | 41128 | 37 | C | I | 0 |
| SO0021 | fatty oxidation complex, alpha subunit (fadB) | 76929 | 276 | C | I | 0 |
| SO0052 | protein-export protein SecB (secB) | 17200 | 64 | IM | U | 0 |
| SO0053 | glycerol-3-phosphate dehydrogenase (NAD(P)+) (gpsA) | 36041 | 89 | IM | C | 0 |
| SO0098 | histidine ammonia-lyase (hutH) | 54682 | 38 | IM | E | 0 |
| SO0102 | formate dehydrogenase, nitrate-inducible, iron-sulfur subunit (fdnH) | 34134 | 44 | P | C | 0 |
| SO0105 | L-seryl-tRNA selenium transferase (selA) | 51765 | 70 | C | E | 0 |
| SO0108 | conserved hypothetical protein | 44635 | 60 | IM | R | 10 |
| SO0162 | phosphoenolpyruvate carboxykinase (ATP) (pckA) | 56190 | 103 | P | C | 0 |
| SO0217 | translation elongation factor Tu (tufB) | 43605 | 3651 | C | J | 0 |
| SO0220 | ribosomal protein L11 (rplK) | 15155 | 59 | C | J | 0 |
| SO0221 | ribosomal protein L1 (rplA) | 24585 | 142 | C | J | 0 |
| SO0222 | ribosomal protein L10 (rplJ) | 17789 | 312 | C | J | 0 |
| SO0223 | ribosomal protein L7/L12 (rplL) | 12509 | 37 | C | J | 0 |
| SO0224 | DNA-directed RNA polymerase, beta subunit (rpoB) | 150601 | 330 | C | K | 0 |
| SO0225 | DNA-directed RNA polymerase, beta subunit (rpoC) | 156220 | 737 | C | K | 0 |
| SO0226 | ribosomal protein S12 (rpsL) | 13812 | 65 | C | J | 0 |
| SO0227 | ribosomal protein S7 (rpsG) | 17777 | 147 | C | J | 0 |
| SO0228 | translation elongation factor G (fusA-1) | 77300 | 41 | C | J | 0 |
| SO0229 | translation elongation factor Tu (tufA) | 43542 | 3903 | C | J | 0 |
| SO0230 | ribosomal protein S10 (rpsJ) | 11761 | 95 | C | J | 0 |
| SO0231 | ribosomal protein L3 (rplC) | 22518 | 198 | C | J | 0 |
| SO0232 | ribosomal protein L4 (rplD) | 21968 | 290 | C | J | 0 |
| SO0233 | ribosomal protein L23 (rplW) | 11061 | 41 | C | J | 0 |
| SO0234 | ribosomal protein L2 (rplB) | 30059 | 432 | C | J | 0 |
| SO0235 | ribosomal protein S19 (rpsS) | 10466 | 33 | C | J | 0 |
| SO0236 | ribosomal protein L22 (rplV) | 12064 | 44 | C | J | 0 |
| SO0237 | ribosomal protein S3 (rpsC) | 25739 | 953 | C | J | 0 |
| SO0238 | ribosomal protein L16 (rplP) | 15443 | 109 | C | J | 0 |
| SO0240 | ribosomal protein S17 (rpsQ) | 9460 | 59 | C | J | 0 |
| SO0241 | ribosomal protein L14 (rplN) | 13561 | 85 | C | J | 0 |
| SO0242 | ribosomal protein L24 (rplX) | 11311 | 38 | C | J | 0 |
| SO0243 | ribosomal protein L5 (rplE) | 20276 | 56 | C | J | 0 |
| SO0246 | ribosomal protein L6 (rplF) | 18968 | 492 | C | J | 0 |
| SO0247 | ribosomal protein L18 (rplR) | 12715 | 46 | C | J | 0 |
| SO0248 | ribosomal protein S5 (rpsE) | 17729 | 402 | C | J | 0 |
| SO0250 | ribosomal protein L15 (rplO) | 15071 | 178 | C | J | 0 |
| SO0251 | preprotein translocase, SecY subunit (secY) | 48709 | 178 | IM | U | 10 |
| SO0253 | ribosomal protein S13 (rpsM) | 13426 | 160 | C | J | 0 |
| SO0254 | ribosomal protein S11 (rpsK) | 13940 | 114 | C | J | 0 |
| SO0255 | ribosomal protein S4 (rpsD) | 23479 | 1109 | C | J | 0 |
| SO0256 | DNA-directed RNA polymerase, alpha subunit (rpoA) | 36309 | 162 | C | K | 0 |
| SO0257 | ribosomal protein L17 (rplQ) | 14762 | 33 | C | J | 0 |
| SO0265 | conserved hypothetical protein | 45291 | 60 | IM | O | 2 |
| SO0266 | cytochrome c-type biogenesis protein CcmF (ccmF-1) | 72323 | 53 | IM | O | 15 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO0274 | phosphoenolpyruvate carboxylase (ppc) | 100336 | 61 | U | C | 0 |
| SO0314 | ornithine decarboxylase, inducible (speF) | 82018 | 516 | C | E | 0 |
| SO0342 | conserved hypothetical protein | 41804 | 39 | U | S | 0 |
| SO0347 | acyltransferase family protein | 34624 | 37 | IM | I | 1 |
| SO0359 | guanosine-3,5-bis(diphosphate) 3-pyrophosphohydrolase (spoT) | 79021 | 45 | C | K | 0 |
| SO0404 | hypothetical protein | 129383 | 713 | OM | NA | 0 |
| SO0405 | transcription termination factor Rho (rho) | 47189 | 191 | IM | K | 0 |
| SO0424 | pyruvate dehydrogenase complex, E1 component, pyruvate dehydrogenase (aceE) | 99781 | 320 | IM | C | 0 |
| SO0425 | pyruvate dehydrogenase complex, E2 component, dihydrolipoamide acetyltransferase (aceF) | 69470 | 54 | IM | C | 0 |
| SO0426 | pyruvate dehydrogenase complex, E3 component, lipoamide dehydrogenase (lpdA) | 50733 | 712 | IM | C | 0 |
| SO0429 | peptidase, M13 family | 76207 | 80 | P | O | 0 |
| SO0432 | aconitate hydratase 2 (acnB) | 93904 | 491 | C | C | 0 |
| SO0435 | uroporphyrinogen decarboxylase (hemE) | 39382 | 243 | U | H | 0 |
| SO0441 | phosphoribosylamine--glycine ligase (purD) | 45738 | 51 | U | F | 0 |
| SO0442 | phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase (purH) | 58641 | 111 | MA | F | 0 |
| SO0506 | conserved hypothetical protein TIGR00148 | 55272 | 33 | P | H | 0 |
| SO0578 | hypothetical protein | 88218 | 36 | OM | NA | 1 |
| SO0605 | hflK protein (hflK) | 42627 | 272 | OM | O | 1 |
| SO0606 | hflC protein (hflC) | 33184 | 166 | OM | O | 1 |
| SO0609 | ubiquinol-cytochrome c reductase, cytochrome b (petB) | 45979 | 38 | IM | C | 10 |
| SO0610 | ubiquinol-cytochrome c reductase, cytochrome c1 (petC) | 26071 | 35 | OM | C | 1 |
| SO0611 | stringent starvation protein a (sspA) | 24122 | 43 | C | O | 0 |
| SO0617 | acetylornithine aminotransferase (argD) | 43415 | 40 | C | E | 0 |
| SO0620 | conserved hypothetical protein | 29635 | 168 | U | NA | 0 |
| SO0624 | catabolite gene activator (crp) | 23862 | 136 | U | T | 0 |
| SO0680 | hypothetical protein | 142069 | 39 | OM | N | 0 |
| SO0704 | chaperonin GroEL (groEL) | 57101 | 179 | MA | O | 0 |
| SO0719 | TonB-dependent receptor, putative | 84248 | 298 | OM | P | 0 |
| SO0770 | malate dehydrogenase (mdh) | 32288 | 186 | MA | C | 0 |
| SO0781 | glycine cleavage system P protein (gcvP) | 105348 | 241 | C | E | 0 |
| SO0815 | TonB-dependent receptor C-terminal domain protein | 71571 | 299 | OM | P | 0 |
| SO0832 | conserved hypothetical protein TIGR00046 | 26813 | 37 | C | S | 0 |
| SO0840 | acetyl-CoA carboxylase multifunctional enzyme accADC, carboxyl transferase subunit alpha/carboxyl transferase subunit beta/biotin carboxylase | 168259 | 7024 | U | I | 0 |
| SO0842 | translation elongation factor G (fusA-2) | 77127 | 1957 | C | J | 0 |
| SO0848 | periplasmic nitrate reductase (napA) | 93099 | 59 | P | C | 0 |
| SO0859 | sensory box histidine kinase/response regulator | 197712 | 31 | IM | T | 1 |
| SO0860 | response regulator | 38176 | 32 | C | T | 0 |
| SO0861 | conserved hypothetical protein | 35130 | 56 | U | R | 0 |
| SO0862 | D-3-phosphoglycerate dehydrogenase (serA) | 44400 | 92 | U | H | 0 |
| SO0876 | peptidase B (pepB) | 45477 | 305 | C | E | 0 |
| SO0904 | NADH:ubiquinone oxidoreductase, Na translocating, gamma subunit (nqrC-1) | 29153 | 35 | IM | C | 1 |
| SO0918 | aculeacin A acylase (aac) | 92593 | 136 | P | R | 0 |
| SO0929 | S-adenosylmethionine synthetase (metK) | 41582 | 168 | C | H | 0 |
| SO0930 | transketolase (tkt) | 72083 | 320 | U | G | 0 |
| SO0932 | phosphoglycerate kinase (pgk) | 40744 | 66 | C | G | 0 |
| SO0933 | fructose-bisphosphate aldolase, class II, Calvin cycle subtype (fba) | 38828 | 295 | MA | G | 0 |
| SO0959 | cytosol aminopeptidase (pepA-1) | 54354 | 90 | OM | E | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO0970 | fumarate reductase flavoprotein subunit precursor | 62865 | 206 | P | C | 0 |
| SO0992 | lysyl-tRNA synthetase (lysS) | 57138 | 48 | C | J | 0 |
| SO1016 | NADH dehydrogenase I, G subunit (nuoG) | 101099 | 159 | P | C | 0 |
| SO1051 | periplasmic glucan biosynthesis protein, putative | 67936 | 82 | P | P | 0 |
| SO1060 | conserved hypothetical protein | 22142 | 37 | U | R | 0 |
| SO1065 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA (fkpA) | 27741 | 203 | OM | O | 0 |
| SO1066 | extracellular nuclease | 93863 | 61 | OM | R | 1 |
| SO1103 | NADH:ubiquinone oxidoreductase, Na translocating, alpha subunit (nqrA-2) | 47999 | 506 | IM | C | 0 |
| SO1108 | NADH:ubiquinone oxidoreductase, Na translocating, beta subunit (nqrF-2) | 46573 | 77 | IM | C | 1 |
| SO1115 | aminoacyl-histidine dipeptidase (pepD) | 53191 | 34 | MA | E | 0 |
| SO1117 | cytosol aminopeptidase, putative | 54905 | 77 | C | E | 0 |
| SO1126 | chaperone protein DnaK (dnaK) | 68876 | 59 | MA | O | 0 |
| SO1139 | peptidyl-prolyl cis-trans isomerase FklB (fklB) | 21855 | 79 | OM | O | 0 |
| SO1142 | carbamoyl-phosphate synthase, large subunit (carB) | 118876 | 307 | C | E | 0 |
| SO1149 | conserved hypothetical protein | 44416 | 101 | C | R | 0 |
| SO1164 | D-alanyl-D-alanine carboxypeptidase (dacA-1) | 43139 | 65 | P | M | 0 |
| SO1174 | leucyl-tRNA synthetase (leuS) | 97336 | 75 | P | J | 0 |
| SO1183 | oxidoreductase, FAD-binding | 44590 | 40 | U | C | 0 |
| SO1185 | conserved hypothetical protein TIGR00092 | 39660 | 40 | C | J | 0 |
| SO1197 | cell division protein FtsH (ftsH) | 71353 | 293 | IM | O | 2 |
| SO1201 | preprotein translocase, SecG subunit (secG) | 11554 | 100 | IM | U | 2 |
| SO1203 | N utilization substance protein A (nusA) | 55353 | 122 | C | K | 0 |
| SO1204 | translation initiation factor IF-2 (infB) | 96225 | 548 | C | J | 0 |
| SO1209 | polyribonucleotide nucleotidyltransferase (pnp) | 75898 | 422 | C | J | 0 |
| SO1210 | TPR domain protein | 32249 | 155 | U | R | 0 |
| SO1215 | outer membrane protein OmpK, putative | 31430 | 31 | OM | M | 0 |
| SO1221 | purine nucleoside phosphorylase (deoD-2) | 25814 | 150 | U | F | 0 |
| SO1222 | hypothetical protein | 45188 | 31 | OM | R | 1 |
| SO1258 | adenylosuccinate synthetase, putative | 46052 | 53 | U | F | 0 |
| SO1276 | 4-aminobutyrate aminotransferase (gabT) | 45159 | 149 | C | E | 0 |
| SO1284 | RNA polymerase sigma-70 factor (rpoD) | 70462 | 97 | C | K | 0 |
| SO1295 | major outer membrane lipoprotein, putative | 9235 | 122 | OM | M | 0 |
| SO1300 | glutamate-1-semialdehyde-2,1-aminomutase (hemL) | 46502 | 143 | IM | H | 0 |
| SO1315 | tyrosyl-tRNA synthetase (tyrS) | 44000 | 161 | C | J | 0 |
| SO1348 | ribonuclease III (rnc) | 25452 | 33 | C | K | 0 |
| SO1356 | signal recognition particle protein Ffh (ffh) | 49880 | 36 | C | U | 0 |
| SO1360 | ribosomal protein L19 (rplS) | 13284 | 164 | C | J | 0 |
| SO1377 | conserved hypothetical protein | 65602 | 32 | IM | N | 1 |
| SO1424 | hypothetical protein | 83796 | 115 | OM | NA | 0 |
| SO1429 | anaerobic dimethyl sulfoxide reductase, A subunit (dmaA-1) | 91916 | 1045 | P | C | 0 |
| SO1430 | anaerobic dimethyl sulfoxide reductase, B subunit (dmsB-1) | 25298 | 67 | IM | C | 0 |
| SO1482 | TonB-dependent receptor, putative | 90371 | 137 | OM | P | 0 |
| SO1490 | alcohol dehydrogenase II (adhB) | 40387 | 343 | C | C | 0 |
| SO1519 | iron-sulfur cluster-binding protein | 52140 | 145 | C | C | 0 |
| SO1521 | iron-sulfur cluster-binding protein | 103410 | 283 | U | C | 0 |
| SO1524 | heat shock protein GrpE (grpE) | 22688 | 151 | MA | O | 0 |
| SO1531 | thiamine biosynthesis protein ThiI (thiI) | 55350 | 36 | C | P | 0 |
| SO1538 | isocitrate dehydrogenase, NAD-dependent | 36173 | 63 | MA | C | 0 |
| SO1550 | conserved hypothetical protein | 50492 | 71 | U | R | 0 |
| SO1580 | TonB-dependent heme receptor | 82653 | 52 | OM | P | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO1602 | multi-domain beta-ketoacyl synthase | 272650 | 79 | P | Q | 0 |
| SO1625 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (dapD) | 30026 | 41 | C | E | 0 |
| SO1629 | ribosomal protein S2 (rpsB) | 26610 | 480 | C | J | 0 |
| SO1630 | translation elongation factor Ts (tsf) | 30509 | 262 | C | J | 0 |
| SO1631 | uridylate kinase (pyrH) | 26406 | 108 | IM | F | 0 |
| SO1637 | bacterial surface antigen | 92574 | 72 | OM | M | 0 |
| SO1638 | outer membrane protein OmpH (ompH) | 18980 | 44 | OM | M | 0 |
| SO1642 | lipid A disaccharide synthase (lpxB) | 42832 | 47 | IM | M | 1 |
| SO1651 | Snf2 family protein | 120656 | 33 | U | K | 0 |
| SO1677 | acetyl-CoA acetyltransferase (atoB) | 40932 | 310 | U | I | 0 |
| SO1678 | methylmalonate-semialdehyde dehydrogenase (mmsA) | 54157 | 63 | C | C | 0 |
| SO1679 | acyl-CoA dehydrogenase family protein | 42345 | 99 | U | I | 0 |
| SO1680 | enoyl-CoA hydratase/isomerase family protein | 27559 | 57 | U | I | 0 |
| SO1682 | 3-hydroxyisobutyrate dehydrogenase (mmsB) | 31085 | 288 | P | I | 0 |
| SO1683 | 3-oxoacyl-(acyl-carrier-protein) reductase, putative | 26564 | 41 | IM | Q | 1 |
| SO1690 | ABC transporter, ATP-binding protein | 59740 | 75 | IM | R | 0 |
| SO1776 | outer membrane protein precursor MtrB (mtrB) | 77743 | 694 | OM | M | 0 |
| SO1778 | decaheme cytochrome c (omcB) | 72503 | 433 | OM | C | 0 |
| SO1779 | decaheme cytochrome c (omcA) | 79992 | 4484 | OM | NA | 0 |
| SO1786 | glutaminyl-tRNA synthetase (glnS) | 64804 | 31 | C | J | 0 |
| SO1793 | trigger factor (tig) | 47567 | 449 | C | O | 0 |
| SO1794 | ATP-dependent Clp protease, proteolytic subunit (clpP) | 22215 | 64 | IM | U | 0 |
| SO1796 | ATP-dependent protease La (lon) | 87696 | 219 | MA | O | 0 |
| SO1798 | peptidyl-prolyl cis-trans isomerase D (ppiD) | 67456 | 63 | P | O | 1 |
| SO1812 | methionine gamma-lyase (mdeA) | 43078 | 38 | C | E | 0 |
| SO1821 | outer membrane porin, putative | 43970 | 81 | OM | M | 0 |
| SO1824 | conserved hypothetical protein | 24943 | 126 | OM | G | 0 |
| SO1825 | MotA/TolQ/ExbB proton channel family protein | 48740 | 591 | IM | U | 3 |
| SO1826 | TonB system transport protein ExbB2 (exbB2) | 18104 | 177 | IM | U | 3 |
| SO1827 | TonB system transport protein ExbD2 (exbD2) | 14602 | 38 | IM | U | 1 |
| SO1829 | TPR domain protein | 46578 | 68 | OM | R | 1 |
| SO1851 | conserved hypothetical protein | 80454 | 38 | U | L | 0 |
| SO1853 | ABC transporter, ATP-binding protein | 72366 | 46 | IM | R | 0 |
| SO1854 | hypothetical protein | 66589 | 58 | OM | NA | 0 |
| SO1870 | biosynthetic arginine decarboxylase (speA) | 71359 | 91 | P | E | 0 |
| SO1894 | acetyl-CoA carboxylase, biotin carboxylase, putative | 75780 | 2027 | U | I | 0 |
| SO1896 | 3-methylcrotonyl CoA carboxylase, beta subunit (pccB-1) | 58462 | 326 | IM | I | 0 |
| SO1897 | isovaleryl-CoA dehydrogenase (ivd) | 42276 | 95 | U | I | 0 |
| SO1926 | citrate synthase (gltA) | 48505 | 1450 | C | C | 0 |
| SO1928 | succinate dehydrogenase, flavoprotein subunit (sdhA) | 64872 | 131 | IM | C | 0 |
| SO1930 | 2-oxoglutarate dehydrogenase, E1 component (sucA) | 105483 | 324 | IM | C | 0 |
| SO1931 | 2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase (sucB) | 42804 | 297 | IM | C | 0 |
| SO1932 | succinyl-CoA synthase, beta subunit (sucC) | 41677 | 133 | C | C | 0 |
| SO1933 | succinyl-CoA synthase, alpha subunit (sucD) | 30030 | 202 | U | C | 0 |
| SO1961 | maltose O-acetyltransferase (maa) | 21320 | 32 | C | R | 0 |
| SO1962 | 4-hydroxyphenylpyruvate dioxygenase | 39407 | 33 | C | E | 0 |
| SO1981 | conserved hypothetical protein | 55596 | 32 | C | H | 0 |
| SO2001 | 5-nucleotidase (ushA) | 61353 | 422 | P | F | 0 |
| SO2016 | heat shock protein HtpG (htpG) | 71779 | 593 | MA | O | 0 |
| SO2085 | phenylalanyl-tRNA synthetase, alpha subunit (pheS) | 37371 | 48 | C | J | 0 |
| SO2086 | phenylalanyl-tRNA synthetase, beta subunit (pheT) | 87147 | 162 | C | J | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO2216 | sensory box protein | 84731 | 35 | U | T | 0 |
| SO2218 | asparaginyl-tRNA synthetase (asnS) | 52442 | 80 | C | J | 0 |
| SO2264 | cysteine desulfurase (iscS) | 45003 | 89 | C | E | 0 |
| SO2274 | nucleoside diphosphate kinase (ndk) | 15530 | 170 | C | F | 0 |
| SO2299 | threonyl-tRNA synthetase (thrS) | 74357 | 88 | C | J | 0 |
| SO2304 | alanine dehydrogenase, authentic point mutation (ald) | 39385 | 47 | C | E | 0 |
| SO2310 | seryl-tRNA synthetase (serS) | 47025 | 100 | C | J | 0 |
| SO2338 | succinylglutamate desuccinylase (astE) | 38987 | 39 | U | E | 0 |
| SO2339 | alpha keto acid dehydrogenase complex, E1 component, alpha subunit | 43870 | 62 | C | C | 0 |
| SO2340 | alpha keto acid dehydrogenase complex, E1 component, beta subunit | 35578 | 138 | C | C | 0 |
| SO2341 | alpha keto acid dehydrogenase complex, E2 component | 57626 | 1557 | C | C | 0 |
| SO2342 | quinolinate synthetase complex, subunit A (nadA) | 39140 | 33 | C | H | 0 |
| SO2345 | glyceraldehyde 3-phosphate dehydrogenase (gapA-2) | 36629 | 41 | C | G | 0 |
| SO2347 | glyceraldehyde 3-phosphate dehydrogenase (gapA-3) | 52427 | 230 | C | G | 0 |
| SO2350 | aspartate aminotransferase (aspC-1) | 43192 | 74 | C | E | 0 |
| SO2361 | cytochrome c oxidase, cbb3-type, subunit III (ccoP) | 35338 | 72 | OM | C | 2 |
| SO2363 | cytochrome c oxidase, cbb3-type, subunit II (ccoO) | 23530 | 150 | OM | C | 1 |
| SO2395 | acyl-CoA dehydrogenase family protein | 63691 | 34 | C | I | 0 |
| SO2397 | oxidoreductase, short-chain dehydrogenase/reductase family | 28477 | 165 | IM | Q | 1 |
| SO2402 | ribosomal protein S1 (rpsA) | 61339 | 1345 | C | J | 0 |
| SO2404 | 3-phosphoshikimate 1-carboxyvinyltransferase (aroA) | 46029 | 60 | U | E | 0 |
| SO2406 | aspartate aminotransferase (aspC-2) | 43778 | 45 | C | E | 0 |
| SO2411 | DNA gyrase, A subunit (gyrA) | 101854 | 77 | C | L | 0 |
| SO2415 | ribonucleoside-diphosphate reductase, alpha subunit (nrdA) | 86399 | 52 | U | F | 0 |
| SO2416 | ribonucleoside-diphosphate reductase, beta subunit (nrdB) | 43809 | 56 | U | F | 0 |
| SO2427 | TonB-dependent receptor, putative | 92130 | 63 | OM | P | 1 |
| SO2433 | aspartyl-tRNA synthetase (aspS) | 66319 | 259 | C | J | 0 |
| SO2440 | thiH protein (thiH) | 42953 | 34 | C | H | 0 |
| SO2469 | conserved hypothetical protein | 97928 | 55 | OM | P | 1 |
| SO2476 | polysaccharide biosynthesis protein | 44072 | 92 | OM | M | 0 |
| SO2477 | alcohol dehydrogenase, iron-containing | 39809 | 59 | C | C | 0 |
| SO2490 | transcriptional regulator, RpiR family | 31597 | 63 | C | K | 0 |
| SO2491 | pyruvate kinase II (pykA) | 51268 | 83 | U | G | 0 |
| SO2492 | oxidoreductase, acyl-CoA dehydrogenase family | 83772 | 154 | IM | I | 1 |
| SO2498 | sensory box protein | 88298 | 34 | U | T | 0 |
| SO2506 | excinuclease ABC, B subunit (uvrB) | 76285 | 34 | U | L | 0 |
| SO2519 | transcriptional regulator, AraC family | 44413 | 32 | U | K | 0 |
| SO2539 | response regulator | 74169 | 35 | U | T | 0 |
| SO2593 | conserved hypothetical protein | 183017 | 2401 | OM | E | 0 |
| SO2602 | conserved hypothetical protein | 23469 | 33 | C | T | 0 |
| SO2619 | methionyl-tRNA synthetase (metG) | 76694 | 119 | C | J | 0 |
| SO2629 | isocitrate dehydrogenase, NADP-dependent (icd) | 80679 | 180 | P | C | 0 |
| SO2635 | adenylosuccinate lyase (purB) | 50947 | 45 | C | F | 0 |
| SO2636 | conserved hypothetical protein | 43070 | 54 | C | S | 0 |
| SO2638 | leucine dehydrogenase (ldh) | 37343 | 837 | C | E | 0 |
| SO2644 | phosphoenolpyruvate synthase (ppsA) | 87121 | 172 | C | G | 0 |
| SO2753 | prolyl endopeptidase | 81113 | 39 | P | E | 0 |
| SO2756 | antioxidant, AhpC/Tsa family, authentic frameshift | 21998 | 807 | C | NA | 0 |
| SO2760 | phosphoribosylformylglycinamidine cyclo-ligase (purM) | 37883 | 152 | U | F | 0 |
| SO2785 | ribonuclease E (rne) | 120535 | 245 | C | J | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO2787 | cold shock domain family protein | 7519 | 172 | C | NA | 0 |
| SO2796 | conserved hypothetical protein | 117008 | 56 | OM | NA | 1 |
| SO2853 | 3-oxoacyl-(acyl-carrier-protein) synthase III, putative | 38288 | 34 | U | I | 0 |
| SO2881 | superoxide dismutase, Fe (sodB) | 21539 | 152 | E | P | 0 |
| SO2886 | Na+/H+ antiporter (nhaB) | 58506 | 33 | IM | P | 9 |
| SO2903 | cysteine synthase A (cysK) | 34456 | 92 | U | E | 0 |
| SO2907 | TonB-dependent receptor domain protein | 95930 | 236 | OM | P | 1 |
| SO2912 | formate acetyltransferase (pflB) | 85205 | 289 | IM | C | 0 |
| SO2915 | acetate kinase (ackA) | 43897 | 209 | C | C | 0 |
| SO2916 | phosphate acetyltransferase (pta) | 78203 | 102 | IM | C | 1 |
| SO2934 | conserved hypothetical protein | 85587 | 249 | P | R | 0 |
| SO2938 | hypothetical protein | 19401 | 77 | E | NA | 0 |
| SO3023 | tryptophan synthase, beta subunit (trpB) | 43131 | 59 | C | E | 0 |
| SO3033 | ferric alcaligin siderophore receptor | 77948 | 90 | OM | P | 0 |
| SO3054 | metallo-beta-lactamase family protein | 23866 | 81 | C | R | 0 |
| SO3064 | amidophosphoribosyltransferase (purF) | 56062 | 189 | C | F | 0 |
| SO3072 | 3-oxoacyl-(acyl-carrier-protein) synthase I (fabB) | 43391 | 384 | C | Q | 0 |
| SO3099 | long-chain fatty acid transport protein, putative | 47034 | 34040 | OM | I | 0 |
| SO3111 | protein-export membrane protein SecD (secD-2) | 66508 | 251 | IM | U | 6 |
| SO3144 | electron transfer flavoprotein, alpha subunit (etfA) | 31477 | 32 | U | C | 0 |
| SO3154 | prolyl-tRNA synthetase (proS) | 63213 | 297 | C | J | 0 |
| SO3157 | lipoprotein, putative | 26308 | 161 | OM | NA | 0 |
| SO3175 | asparagine synthetase, glutamine-hydrolyzing (asnB-2) | 73643 | 419 | U | E | 0 |
| SO3183 | perosamine synthetase-related protein | 36433 | 35 | OM | M | 0 |
| SO3188 | dTDP-glucose 4,6-dehydratase (rfbB) | 42812 | 35 | OM | M | 0 |
| SO3190 | polysaccharide biosynthesis protein | 47603 | 103 | IM | M | 1 |
| SO3193 | polysaccharide biosynthesis protein | 100313 | 68 | OM | M | 1 |
| SO3207 | chemotaxis protein | 80828 | 48 | IM | N | 0 |
| SO3211 | flagellar biosynthetic protein FlhG (flhG) | 30889 | 33 | C | D | 0 |
| SO3265 | conserved hypothetical protein | 44105 | 33 | U | I | 0 |
| SO3286 | cytochrome d ubiquinol oxidase, subunit I (cydA) | 58135 | 402 | IM | C | 9 |
| SO3287 | phosphoribosylformylglycinamidine synthase (purL) | 141479 | 35 | U | F | 0 |
| SO3292 | GMP synthase (guaA) | 58560 | 60 | C | F | 0 |
| SO3293 | inosine-5-monophosphate dehydrogenase (guaB) | 51937 | 621 | U | F | 0 |
| SO3308 | GTP-binding protein EngA (engA) | 54554 | 32 | C | R | 0 |
| SO3311 | histidyl-tRNA synthetase (hisS) | 47608 | 56 | C | J | 0 |
| SO3315 | conserved hypothetical protein TIGR00048 | 42024 | 32 | C | R | 0 |
| SO3340 | conserved hypothetical protein | 29984 | 149 | IM | M | 3 |
| SO3343 | conserved hypothetical protein | 20697 | 83 | OM | NA | 0 |
| SO3347 | conserved hypothetical protein TIGR00250 | 15475 | 34 | U | L | 0 |
| SO3411 | protease, putative | 121238 | 55 | P | M | 0 |
| SO3417 | peptidyl-prolyl cis-trans isomerase SlyD (slyD) | 23989 | 78 | P | O | 0 |
| SO3424 | valyl-tRNA synthetase (valS) | 109029 | 56 | C | J | 0 |
| SO3428 | alanyl-tRNA synthetase (alaS) | 95290 | 201 | C | J | 0 |
| SO3430 | recA protein (recA) | 38220 | 432 | MA | L | 0 |
| SO3440 | enolase (eno) | 42239 | 241 | C | G | 0 |
| SO3441 | CTP synthase (pyrG) | 60453 | 92 | IM | F | 1 |
| SO3455 | GTP pyrophosphokinase (relA) | 83654 | 32 | C | T | 0 |
| SO3466 | riboflavin synthase, beta subunit (ribH) | 16736 | 39 | C | H | 0 |
| SO3467 | 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II, putative (ribBA) | 40513 | 43 | C | H | 0 |
| SO3513 | tryptophan halogenase, putative | 57308 | 40 | C | NA | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO3517 | NADH dehydrogenase (ndh) | 47573 | 38 | IM | C | 0 |
| SO3532 | isoleucyl-tRNA synthetase (ileS) | 106467 | 238 | C | J | 0 |
| SO3545 | OmpA family protein | 40271 | 159 | OM | M | 0 |
| SO3549 | conserved hypothetical protein | 20816 | 39 | C | S | 0 |
| SO3560 | peptidase, M16 family | 104713 | 77 | OM | R | 0 |
| SO3565 | 2,3-cyclic-nucleotide 2-phosphodiesterase (cpdB) | 75167 | 460 | P | F | 0 |
| SO3652 | ribosomal protein L21 (rplU) | 11338 | 157 | C | J | 0 |
| SO3669 | heme transport protein (hugA) | 76438 | 177 | OM | P | 0 |
| SO3690 | ABC transporter, permease protein | 44303 | 34 | IM | V | 4 |
| SO3706 | NupC family protein | 44460 | 72 | IM | F | 10 |
| SO3733 | hypothetical protein | 62873 | 92 | P | NA | 0 |
| SO3772 | conserved hypothetical protein | 20142 | 76 | OM | T | 1 |
| SO3774 | proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase, putative | 115399 | 272 | C | C | 0 |
| SO3779 | ABC transporter, ATP-binding protein CydC (cydC) | 64719 | 49 | IM | C | 6 |
| SO3780 | ABC transporter, ATP-binding protein CydD (cydD) | 70547 | 58 | IM | C | 6 |
| SO3833 | peptide chain release factor 1 (prfA) | 40376 | 55 | C | J | 0 |
| SO3834 | glutamyl-tRNA reductase (hemA) | 45886 | 33 | C | H | 0 |
| SO3837 | ribose-phosphate pyrophosphokinase (prsA) | 34249 | 72 | C | E | 0 |
| SO3844 | peptidase, M13 family | 77272 | 83 | P | O | 0 |
| SO3855 | malate oxidoreductase (sfcA) | 62750 | 47 | U | C | 0 |
| SO3870 | disulfide bond formation protein | 24278 | 34 | IM | O | 4 |
| SO3896 | outer membrane porin, putative | 39874 | 509 | OM | M | 0 |
| SO3897 | DNA topoisomerase IV, A subunit (parC) | 84765 | 55 | C | L | 0 |
| SO3906 | conserved hypothetical protein | 35988 | 39 | OM | S | 1 |
| SO3915 | conserved hypothetical protein | 16558 | 89 | E | N | 0 |
| SO3927 | ribosomal protein L9 (rplI) | 15653 | 80 | C | J | 0 |
| SO3930 | ribosomal protein S6 (rpsF) | 14993 | 37 | C | J | 0 |
| SO3937 | adenylosuccinate synthetase (purA) | 47239 | 125 | C | F | 0 |
| SO3940 | ribosomal protein L13 (rplM) | 15733 | 325 | C | J | 0 |
| SO3942 | serine protease, HtrA/DegQ/DegS family | 46522 | 329 | P | O | 1 |
| SO3988 | aerobic respiration control protein ArcA (arcA) | 27260 | 429 | C | T | 0 |
| SO4012 | hypothetical protein | 10465 | 34 | IM | NA | 1 |
| SO4030 | excinuclease ABC, A subunit (uvrA) | 105924 | 41 | C | L | 0 |
| SO4034 | ATP-dependent RNA helicase DeaD (deaD) | 69567 | 96 | C | K | 0 |
| SO4039 | hydrolase, haloacid dehalogenase-like family | 22728 | 41 | C | R | 0 |
| SO4054 | 5,10-methylenetetrahydrofolate reductase (metF) | 33192 | 33 | MA | E | 0 |
| SO4077 | TonB-dependent receptor, putative | 78344 | 98 | OM | P | 0 |
| SO4080 | conserved hypothetical protein | 85489 | 34 | E | NA | 0 |
| SO4090 | outer membrane efflux protein | 52040 | 33 | OM | M | 0 |
| SO4098 | rod shape-determining protein MreB (mreB) | 37363 | 103 | IM | D | 0 |
| SO4105 | MSHA pilin protein MshA (mshA) | 17759 | 67 | OM | NA | 1 |
| SO4106 | MSHA pilin protein MshB (mshB) | 21375 | 31 | OM | NA | 1 |
| SO4109 | MSHA biogenesis protein MshE (mshE) | 65453 | 65 | P | U | 0 |
| SO4133 | uridine phosphorylase (udp) | 27121 | 158 | C | F | 0 |
| SO4144 | cytochrome c, putative | 52336 | 32 | OM | NA | 0 |
| SO4163 | heat shock protein HslVU, ATPase subunit HslU (hslU) | 49940 | 84 | MA | O | 0 |
| SO4199 | ubiquinone/menaquinone biosynthesis methlytransferase UbiE (ubiE) | 28047 | 31 | C | H | 0 |
| SO4202 | Sec-independent protein translocase protein TatA (tatA) | 9333 | 61 | E | U | 1 |
| SO4211 | preprotein translocase, SecA subunit (secA) | 102787 | 858 | IM | U | 0 |
| SO4215 | cell division protein FtsZ (ftsZ) | 40910 | 133 | P | D | 0 |
| SO4218 | UDP-N-acetylmuramate--alanine ligase (murC) | 53421 | 43 | IM | M | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SO4219 | UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (murG) | 38528 | 73 | IM | M | 0 |
| SO4230 | glycerol kinase (glpK) | 54162 | 39 | C | C | 0 |
| SO4232 | long-chain fatty acid transport protein | 46765 | 39 | OM | I | 0 |
| SO4267 | type I restriction-modification system, R subunit (hsdR-2) | 135459 | 33 | C | V | 0 |
| SO4317 | RTX toxin, putative | 284982 | 35 | E | U | 0 |
| SO4343 | aminotransferase, class V | 41282 | 151 | MA | E | 0 |
| SO4408 | virulence regulator BipA (bipA) | 66111 | 200 | C | T | 0 |
| SO4410 | glutamine synthetase, type I (glnA) | 51912 | 45 | C | E | 0 |
| SO4473 | outer membrane protein, putative | 20074 | 104 | OM | M | 0 |
| SO4513 | formate dehydrogenase, alpha subunit | 106403 | 122 | P | C | 0 |
| SO4520 | oxygen-independent coproporphyrinogen III oxidase, putative | 51142 | 50 | C | H | 0 |
| SO4557 | methyl-accepting chemotaxis protein | 67264 | 157 | IM | N | 2 |
| SO4602 | glycerol-3-phosphate acyltransferase (plsB) | 91846 | 133 | IM | I | 0 |
| SO4669 | DNA polymerase I (polA) | 101959 | 66 | C | L | 0 |
| SO4673 | threonine 3-dehydrogenase (tdh) | 37573 | 62 | C | E | 0 |
| SO4680 | conserved hypothetical protein | 41525 | 33 | IM | M | 0 |
| SO4686 | NAD dependent epimerase/dehydratase family protein | 37390 | 32 | IM | M | 0 |
| SO4693 | multidrug resistance protein, AcrA/AcrE family | 40219 | 102 | OM | M | 0 |
| SO4719 | conserved hypothetical protein | 29348 | 131 | P | H | 0 |
| SO4730 | oxygen-independent coproporphyrinogen III oxidase (hemN) | 53480 | 164 | C | H | 0 |
| SO4731 | adenosine deaminase (add) | 36267 | 34 | C | F | 0 |
| SO4741 | glucosamine--fructose-6-phosphate aminotransferase (isomerizing) (glmS) | 67029 | 136 | IM | M | 0 |
| SO4747 | ATP synthase F1, beta subunit (atpD) | 49856 | 2966 | OM | C | 0 |
| SO4748 | ATP synthase F1, gamma subunit (atpG) | 31609 | 382 | OM | C | 0 |
| SO4749 | ATP synthase F1, alpha subunit (atpA) | 55228 | 1555 | OM | C | 0 |
| SO4758 | glucose-inhibited division protein A (gidA) | 69435 | 130 | C | D | 0 |
| SOA0004 | type II DNA modification methyltransferase | 66390 | 37 | C | NA | 0 |
| SOA0016 | TnSon1, transposase | 112007 | 31 | U | NA | 0 |
| SOA0048 | prolyl oligopeptidase family protein | 72242 | 72 | P | NA | 1 |
| SOA0099 | conserved hypothetical protein | 63665 | 164 | P | NA | 0 |
| SOA0100 | conserved hypothetical protein | 65670 | 38 | P | NA | 0 |
| SOA0106 | methyl-accepting chemotaxis protein | 59803 | 49 | IM | N | 2 |
| SOA0110 | lipoprotein, putative | 146008 | 81 | OM | M | 0 |
| SOA0112 | lipoprotein, putative | 145698 | 278 | OM | M | 0 |
| SOA0140 | hypothetical protein | 87755 | 86 | P | NA | 0 |
| SOA0161 | zinc-binding dehydrogenase | 40772 | 48 | IM | NA | 1 |

[a] Accession no of proteins according to TIGR annotation (http://www.tigr.org/). [b] Mascot protein score. [c] Subcellular location. C: cytoplasmic; E: extracellular; IM: inner membrane; OM: outer membrane; P: periplasmic; MA: membrane associated; and U: unknown. [d] Protein function classification based on COG (http://www.ncbi.nlm.nih.gov/sutils/coxik.cgi?gi=261). C: energy production and converstion; D: cell cycle control, mitosis and meiosis; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; G: carbohydrate transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; J: translation; K: transcription; L: replication, recombination, and repair; M cell wall/membrane biogenesis; N: cell motility; NA: not in COGs; O: posttranslational modification, protein turnover, chaperones; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown; T: signal transduction mechanisms; U: intracellular trafficking and secretion; V: defense mechanisms; and W: extracellular structures. [e] Predicted number of transmembrane helices according to TMHMM (http://www.cbs.dtu.dk/services/TMHMM/).

## CHAPTER 4

# Identification of Protein Interaction Sites, Accessible Residues and Protein Topology in Living Cells by Chemical Cross-Linking


## Abstract

Current chemical cross-linking methods are commonly employed for mapping sites of interaction in purified, known protein complexes. When applied *in vivo*, information on the sites of interaction are unattainable due to overwhelming sample complexity. We present results from a novel cross-linking strategy that allow simultaneous protein-protein interaction and surface topology measurement *in vivo* without prior knowledge of the system. The strategy consists of: (i) cross-linking reaction: intact cell labeling with protein interaction reporters (PIRs); (ii) two-stage mass spectrometric analysis: stage 1 identification of PIR-labeled proteins and construction of a restricted database by 2D-LC/MS/MS; and stage 2 analysis of PIR-labeled peptides by multiplexed LC/FTICR-MS; (iii) data analysis: identification of cross-linked peptides and proteins of origin using accurate mass and other constraints. This strategy was applied to *Shewanella oneidensis* MR-1 bacteria and initial results demonstrate for the first time simultaneous identification of protein-protein interactions and sites of interaction *in vivo*.

## Introduction

An essential component of the goal to elucidate global biological function is the determination of networks of interactions between proteins and among proteins and other biological molecules such as DNAs, RNAs, and ligands. Proteins interact with others through their surface residues to perform various functions. These accessible surface areas are most sensitive to environmental changes and provide important recognition sites for interacting protein partners. Therefore, the determination of these functionally-important amino acids at the protein surface and detection of protein-protein interactions are of critical importance for better understanding global function within biological systems and ultimately, for better biomedical applications, e.g., drug designs in diagnostic and therapeutic research.

Current methods for probing protein-protein interactions or protein surface residues are both experimental and computational. Experimental methodologies include mutagenesis and chemical modification. In the mutation method, functionally-important surface residues are indicated if the replacement of these residues results in changes in protein activity.[1,2] Chemical modification is an alternative method which has long been employed in protein topology studies. Recent advances in mass spectrometry technologies have made it possible to identify sites of incorporation in proteins which are modified with chemical probes. Several commercially-available chemical reagents have been utilized as topological tools to probe solvent accessibility of a variety of amino acid

side chains[3-7] including lysine, arginine, tyrosine, cysteine, etc. Computational methods, on the other hand, provide rapid and high throughput prediction of protein surface and functional residues without tedious lab work. These methods rely on assumptions about protein structures to be applicable to large-scale analysis. For example, one underlying premise of many computational methods is that the important functional residues in a protein sequence that are critical to binding with other proteins or ligands are highly conserved. Therefore, structural comparisons or multiple sequence alignments using evolutionary data can result in identification of these functionally-important surface residues.[8-11] As more and more genomic data are compiled, these approaches will be further refined, but will not diminish the need for experimental approaches that can identify these residues in living cells. In general however, most experimental and computational methods have been applied to individual target proteins in their isolated forms. Such characterization within the interacting, native environment of a protein, where information on protein topology and interactions is most important, is often neglected. This latter fact stems from the difficulty in obtaining this type of information on a complex system and the need for advanced technology that can enable such research.

Chemical cross-linking methods in combination with mass spectrometry have become increasingly important tools for mapping low-resolution topology of proteins and protein complexes (see reviews[12-14]). These approaches are either employed to study protein complexes in purified form,[15-17] or coupled with immuno-affinity methods to probe the potential binding partners for a specific protein of interest.[18-20] The sites of interaction or cross-linker attachments are often successfully identified for purified

protein complex systems with relatively large quantity of proteins. These interacting sites are of particular importance for drug designs since they are potential targets for therapeutic intervention. However, in most *in vivo* applications which can effectively detect protein-protein interactions with the observation of higher MW bands on SDS gels or Western blot images, the information for the sites of interaction is less attainable due to the complex nature inherent with the cross-linking approach. Cross-linking reactions are often carried out with large excessive amount of cross-linkers due to low reaction yield, thus resulting in highly complex mixtures. Three basic types of cross-linking products are most useful: (i) inter-cross-link species allow identification of protein-protein interactions; (ii) intra-cross-link products provide important through-space measures between two cross-linked residues in 3-D structural study of proteins; and (iii) dead-end labeled peptides reveal the solvent accessible residues on protein surfaces. Proteolysis performed prior to MS analysis further increases the degree of complexity since in general, the number of possible cross-linked peptides is equal to the square of the number of tryptic peptides. The problem can be further exacerbated by other factors such as incomplete digestion and posttranslational modifications. Finally, interpretation of MS/MS fragmentation spectra of the cross-linked peptides is challenging since simultaneous fragmentation of the cross-linker itself with one or more peptides can result in spectral complexity that precludes interpretation and/or protein identification.

To overcome the limitations of current cross-linking strategies, we have pursued concepts embodied in a novel class of cross-linkers called Protein Interaction Reporters (PIRs). The spacer chain chemistry of traditional cross-linker is modified in our PIR compounds with the incorporation of two specific mass spectrometry-cleavable bonds, a

mass-encoded reporter tag, and an affinity tag (**Fig. 1**). We applied our first-generation PIR compound to ribonuclease S (RNase S), a known protein complex of S-peptide and S-protein, and demonstrated the utility of the PIR cross-linking concept for detecting protein-protein interactions.[21] When further combining the advantageous features of PIR cross-linkers with software tools and accurate peptide mass-based protein identification, the PIR strategy can enable analysis of protein surface residues and protein-protein interactions for a specific tissue, cell type, or organism *in situ* without any prior knowledge. Here, we describe the application of the PIR strategy to the microbial system, *Shewanella oneidensis* MR-1 and successful identification of all three kinds of PIR-labeled peptides as well as a protein-protein interaction between SecA and a small outer membrane lipoprotein and sites of interaction from intact living cells.

## Results

### PIR structure

The PIR strategy is enabled by engineering two labile bonds in the spacer chain of the cross-linker that can be cleaved with high specificity in the mass spectrometer. This allows the detection of the cross-linked peptide or peptide complex during a low-energy precursor scan and then, detection of the intact peptide masses that are released upon PIR activation. The conceptual mathematical relationships that exist between PIR-labeled precursor, reporter ion and released peptide masses, and how these relationships can be used to facilitate reaction product differentiation are described elsewhere (see **Fig. 6** of ref 21). The measured, released peptide mass information, enzyme digestion specificity,

together with an additional search constraint such as an internal lysine residue, allow unambiguous peptide and protein identification without further mass spectrometry as discussed below. The information from PIR experiments allows protein interactions, actual interacting protein regions, exposed amino acids and close proximity amino acids to be determined from *in vivo* cross-linking experiments. For the current goal of studying interactions in living cells, the second-generation PIR structure incorporates a biotin tag (**Fig. 1**) for affinity enrichment of PIR-labeled products from complex mixtures. RNase S complex was labeled with this biotinylated PIR compound and the same cross-linked peptides were observed as those using the first-generation PIR compound,[21] suggesting that biotin incorporation did not alter the reactivity of the compound.

**Cell membrane penetration**. The cell penetration properties of the PIR compound were investigated with electron microscopy (EM) imaging of PIR-labeled cells. These experiments included both whole cell and thin section EM imaging to ascertain if the proteins on the cell membranes and in the cytoplasm were labeled with PIRs. **Fig. 2** illustrates EM images generated from *S. oneidensis* cells that were reacted with the PIR and from control cells that were not treated with PIR. Anti-biotin primary antibodies and immunogold nanoparticle-conjugated secondary antibodies were used to visualize PIR-labeled proteins. Only those proteins that are PIR-labeled or endogenously biotinylated react and bind with the anti-biotin antibodies and subsequent IgG-conjugated immunogold nanoparticles to result in detected signals (dark dots in the images). Since the thin sections (~100 nM) are much smaller than the average cell dimensions, these experiments allow interrogation inside the cells. As shown in **Fig. 2**, the control cells

exhibited very little reactivity with the anti-biotin and immunogold nanoparticles, while the PIR-treated cells showed significant levels of reactivity on exposed cell surfaces and cell cytosolic regions in thin-section images. Additional control experiments performed with PIR-labeled cells that involved the same immunogold nanoparticle treatment, but excluded the anti-biotin antibody treatment were performed and resulted in no observed reactivity. Therefore, these data indicate that the PIR compound is able to penetrate *S. oneidensis* cell membranes and label both surface proteins and cytosolic proteins in intact *S. oneidensis* cells. From the mass spectrometry experiments described below, we identified many PIR-labeled proteins that are localized on outer and inner membranes, as well as many cytosolic proteins, in agreement with the data from these EM studies. The observation of some abundant cytosolic proteins was also reported by others in their membrane-prepared samples with labeling intact cells by a membrane-impermeable chemical reagent.[22,23] It was thought these cytosolic proteins could come from contamination or non-viable cells. We do not eliminate these possibilities for our case although we washed labeled cells extensively prior to cell lysis to reduce the amount of labeled protein that could have originated from secreted proteins or inadvertently disrupted cells, as discussed below.

**Two-stage analysis**

To fully utilize the advanced features of PIRs, a two-stage mass spectrometric approach was developed (**Fig. 3**). Cells, tissues, or biological fluids of an organism can be labeled with PIR compounds under near physiological conditions. After labeling, proteins are extracted and are divided into two parts for the subsequent stage 1 and stage

2 analyses. Stage 1 analysis is carried out with affinity capture of labeled proteins, followed by digestion and shotgun 2D-LC/MS/MS for protein identification.[24] Stage 1 analysis results in a restricted database that includes primarily PIR-labeled proteins, which provides a "lookup" table for stage 2 analyses. Stage 2 is performed with digestion of all proteins after PIR-labeling, followed by affinity capture of the labeled peptides. These peptides are then subjected to multiplexed LC/FTICR-MS experiments to measure both the intact PIR-labeled peptide masses in the low-energy precursor scan and the masses of the intact peptides in the PIR activation scan. Finally the accurate peptide masses, a software algorithm that employs the specific relationships between precursor, reporter and peptide ion masses, and the restricted database resultant from stage 1 analysis are utilized to identify the type of PIR-labeled peptides, peptide identities, sites of labeling, and proteins of origin. We applied the PIR strategy to determine protein-protein interactions and protein surface topology in the living *S. oneidensis* MR-1 cells using this two-stage approach. The cells were labeled intact with PIR compound for 5 min and then lysed in PBS buffer with inclusion of the detergent NP-40 (2% v/v) to extract membrane and soluble proteins. Before lysis, 5 washing steps were performed with PBS to eliminate any proteins from the labeling solution that were produced either by in inadvertent cell disruption or from normal cellular secretion. We carried out several experiments to optimize PIR labeling reactions and used SDS-PAGE and anti-biotin Western blot analysis to provide relative quantitation of the extent of labeling. For example, we found that 1 mM PIR concentration resulted in greater detected signal, but concentrations higher than this did not further increase label incorporation. 5-min reaction produced nearly the same amount of labeled proteins as

was observed from a 30-min reaction with anti-biotin Western blot analysis, indicating the rapid reaction of this compound. Selected optimization results are provided in **Supplementary Fig. 1**.

Half of the cell lysate was used for stage 1 mass spectrometric analysis and the enriched PIR-labeled proteins were utilized for protein identification through shotgun 2D-LC/MS/MS analysis of tryptic peptides (predominantly unmodified) from this subset of proteins. The goal of stage 1 analysis is to construct a candidate protein pool as complete as possible. Stage 2 analysis provides much more stringent criteria to eliminate false positives. Therefore, all proteins identified by Mascot with significance threshold $p < 0.05$ were included in the restricted database without further filtering. A few proteins that were also identified in the parallel control sample (without PIR labeling), such as some endogenously biotinylated proteins (acetyl CoA carboxylase and acetyl-CoA carboxylase multienzyme complex) and some highly abundant proteins (translation elongation factor Tu and ribosomal proteins), were not excluded from the restricted database. Overall the restricted database consists of a total number of 364 proteins (see **Supplementary Table 1).** Among these, approximately equal proportions (~40%) were classified as cytoplasmic proteins and the cell envelope proteins (outer membrane, periplasmic, and inner membrane), which agrees with the EM immunogold analysis (**Fig. 2**). In addition, the number of total proteins and the fraction of membrane proteins are comparable to the results reported by others who used commercial NHS ester reagents for intact cell labeling,[23,25-27] indicating our PIR compound can effectively label proteins in living cells. A separate manuscript describes detailed classification of the types of

proteins that are labeled by this PIR compound as well as two other PIR compounds (submitted to Journal of Proteome Research).

The novel utility of the PIR approach is based on the mass spectrometry-cleavable features of these compounds. This element allows the precise identification of the sites of labeling in a new way, since after PIR activation, accurate mass measurement and/or MS/MS of the released peptides can then be carried out to establish protein identity and pinpoint labeling sites. Multiplexed LC/MS/MS[28] is best suited for the PIR strategy since multiple co-eluting PIR-labeled peptides can simultaneously be activated. Most unlabeled peptides remain intact with the CID energy tuned to activate labile PIR bonds. Thus, we performed stage 2 analyses using multiplexed LC/FTICR-MS on the second half of PIR-labeled *S. oneidensis* cell lysate. In this case, the labeled cell lysate was digested with trypsin prior to avidin capture. During LC separation, alternating conditions were applied to allow either PIR-labeled peptide precursor ion detection or released reporter and intact peptide ion detection. This is accomplished by alternating collision energies between two scripts, one with low-energy ion accumulation (-4 v), the other with PIR activation energy (-18 v) to efficiently fragment PIR labile bonds but not peptide bonds. The second-generation PIR used in this work fragments at the same labile bonds as was shown with the first-generation non-biotinylated PIR[21] (also see **Supplementary Fig. 2). Fig. 4** illustrates an example dataset from stage 2 analysis of *S. oneidensis* cells, which shows the base peak chromatogram and the extracted ion chromatogram (EIC) for the expected PIR reporter ion (singly-charged ion at nominal *m/z* 1122). The inset in **Fig. 4a** shows the overlaid EICs for an observed PIR-labeled species (dashed-line trace) and the released peptide observed upon PIR activation (solid-line

trace). The observed complementarity of the peak intensity patterns and the high degree of overlap of the elution profiles for these two ions strongly suggest a precursor-product relationship exists between these two species. This information can supplement multiplexed MS/MS analyses to help define precursor-product ion relationships and can be used to further constrain our search results.

**Mass accuracy and protein identification**

To facilitate automated analysis of stage 2 datasets, we have developed a software program called "*X-links*" that reads in the mass lists resultant from multiplexed LC/FTICR-MS experiments described above. The program first segregates the lists of precursor ion masses that are recorded during scans with low-energy conditions (odd-scans) and then lists the masses recorded during PIR activation scans (even scans) (**Supplementary Fig. 3)**. If the reporter ion (neutral mass 1121) is detected in the even scan within the given mass error tolerance (5 ppm), the program will search the data for relationships between precursor, reporter, and released neutral peptide masses to identify dead-end, intra- and inter-cross-linked species using the relationships defined in ref 21. All these calculations are constrained with error tolerance of 10 ppm or less. It should be noted that we use this relatively large tolerance (by FTICR-MS standards) since these experiments are not performed with automated gain control[29] and have not yet utilized any advanced space charge-induced frequency shift algorithms.[30,31] These strategies and others can further decrease mass tolerance of these searches. If a corresponding match is found, the result will be reported in the spreadsheet with inclusion of the type of labeled products, scan number, measured neutral masses of reporter ion, precursor ion, and

product ions, peak intensity, mass measurement error of reporter ion, calculation error between the precursor and its products, etc. Once the PIR-labeled peptide is identified, the peptide mass (without label) is calculated by subtracting the mass of the remaining tag on the peptide, which is known to be 99.0320, from the reported product neutral mass. Then, the single accurate peptide mass is used to search the restricted database to determine peptide identity, sites of PIR incorporation, and protein of origin. The current version of *X-links* identifies PIR-labeled peptides based solely on these calculations as described above with error tolerance of 10 ppm. The *X-links* search results were manually inspected to eliminate false positive hits by examining the original spectra and elution profiles. For example, if the identified precursor did not co-elute exactly with the identified product, then the assigned match was considered to be false positive and was thus eliminated. In the future, additional criteria such as peak intensity comparison and elution profile comparison between the precursor and product will be incorporated in *X-links* to improve the specificity of the search for large-scale, high throughput applications.

Traditionally, mass spectrometry-based protein identification is achieved either by peptide mass fingerprinting of three or more peptides or tandem MS fragmentation of a few peptides (frequently even a single peptide) from an enzymatic digestion. Recently, growing interest in the high performance FTICR-MS technique and more rapid incorporation of FTICR-MS in the field of proteomics have resulted in increasing attention and efforts focused on the effects of mass measurement accuracy on protein identification. Conrads *et al*[32] have proposed that it is possible to use a single peptide mass to identify a protein in the context of a given organism, provided mass measurement accuracy is sufficiently high. The specificity of the accurate mass of a single peptide for

protein identification can be further extended with improved mass measurement accuracy and additional constraints, such as an internal cysteine as reported by Goodlett *et al*[33] or LC retention time as employed in accurate mass and time tag (AMT) strategy pioneered by the Smith group.[34,35]

To assess the feasibility of identifying proteins in *S. oneidensis* using a single PIR-labeled peptide, a simulation analysis (**Fig. 5**) was performed with *in silico* tryptic digestion of all predicted ORFs (4,854 protein sequences) from *S. oneidensis* and the restricted database (364 protein sequences) generated from stage 1. Unambiguous peptide identification using high mass accuracy alone is a very challenging task; however, the PIR strategy reduces the problem by providing an additional search constraint since each peptide contains an internal lysine. In addition, the PIR strategy can further increase confidence of single peptide-based protein identification with the use of a restricted database. In summary, with 1 ppm mass measurement accuracy, the percentage of unique tryptic peptide masses with up to one missed cleavage is 39% for the complete database of *S. oneidensis* and 74% for the restricted database, covering 95% and 100% of all predicted proteins, respectively. In other words, 39% of all the tryptic peptides of *S. oneidensis,* if measured with random mass errors of 1 ppm or less, will uniquely identify a single ORF from the complete genome and 95% of all ORFs will have at least one such peptide. These percentages increase substantially (to 74% and 100% respectively), with consideration of the restricted database. With an additional internal lysine constraint, the percentage of unique peptide masses becomes 68% for the whole proteome and 91% for restricted proteome, with coverage of 92% and 100% of these databases, respectively. Taken together, these simulation results strongly suggest that the use of a single accurate

peptide mass with an internal lysine constraint and a restricted database is a highly stringent set of search constraints. Importantly, the use of the two-stage approach and the restricted protein database will allow extension of the PIR strategy to more complex biological organisms and samples. To investigate this we have begun testing of the current PIR compound with HeLa cells, recorded confocal microscopy images showing label incorporation on the cell surfaces (**Supplementary Fig. 4**), and have identified more than 200 proteins in stage 1 analysis.

Since mass measurement accuracy is the most important factor in single peptide mass-based protein identification, the FTICR-MS instrument was carefully calibrated to mass accuracy of 1 ppm or less with direct infusion of BSA tryptic peptide mixture immediately prior to LC/FTICR-MS analysis. Sometimes it was necessary to lower the trapping potentials and excite potential to get better mass accuracy at the sacrifice of sensitivity. Thus, the detection of peptides in stage 2 analysis with current instrumentation is less sensitive than stage 1 measurements. Currently with external calibration, the average mass accuracy for LC/FTICR-MS data we can achieve with our instrument is approximately ± 5 ppm. Thus, 5 ppm is the mass tolerance limit we used for protein identification with a single peptide mass search against the restricted database. However, this results in tight search constraints and reduction in false positive identification, since only species found to be labeled in both stages result in identification of labeled peptides. A summary of peptides and proteins identified in stage 2 analyses of *S. oneidensis* cells is shown in **Table 1**.

**Validation of PIR-labeled peptides**

Although the single peptide mass-based strategy for protein identification is highly reliable as discussed above, validation of a few examples was necessary for the initial application of this technology. This requires MS/MS activation of the PIR-labeled peptide, isolation and fragmentation of the released peptide, thus $MS^3$. **Fig. 6** illustrates the $MS^3$ validation of a dead-end labeled peptide found in *X-links* analysis and identified by accurate mass to be a tryptic peptide of an outer membrane protein, decaheme cytochrome *c* (OmcA). **Fig. 6a** shows the multiplexed LC/FTICR-MS data used to identify this peptide. Shown are two overlaid spectra from low-energy (red trace) and PIR activation (black trace) conditions. **Fig. 6b** illustrates the $MS^3$ results obtained by isolating the PIR-labeled peptide (nominal *m/z* $1052^{3+}$), activating the PIR bonds, releasing the peptide, isolating the intact peptide ion at *m/z* $958^{2+}$ and then fragmenting it. These results were obtained during LC separation of stage 2 samples using a quadrupole ion trap mass spectrometer. The observed fragmentation pattern shows excellent agreement with the identified sequence based on *X-links* and accurate mass analysis. A prominent series of *y* ions (consecutive from $y_3$ to $y_{13}$) were observed, thus confirming the assignment of this dead-end labeled species with the labeling site at the internal lysine residue.

**Identification of an inter-cross-linked peptide complex**

The initial example of a cellular inter-cross-linked product was found to be that of a dimeric species where both cross-linked peptides are of the same mass. The overlaid EICs of the precursor ion (nominal *m/z* $732^{3+}$) and the released peptide ion (nominal *m/z* 537) are shown in **Fig. 4a** inset. *X-links* analysis automatically tested multiplexed

LC/FTICR-MS data for this possibility to allow this identification. Two degenerate mass sequences KYK and YKK were identified when the unmodified peptide mass, 437.2634, was used to search the restricted database. First consideration suggested that this inter-cross-linked peptide pair might be from a homo-dimeric protein complex. However when further validation of KYK or YKK sequence assignment was performed with $MS^3$ analyses, fragments from both labeled KYK and YKK were observed. **Fig. 7a** illustrates the MS/MS pattern observed after isolating the PIR-labeled peptide complex (nominal *m/z* $732^{3+}$) and activating the cleavable bonds in the PIR. This spectrum confirms the *X-links* assignment of an inter-cross-linked species of the same mass from the multiplexed LC/FTICR-MS data, since the spectrum shows peaks from both a released intact peptide (nominal *m/z* 537) and a remaining PIR-labeled peptide (nominal *m/z* $830^{2+}$). Further MS/MS of the ion at *m/z* $830^{2+}$ released the reporter ion (nominal *m/z* 1122) and intact peptide ion (nominal *m/z* 537). These data definitively identified an inter-cross-linked species with identical mass peptides from PIR-labeled *S. oneidensis* cells. Next, the released peptide ions (nominal *m/z* 537) were isolated and fragmented resulting in a spectrum that showed a majority of intense peaks that could be assigned to both sequences KYK and YKK with two peaks unique to each. Peak assignments for both sequences are illustrated in **Fig. 7b**. The sequence YKK could be attributed to only one protein from the restricted database, major outer membrane lipoprotein (SO1295), while the sequence KYK resulted in three possible proteins, i.e., SecA, FabB, and TPR domain protein, from the restricted database (**Table 1**). Interestingly, the observation of the $y_1$ (*m/z* 246.19) and $y_2$ (*m/z* 374.16) ions demonstrated that the labeling occurred at the carboxyl terminal lysine, instead of the middle lysine, of the sequence YKK. Thus, this

labeled lysine must be the carboxyl terminus of the entire protein sequence. Otherwise, the modification to this lysine would have prevented trypsin cleavage at this position and the tryptic YKK peptide would not be generated. Conversely, in the entire *S. oneidensis* genome database, only major outer membrane lipoprotein has a carboxyl terminal YKK sequence, thus we can conclude one peptide from this inter-cross-linked pair was definitely from major outer membrane lipoprotein. It has been reported that translocation of the homologous lipoprotein sequence across the inner membrane of *E. coli* is dependent on the functionality of SecA.[36-38] SecA is a major component of multi-subunit translocase enzyme system that catalyzes protein translocation across inner cell membranes.[39] The interactions between lipoprotein and the other two proteins FabB and TPR domain protein, to our knowledge, have not been reported. When searching the whole *S. oneidensis* database, three other proteins could conceivably produce the tryptic peptide KYK. These are type I restriction-modification system, S subunit (hsdS-2) (MW 43 kD), sigma-54 dependent response regulator (MW 55 kD), and chemotaxis motB protein (motB) (MW 31 kD). All these three proteins have reasonable size and fairly large number of lysine and arginine residues. However, they were not identified in our stage 1 analysis, suggesting that these proteins were less accessible to our PIR compound. Furthermore, since the shotgun LC/MS/MS by ion trap mass spectrometer used in the stage 1 analysis currently showed greater sensitivity than multiplexed LC/MS/MS with the FTICR mass spectrometer used in stage 2 analysis, it seems very unlikely that we would detect a PIR-labeled peptide from one of these three proteins in stage 2 analysis only. Thus it is most likely the inter-cross-linked peptides were from SecA and the outer membrane lipoprotein. To further confirm these results, we performed anti-SecA

Western blot analysis of the gel-separated PIR-labeled proteins from *S. oneidensis* cells (**Fig. 7c**). The PIR-labeled cell lysate clearly showed a band for SecA monomer (MW 103 kDa), a cross-linked band above SecA at approximately MW 112 kDa, which corresponds approximately to the sum of the molecular weight of SecA and the major outer membrane lipoprotein (MW 9 kDa). These results constitute the first *in vivo* evidence suggesting a protein-protein interaction between SecA and outer membrane lipoprotein in *S. oneidensis* together with the visualization of critical sites of interaction information.

**Discussion**

Chemical cross-linking strategies are common methods to probe protein-protein interactions and protein complex topologies. Most applications of cross-linking have revolved about a target protein or protein complex in a purified form or in the context of limited environment. However, a living organism is a dynamically changing system, in which many protein-protein interactions are transient and protein conformations change with time. Immunoprecipitation of a target protein coupled with SDS-PAGE, mass spectrometry, and cross-linking methods allow one to identify the interacting partners of the target protein *in vivo*. However, surface topological information for *in vivo* protein-protein interactions has been un-obtainable. Determination of these sites of interaction for protein complexes is critical for understanding biological function and for drug design. To this end, we present a general strategy that shows great promise to map protein-protein interaction and protein surface topology in cells or organisms under their

native conditions without prior knowledge of the protein locations, activities, and functions.

In our PIR experiments, stage 1 analysis resulted in identification of 364 proteins. Among these proteins, approximately 40% are membrane proteins; some are known to be critical for electron transport; and some are involved in protein synthesis and transport. For example, outer membrane cytochromes OmcA and OmcB are exposed on the outer membrane surfaces to allow direct electron transfer to extracellular electron acceptors;[40] outer membrane protein MtrB is required for proper localization of OmcA and OmcB to outer membrane;[41] and cytosolic proteins SecA and SecB are known to interact with SecYEG complex for preprotein translocation.[39] Stage 2 analysis resulted in identification of 17 PIR-labeled peptides, of which most were dead-end labeled. As compared to stage 1 analysis, the throughput of stage 2 seems very low. Several factors contribute to this result. One primary reason for this was that the instrument used for accurate mass analysis in stage 2 was substantially less sensitive than the instrument used for stage 1 analysis based on our routine observations. A second reason is that NHS esters hydrolyze rapidly in aqueous solution. NMR and LC/MS analyses of the material cleaved from the resin following the final addition of the N-hydroxy succinamides suggest that 60 to 70% of the compound cleaved from the synthesis resin was produced with both NHS esters intact. The primary impurity with the current PIR is a compound with a single NHS ester and the other cross-linker arm hydrolyzed to succinic acid. We were able to purify the compound with 2 intact esters using HPLC. However, the measured rates of hydrolysis of the compound at pH 7.2 suggest a significant fraction will hydrolyze during reaction, resulting in dead-end labeling. On the other hand, the fast

reaction rate of the NHS ester with primary amines at pH 7.2 makes this an attractive group to utilize for chemical cross-linkers. Therefore the initial PIR development focused on this functional group. Future studies will evaluate additional functional groups with the PIR strategy. A third consideration involves the number of peptides per protein that are available for detection in stage 1 analysis. This number may be as high as 50 to 100 for each large protein that is affinity enriched, while those that are purified through stage 2 are much smaller in number, normally fewer than 10. Furthermore, any individual labeled protein may be PIR-labeled in a variety of ways, i.e., dead-end, intra- and inter-linked. All these products are purified through stage 1 analysis and give rise to a common set of unmodified tryptic peptides. Stage 2 analysis however, results in division of each of these products into uniquely detected species and therefore, lower amounts of each species are available for detection. Thus, fewer species are expected in stage 2 analysis as compared to stage 1. With increased labeling efficiency and increased sensitivity of newer high performance mass analyzers, the efficiency of stage 2 analysis can be improved. In addition, future experiments can employ larger sample sizes for stage 2 analysis to improve the ability to detect labeled peptides. Finally, the current PIR compound is relatively large (MW 1550 Da), which results in an average MW greater than 5000 Da for two average-sized inter-cross-linked tryptic peptides. Current LC/MS/MS methods are most sensitive for the peptides with MW less than 3000 Da. Therefore, further optimizing PIR structure with reduced size will increase efficiency of analysis and improve throughput. Nevertheless, as the first proof-of-concept, identification of these 17 PIR-labeled peptides with the current compound illustrates the

PIR strategy and shows how this chemical-based approach can enable determination of protein-protein interactions with key topological information *in vivo*.

As mentioned earlier, dead-end labeling can occur in two forms because the NHS ester can hydrolyze in water before reacting with proteins, or can rapidly react by displacement with low level of ammonia impurity present in the reaction buffer.[42] These two forms of dead-end products differ by 1 Da (OH versus $NH_2$). The given example of dead-end labeled peptide (**Fig. 7a**) was actually an ammonia displacement dead-end instead of water displacement (hydrolysis) dead-end. Most PIR-labeled peptides from the present study were identified in both forms which generated the same intact peptide with the remaining tag after PIR activation. Dead-end labeling reveals pertinent information about residue exposure and is useful for establishing protein topological features within cellular environments.

Importantly our initial PIR application has resulted in the identification of inter-cross-linked peptides, both of which have the same mass. A likely candidate for this type of observation would be from a cross-linked homodimeric species. Surprisingly, further $MS^3$ experimental results revealed the presence of a heterodimeric species with peptide sequences KYK and YKK. The result obtained from anti-SecA Western blot analysis further supported the observation that a heterodimer between SecA and major outer membrane lipoprotein was cross-linked by the PIR. Several groups have reported that SecA is involved in translocation of a small outer membrane lipoprotein (murein lipoprotein) in *E. coli*.[36-38] SecA from *S. oneidensis* is highly similar (72% sequence identity) to its *E. coli* homologue. Furthermore, the PIR-labeled sequence KYK in *S. oneidensis* (from residue 888 to 901) located near carboxyl terminus of SecA is

conserved in all bacterial species as shown by BLAST sequence alignments, suggesting strong functional significance of this region of the protein. Previous studies by others have also shown that the carboxyl terminus of SecA is exposed and accessible for modifications or enzyme activities.[43,44] On the other hand, the putative major outer membrane lipoprotein from *S. oneidensis* was identified with mass spectrometric studies by others as well.[45] BLAST sequence alignment suggests that major outer membrane lipoprotein in *S. oneidensis* shares 35% sequence identity and 57% similarity to murein lipoprotein in *E. coli*. This small lipoprotein, one of the most abundant proteins in bacteria, is synthesized in the cytoplasm and exported to the outer membrane through a Sec translocase-dependent mechanism.[36-38] Previous studies showed that the lipoprotein exists as a homotrimer in the outer membrane with one of the three subunits covalently linked to the periplasmic peptidoglycan (murein) through the ε-amino group of the carboxyl-terminal lysine.[46-48] The carboxyl terminal sequence YKK of the lipoprotein in *S. oneidensis* is conservatively substituted with YRK in *E. coli.* This carboxyl terminal sequence, YRK or YKK, conserved in most gram-negative bacteria, has been suggested by others that it can serve as a recognition site for other protein binding activity[46] and an attractive target for therapeutic design.[48] In our PIR experiments, aside from the detection of the inter-cross-linked peptide pair, the dead-end labeled sequence KYK and YKK ($MS^3$ data showed fragments from both sequences) both in ammonia displacement version (*m/z* $887.9^{2+}$) and water displacement version (*m/z* $888.4^{2+}$) were observed from multiplexed LC/FTICR-MS analysis as well (**Supplementary Fig. 5**). In addition, the SecA monomer band appeared thicker and at higher MW in the PIR-treated sample than that in the control sample (without PIR) (**Fig. 7c**), indicating dead-end labeling occurred

in SecA. All these observations from us and others[43,44,46-48] strongly support the conclusion that the C-terminus of both SecA and lipoprotein are exposed and accessible. The present study shows the first *in vivo* evidence of the interaction of SecA and major outer membrane lipoprotein and indicates that the C-terminal regions exist in close proximity and likely help facilitate their interaction.

Taken together, our initial results using the PIR strategy on intact living cells of *S. oneidensis* have demonstrated the capability for probing protein surface topology and protein-protein interactions simultaneously without the need of any prior knowledge of the system. This strategy can be further used as an *in vivo* tool to investigate, snapshot, and compare surface-exposed residues and potential binding sites of proteins and protein complexes in cells at different growth phases, culture conditions, cell cycles, external stresses, etc. Future work includes further improving the throughput, efficiency, and mass accuracy for stage 2 analysis, and optimization of the PIR structure with reduced size and better solubility, which will ultimately allow large-scale, high throughput profiling protein surface topologies and protein-protein interactions in living cells.

## Methods

**Materials.** All chemicals were purchased from Sigma (St. Louis, MO) unless otherwise noted. Fmoc-protected amino acids and HMPB-MBHA resins used for PIR synthesis were purchased from Novabiochem (San Diego, CA). Water used for preparing buffer and solution was 18-MΩ deionized water prepared with Barnstead Nanopure Water Systems.

**PIR synthesis.** The PIR is synthesized as described previously[21] with two additional coupling steps. Glycine was coupled to HMPB-MBHA (4-hydroxymethyl-3-methoxyphenoxybutyric acid) resin and then the biotin group was introduced in the form of N-α-Fmoc-N-ε-biotinyl-L-lysine. The second lysine in the form of N-α,ε-di-Fmoc-L-lysine was coupled to the biotinylated lysine and was then used as the branch point for the cross-linker. The Rink groups and N-hydroxysuccinimides (NHS) were then coupled as described previously.[21] Coupling of Fmoc derivatives was monitored by absorbance of the released Fmoc group at 301 nm after each deprotection step (**Supplementary Fig. 6**) to ascertain reaction step efficiency. Additionally, LC/MS and NMR were used to ascertain compound purity following cleavage from the synthesis resin, drying and redesolving in Dimethylsulfoxide (**Supplementary Fig. 7**). The original crude product appeared 50 to 70% pure after cleavage from the resin, due to incomplete NHS ester formation on the succinic acid termini. The primary impurity is the PIR compound with only a single ester and a succinic acid group for the other arm. This product can only yield dead-end labeled product and will increase the fraction of these products that are detected in stage 2 analysis. Consequently, semi-preparative scale LC was carried out on the crude product to reduce single ester products and estimated purity of the LC-purified compound was ~90% (**Supplementary Fig. 8)**. LC/MS analysis was also used to estimate rates of hydrolysis of the PIR. The half-lives for hydrolysis of the current compound at pH = 3.0 and 7.5 are approximately 8 hours and 10 minutes, respectively (**Supplementary Fig. 9**). It should be noted that all protein identification results reported here were achieved with crude PIR product. However, we performed collection of the

LC-purified peak of the intact PIR compound, repeated on-cell labeling and mass spectrometry analysis and detected the same YKY-KYK interlinked product as was originally discovered with the unpurified PIR compound (**Supplementary Fig. 10**). In addition, the purified compound was used to determine the molar absorption coefficient (ε) at 280 nm, which is approximately 8,400.

**Electron microcopy (EM) cell imaging.** *S. oneidensis* MR-1 cells in PBS suspension with or without PIR labeling were loaded onto the carbon-coated nickel grids for immunogold EM detection of surface proteins labeling. Fixation buffer (2% paraformaldehyde, 1.25% glutaraldehyde, 0.05 M phosphate buffer, pH 7.2) was used to incubate with sample grids for 15 min with subsequent PBS washing 2 times. The residual aldehyde was then quenched by incubation with 0.02 M glycine in PBS for 5 min. Anti-biotin monoclonal antibody (Sigma, St. Louis, MO) was added in PBS-BSA (20 mM phosphate, 500 mM NaCl, 0.5% BSA, 0.1% Gelatin, 0.05% Tween20, pH 7.4) at 1:25 ratio for primary antibody incubation for 10 min with occasional shaking at room temperature. After washing with PBS-BSA 3 times, 10 nm colloidal gold particle conjugated with goat-anti-mouse IgGs (Sigma, St. Louis, MO) at 1:50 ratio in PBS-BSA was added for secondary antibody incubation for 30 min. Cells were washed 3 times with PBS-BSA before the final fixation of cells and antibodies using 1% glutaraldehyde in PBS for 15 min. The fixatives were removed by buffer (20 mM phosphate, 150 mM NaCl, 4% BSA, 2 mM $NaN_3$, pH 7.4). The sample grids were examined by EM using JEOL 1200 EX transmission electron microscope after staining with uranyl acetate for contrast enhancement. To validate PIR permeability, the immunogold EM assay was also

carried out using thin sections of *S. oneidensis* MR-1 cells. The labeled and unlabeled cells were treated with a fixation buffer (2% paraformaldehyde, 1.25% glutaraldehyde, 0.05 M phosphate buffer, pH 7.2) for an hour. Sample dehydration was performed in a series of graded ethanol (10%, 25%, 50%, 75%, 90%, and 100%) for consecutive overnight periods. The dehydrated samples were embedded in a LR White resin for infiltration and polymerization. The thin sections were generated with a diamond knife mounted in an ultracut microtome for loading onto carbon-coated nickel grids. First, the grids were floated on a drop of water for 10 min. PBS-BSA was then incubated with sample grids for 5 min to block non-specific binding. A solution of 1:25 diluted anti-biotin monoclonal antibody was added in PBS-BSA for a one hour incubation period with the section grids. After rinsing with PBS-BSA 3 times, 10 nm colloidal gold particle labeled goat-anti-mouse IgGs at 1:50 dilution ratio were incubated with sample grids for 30 min. Sections were then washed with PBS 3 times and post-fixed by 1% glutaraldehyde in PBS for 3 min. Deionized water droplets were used to rinse the section grids for 10 min before EM examination. Uranyl acetate was also used to increase the contrast of section images. Negative controls include the unlabeled cells and the labeled cells without primary antibody incubation. JEOL 1200 EX transmission electron microscope was utilized to perform EM imaging.

**Western blot analysis.** Equal amounts of lysates from cells treated with and without PIR labeling were separated by 8% SDS-PAGE gel and then transferred to a nitrocellulose membrane (Whatman, Sanford, ME) with a Trans-Blot® semidry transfer cell (BioRad, Hercules, CA). After transferring, the membrane was blocked with 5%

nonfat milk in TBS overnight at 4$^{o}$C followed by incubating with the primary antibody anti-SecA serum at 1:5000 dilution for an hour at room temperature. Anti-SecA serum was kindly provided by Dr. Donald B. Oliver (Wesleyan University). And finally the membrane was probed by HRP-conjugated secondary antibody and chemiluminescence peroxidase substrate (Sigma, St. Louis, MO).

**Sample preparation.** *S. oneidensis* MR-1 cells, ATCC 700550, were obtained from American Type Culture Collection (Manassas, VA) and maintained in Luria Bertani (LB) broth. 200 mL sterile LB broth was inoculated with 4 mL of overnight culture and stirred with a rotary shaker at 100 rpm at room temperature. Cells were harvested at mid-log phase ($OD_{600}$ = 1.2) by centrifuging at 3,200 rpm for 20 min. Pelleted cells were washed three times in 40 mL ice-chilled PBS buffer (150 mM sodium phosphate, 100 mM NaCl, pH 7.5). After the final wash, the cells are resuspended in 1 mL PBS and the PIR compound was added to cell suspension to a final concentration of 1 mM. This concentration appeared to be optimal, as labeling at 0.1 mM produced fewer labeled proteins (**Supplementary Fig. 1**) and 10mM did not result in more labeled proteins. Additionally, we investigated multiple labeling steps where, after an initial reaction period of 5 to 15 minutes, the cells were pelleted, supernatant removed, the cells were resuspended in PBS and another PIR reaction was carried out (**Supplementary Fig. 1**). Two consecutive reactions did produce a slight increase in the detected PIR-labeled proteins, but subsequent additions did not. All subsequent experiments involved only a single reaction step. The reactions were carried out at room temperature with shaking for 5 min and then terminated by quenching with 50 μL 1 M Tris buffer (pH 7.5). The cells

were then washed several times with 1 mL additions of PBS to remove secreted proteins and any unreacted PIR. Each wash solution was collected and tested for the presence of labeled proteins and these studies indicated that by the 3rd wash, no labeled proteins were detected by Western blot analysis (**Supplementary Fig. 1**). Therefore, 5 wash steps were employed following PIR reaction for all subsequent studies. The cells were then lysed in 1 mL PBS/2% NP-40 (v/v) by sonication for 2 min. The lysed cells were centrifuged at 15,000 rpm at 4$^{o}$C for 40 min. The pellets were discarded and supernatants were collected in a clean tube. Protein concentration was determined by Bradford protein assay (Bio-Rad, Hercules, CA). The labeled cell lysates were passed through a Dextran de-salting column (Pierce, Rockford, IL) to remove residual unreacted PIR and salts and were then split to two parts for the following stage 1 and stage 2 sample preparations. For stage 1, the labeled proteins were captured and separated from whole cell lysate by an immobilized monomeric avidin column (Pierce, Rockford, IL) with a 1 mL bed volume. After incubation of labeled cell lysate in avidin column for 30 min, the column was washed extensively with 1 mL buffer A (150 mM NaCl, 10 mM Tris, 2 mM EDTA, 0.1% NP-40, and 0.1% Triton X-100, pH 7.5) 4 times, 1 mL buffer B (500 mM NaCl, 10 mM Tris, 2 mM EDTA, 0.1% NP-40, and 0.1% Triton X-100, pH 7.5) 3 times, and 1 mL buffer C (10mM Tris, pH 7.5) 3 times. The labeled proteins were then eluted with 100 μL of elution buffer (8 M urea, 2% SDS, and 5 mM d-biotin). The eluent was TCA precipitated to remove salts and detergent. The washed pellet was resuspended in a digestion buffer (100 uL 100 mM $NH_4HCO_3$, pH 7.8, and 8 M urea). Proteins are reduced with DTT at final concentration of 5 mM and alkylated with iodoacetamide at final concentration of 25 mM. Then the solution was diluted 4 times with 100 mM

$NH_4HCO_3$ buffer to a total volume of 400 uL and 20 μg trypsin (Promega, Madison, WI) was added for overnight digestion at 37°C. 100 μL of tryptic digestion was used for C18 ZipTip® (Millipore, Billerica, MA) to remove salts. The concentrated peptides were finally resuspended in 20 μL 0.1% TFA for 2D-LC/MS/MS analysis.

For stage 2 analysis, the labeled cell lysate was digested before avidin capture. First, the cell lysate was precipitated with TCA to remove detergents and salts and resuspended in 8M urea buffer (volume 200 uL). The pH was adjusted to 8.0 using solid $NH_4HCO_3$. Proteins are then reduced in 10 mM DTT and alkylated in 25 mM iodoacetamide. Next, the solution was diluted 4-fold with 100 mM $NH_4HCO_3$ buffer and 20-50 μg trypsin was added for overnight digestion at 37°C. The digest solution was then passed through the immobilized avidin column using the same capture and washing procedure as described in stage 1. In this case however, the labeled peptides were eluted in 70% acetonitrile and 0.5% TFA. The eluent was evacuated to dryness and then resuspended in 20 μL 10% acetonitrile and 0.1% TFA for multiplexed LC/FTICR-MS.

**2D-LC/MS/MS**. Data-dependent nano 2D-LC/MS/MS was performed using ion trap mass spectrometer (Esquire HCT, Bruker Daltonics, Billerica, MA) equipped with a nano-ESI source and nano-HPLC systems (Ultimate, Dionex, Sunnyvale, CA). For 2D-LC/MS/MS, a strong cation exchange (SCX) cartridge (500 μm × 15 mm, 5 μm, Dionex) was used for the first dimension separation and a nano C18 reversed phase column (C18 PepMap, 75 μm × 150 mm, 3 μm, 100 Å, Dionex) was used for the second dimension separation. 10 μL salt plug (10, 25, 50, 100, 250, 500, and 1000 mM ammonium acetate in 0.1% TFA and 2% acetonitrile) was used for each salt step elution of SCX cartridge.

Samples and salt plugs were injected to the SCX cartridge by autosampler and the eluent after SCX cartridge was trapped in a micro pre-column (C18 PepMap, 300 μm × 1 mm, 5 μm, Dionex). Pre-column was washed with solvent A (0.1% TFA, 2% acetonitrile in water) at a flow rate of 50 μL/min for 3 min and then peptides were eluted at a flow rate of 200 nL/min to the reversed phase nano-column and separated using the following gradient: 0% B for 0-3 min, 0-15% B for 3-15 min, 15-25% B for 15-60 min, 25-40% B for 60-90 min, 95% B for 90-105 min, and 0% B for 105-120 min. Solvent B was 0.1% TFA in 95% acetonitrile. Data analysis and protein identification were performed using Bruker Daltonics *DataAnalysis* software (version 3.1) and *BioTools* software (version 2.2).

**Multiplexed LC/FTICR-MS.** A Bruker Daltonics Apex-Q 7T FTICR mass spectrometer (Billerica, MA) was used for multiplexed LC/FTICR-MS experiments. The nano-ESI source, nano-LC systems, and the reversed phase nano-column were the same as those used with the ion trap mass spectrometer. LC gradient was the same as described above. Ions formed by nano-ESI source enter the instrument through a glass capillary and then pass through a hexapole followed by a quadrupole and then a second hexapole, which is used for ion accumulation and PIR activation. After the second hexapole, ions are guided through a series of ion transfer optics and enter the ICR cell where ions are detected after excitation of cyclotron motion, resulting in accurate mass measurements. *Xmass* 7.0.6 and *HyStar* 3.0 software programs were used to acquire LC/FTICR-MS data. All datasets were acquired with 128 k points and each spectrum was resultant from averaging 4 scans. The optimized collision energy, -18 volts, which

efficiently fragmented the labile bonds in PIR but not peptide bonds, was applied during alternating scans throughout the LC elution. External calibration of the instrument was performed with direct infusion of BSA tryptic peptides immediately before the LC/FTICR-MS experiments.

**Data analysis.** FTICR-MS data were analyzed using the program *ICR-2LS*. The time-domain ICR signals were apodized with a Welch function, zero-filled once and then Fourier transformed before calibration. The spectrum was calibrated with 6 or more peaks to about 1 ppm or less and then the calibration equation was applied to instrument. All masses were reported in neutral mass forms in a .PEK text file which was further used by the program *X-links*. Protein identification using single peptide mass searching against the restricted database was also performed with the *ICR-2LS* program. GPMAW (version 6.21) was used to calculate MS/MS fragmentation. The database containing the entire genome of *S. oneidensis* MR-1 (4,854 ORFs) was downloaded from The Institute for Genomic Research (TIGR) (http://www.tigr.org/).

**Acknowledgements**

We thank Dr. Donald B. Oliver (Wesleyan University) for providing us anti-SecA serum. This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

## References


1. Wells, J.A. Binding in the growth hormone receptor complex. *Proc Natl Acad Sci U S A* **93**, 1-6 (1996).
2. Clarkson, M.J., Wells, J.R., Gibson, F., Saint, R. & Tremethick, D.J. Regions of variant histone His2AvD required for Drosophila development. *Nature* **399**, 694-7 (1999).
3. Suckau, D., Mak, M. & Przybylski, M. Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc Natl Acad Sci U S A* **89**, 5630-4 (1992).
4. Ohguro, H., Palczewski, K., Walsh, K.A. & Johnson, R.S. Topographic study of arrestin using differential chemical modifications and hydrogen/deuterium exchange. *Protein Sci* **3**, 2428-34 (1994).
5. ZAPPACOSTA, F. *et al*. Surface topology of Minibody by selective chemical modifications and mass spectrometry. *Protein Sci* **6**, 1901-1909 (1997).
6. Leite, J.F. & Cascio, M. Probing the topology of the glycine receptor by chemical modification coupled to mass spectrometry. *Biochemistry* **41**, 6140-8 (2002).
7. Silverman, J.A. & Harbury, P.B. Rapid Mapping of Protein Structure, Interactions, and Ligand Binding by Misincorporation Proton-Alkyl Exchange. *J. Biol. Chem.* **277**, 30968-30975 (2002).
8. ROST, B. & Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins* **20**, 216-226 (1994).
9. Bogan, A.A. & Thorn, K.S. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **280**, 1-9 (1998).

10. Lockless, S.W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-9 (1999).

11. Ferre, F., Ausiello, G., Zanzoni, A. & Helmer-Citterich, M. SURFACE: a database of protein surface regions for functional annotation. *Nucleic Acids Res* **32**, D240-4 (2004).

12. Back, J.W., de Jong, L., Muijsers, A.O. & de Koster, C.G. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* **331**, 303-13 (2003).

13. Trakselis, M.A., Alley, S.C. & Ishmael, F.T. Identification and Mapping of Protein-Protein Interactions by a Combination of Cross-Linking, Cleavage, and Proteomics. *Bioconjugate Chem.* **16**, 741-750 (2005).

14. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* **25**, 663-682 (2006).

15. Trester-Zedlitz, M. *et al*. A modular cross-linking approach for exploring protein interactions. *J Am Chem Soc* **125**, 2416-25 (2003).

16. Chang, Z., Kuchar, J. & Hausinger, R.P. Chemical cross-linking and mass spectrometric identification of sites of interaction for UreD, UreF, and urease. *J Biol Chem* **279**, 15305-13 (2004).

17. Petrotchenko, E.V., Olkhovik, V.K. & Borchers, C.H. Isotopically Coded Cleavable Cross-linker for Studying Protein-Protein Interaction and Protein Complexes. *Mol Cell Proteomics* **4**, 1167-1179 (2005).

18. Chen, Y., Ebright, Y.W. & Ebright, R.H. Identification of the target of a transcription activator protein by protein-protein photocrosslinking. *Science* **265**, 90-2 (1994).

19. Wells, J. & Farnham, P.J. Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation. *Methods* **26**, 48-56 (2002).

20. Schmitt-Ulms, G. *et al*. Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. *Nat Biotechnol* **22**, 724-31 (2004).

21. Tang, X., Munske, G.R., Siems, W.F. & Bruce, J.E. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem* **77**, 311-8 (2005).

22. Zhao, Y., Zhang, W. & Kho, Y. Proteomic analysis of integral plasma membrane proteins. *Anal Chem* **76**, 1817-23 (2004).

23. Nunomura, K. *et al*. Cell Surface Labeling and Mass Spectrometry Reveal Diversity of Cell Surface Markers and Signaling Molecules Expressed in Undifferentiated Mouse Embryonic Stem Cells. *Mol Cell Proteomics* **4**, 1968-1976 (2005).

24. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7 (2001).

25. Chen, W.N. *et al*. Detection of in situ labeled cell surface proteins by mass spectrometry: application to the membrane subproteome of human mammary epithelial cells. *Proteomics* **3**, 1647-51 (2003).

26. Jang, J.H. & Hanash, S. Profiling of the cell surface proteome. *Proteomics* **3**, 1947-54 (2003).

27. Goshe, M.B., Blonder, J. & Smith, R.D. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J Proteome Res* **2**, 153-61 (2003).

28. Masselon, C. *et al*. Identification of tryptic peptides from large databases using multiplexed tandem mass spectrometry: simulations and experimental results. *Proteomics* **3**, 1279-86 (2003).

29. Peterman, S.M., Dufresne, C.P. & Horning, S. The Use of a Hybrid Linear Trap/FT-ICR Mass Spectrometer for On-Line High Resolution/High Mass Accuracy Bottom-Up Sequencing. *J Biomol Tech* **16**, 112-124 (2005).

30. Kaiser, N.K., Anderson, G.A. & Bruce, J.E. Improved mass accuracy for tandem mass spectrometry. *J Am Soc Mass Spectrom* **16**, 463-70 (2005).

31. Wu, S. *et al*. Increased protein identification capabilities through novel tandem MS calibration strategies. *J Proteome Res* **4**, 1434-41 (2005).

32. Conrads, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L. & Smith, R.D. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem* **72**, 3349-54 (2000).

33. Goodlett, D.R. *et al*. Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal Chem* **72**, 1112-8 (2000).

34. Smith, R.D. *et al*. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**, 513-23 (2002).

35. Pasa-Tolic, L., Masselon, C., Barry, R.C., Shen, Y. & Smith, R.D. Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* **37**, 621-4, 626-33, 636 passim (2004).

36. Hayashi, S. & Wu, H.C. Accumulation of prolipoprotein in Escherichia coli mutants defective in protein secretion. *J Bacteriol* **161**, 949-54 (1985).

37. Watanabe, T., Hayashi, S. & Wu, H.C. Synthesis and export of the outer membrane lipoprotein in Escherichia coli mutants defective in generalized protein export. *J Bacteriol* **170**, 4001-7 (1988).

38. Froderberg, L., Houben, E.N., Baars, L., Luirink, J. & de Gier, J.W. Targeting and translocation of two lipoproteins in Escherichia coli via the SRP/Sec/YidC pathway. *J Biol Chem* **279**, 31026-32 (2004).

39. Wickner, W. & Schekman, R. Protein translocation across biological membranes. *Science* **310**, 1452-6 (2005).

40. Myers, C.R. & Myers, J.M. Shewanella oneidensis MR-1 restores menaquinone synthesis to a menaquinone-negative mutant. *Appl Environ Microbiol* **70**, 5415-25 (2004).

41. Myers, C.R. & Myers, J.M. MtrB is required for proper incorporation of the cytochromes OmcA and OmcB into the outer membrane of Shewanella putrefaciens MR-1. *Appl Environ Microbiol* **68**, 5585-94 (2002).

42. Swaim, C.L., Smith, J.B. & Smith, D.L. Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *J Am Soc Mass Spectrom* **15**, 736-49 (2004).

43. Kim, Y.J., Rajapandi, T. & Oliver, D. SecA protein is exposed to the periplasmic surface of the E. coli inner membrane in its active state. *Cell* **78**, 845-53 (1994).

44. van der Does, C. *et al*. SecA is an intrinsic subunit of the Escherichia coli preprotein translocase and exposes its carboxyl terminus to the periplasm. *Mol Microbiol* **22**, 619-29 (1996).

45. Elias, D.A. *et al*. Global detection and characterization of hypothetical proteins in Shewanella oneidensis MR-1 using LC-MS based proteomics. *Proteomics* **5**, 3120-30 (2005).

46. Braun, V. & Bosch, V. Repetitive sequences in the murein-lipoprotein of the cell wall of Escherichia coli. *Proc Natl Acad Sci U S A* **69**, 970-4 (1972).

47. Bosch, V. & Braun, V. Distribution of murein-lipoprotein between the cytoplasmic and outer membrane of Escherichia coli. *FEBS Lett* **34**, 307-10 (1973).

48. Shu, W., Liu, J., Ji, H. & Lu, M. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. *J Mol Biol* **299**, 1101-12 (2000).

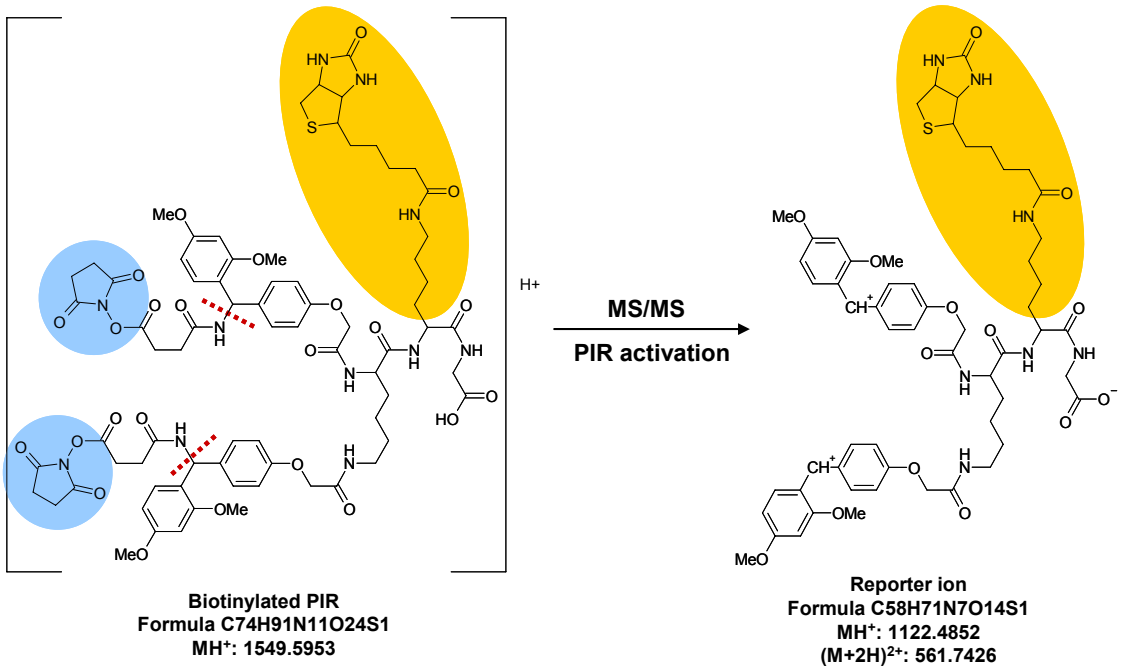



**Figure 1**. Structures of the biotinylated PIR compound and its reporter ion. The MS/MS labile bonds are indicated by the red dashed lines, the reactive groups are NHS-esters highlighted in blue, and the affinity group is biotin highlighted in orange.

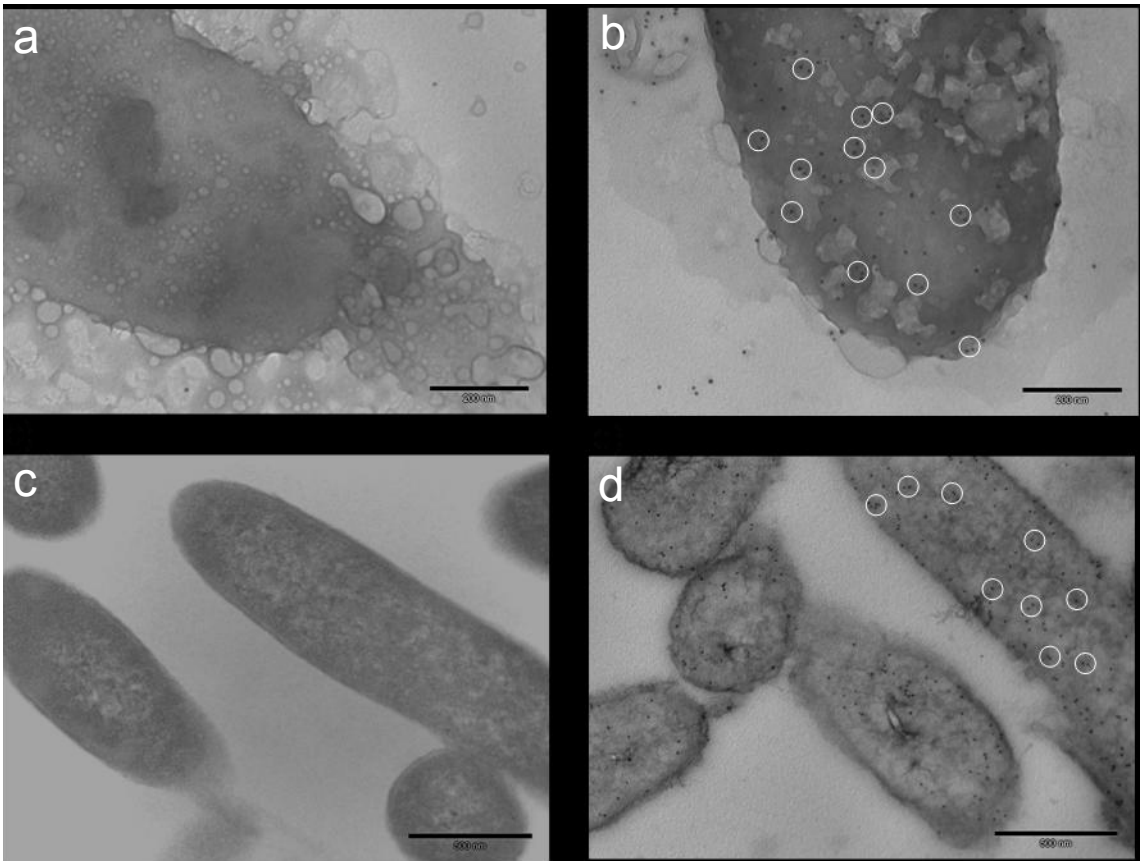



**Figure 2.** Electron microcopy (EM) immunogold detection of PIR-labeled proteins on *S. oneidensis* cells. a) control cell surface; b) PIR-labeled cell surface; c) control cell thin section; and d) PIR-labeled cell thin section. The bar scale in each image is 200 nm. Dark dots indicate the detection of immunogold nanoparticles. To help better visualization, white circles were used to highlight several detected nanoparticles.

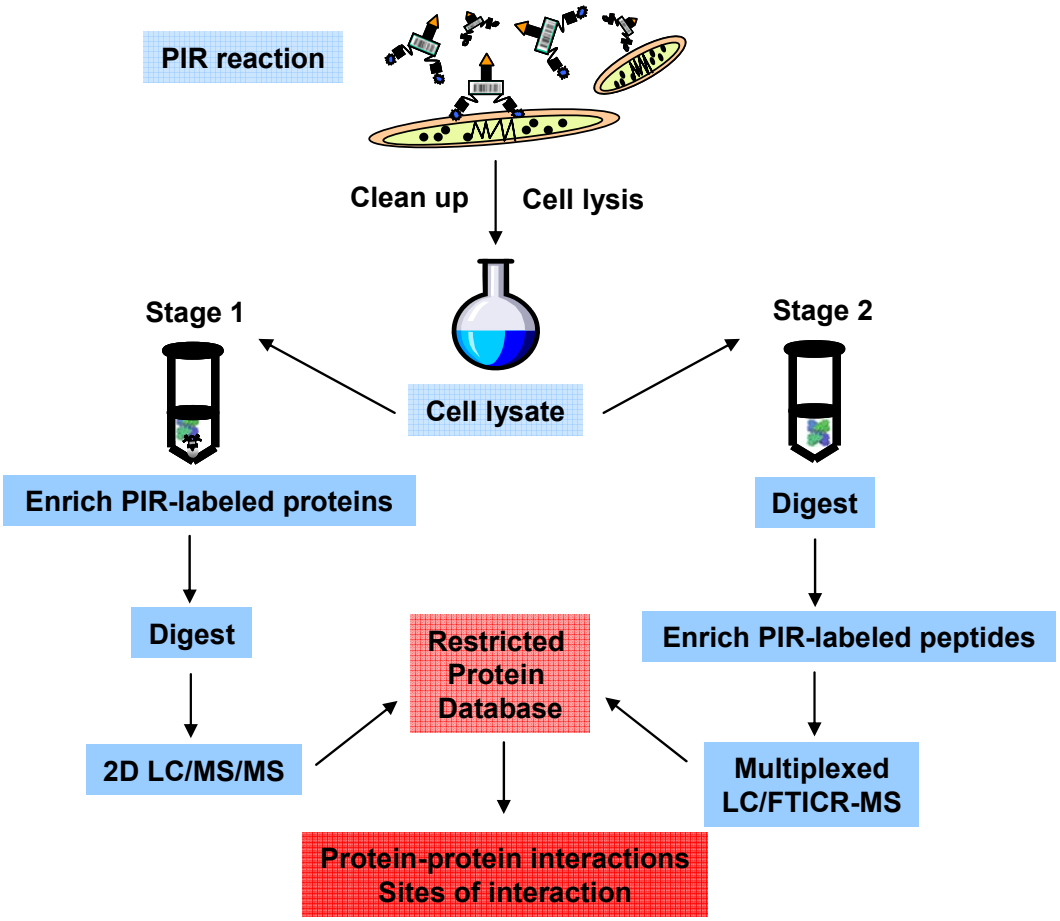



**Figure 3.** Diagram of two-stage mass spectrometric strategy. After intact cell or tissue labeling and protein extraction, proteins are divided into two parts. Stage 1 involves affinity capture of labeled proteins, followed by digestion and shotgun LC/MS/MS for protein identification to constitute a restricted protein database. Stage 2 is performed by digesting proteins first, followed by affinity enrichment of PIR labeled peptides. Analysis of the labeled peptides is carried out with multiplexed LC/FTICR-MS. Accurate masses of the labeled peptides are measured and used for protein identification by searching against the restricted database compiled from stage 1.

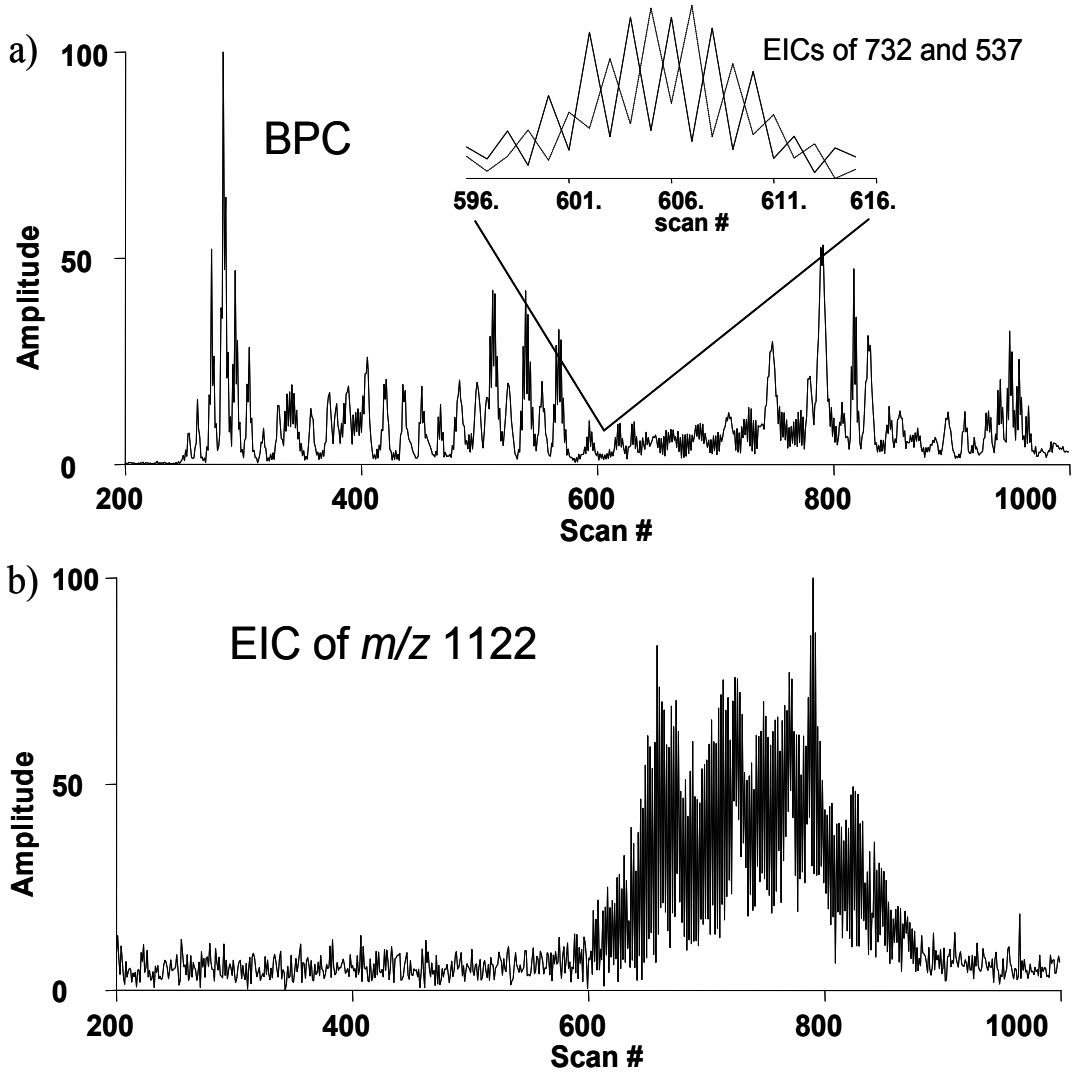



**Figure 4.** Multiplexed LC/FTICR-MS of PIR-labeled peptides with alternating low-energy (-4 v) and PIR activation energy (-18 v) applied in the collision cell. a) Base peak chromatogram (BPC). b) Extracted ion chromatogram (EIC) of singly-charged reporter ion at *m/z* 1122. The inset in a) is the overlaid EICs of a PIR-labeled precursor ion at *m/z* $732^{3+}$ (dashed-line trace) and its product ion at *m/z* 537 (solid-line trace).

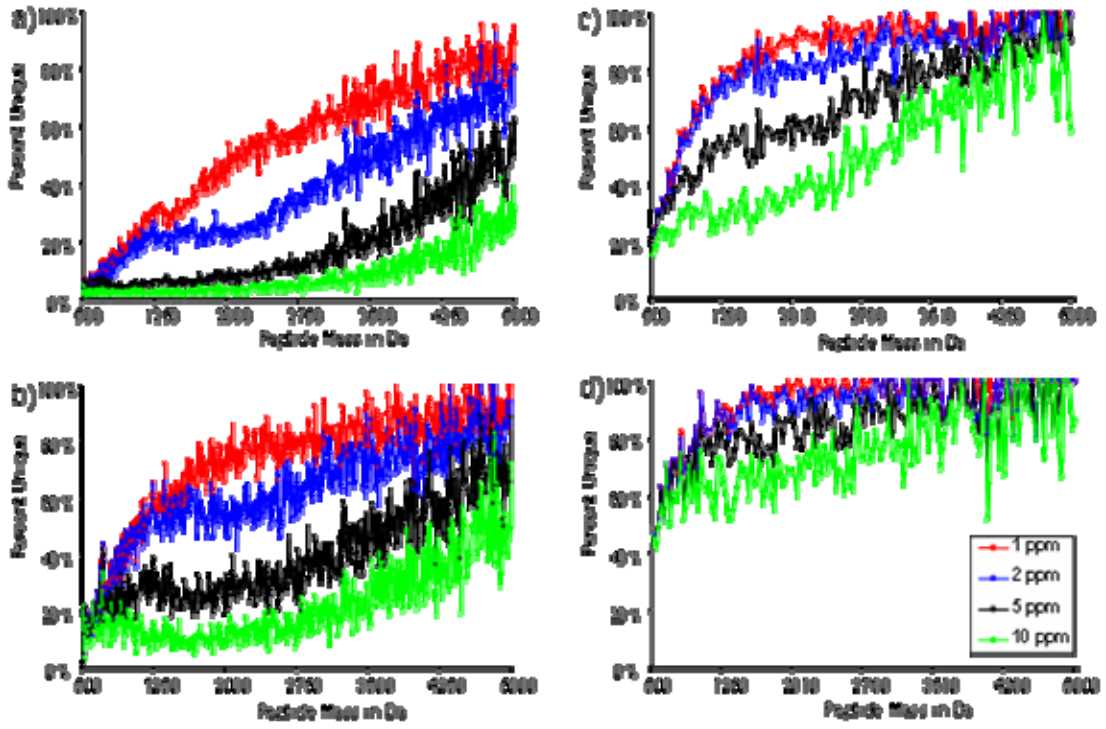



**Figure 5**. Percent unique of tryptic peptides (up to one missed cleavage point) as a function of peptide masses at various levels of mass measurement accuracy (red 1 ppm, blue 2 ppm, black 5 ppm, and green 10 ppm). a) all ORFs (4,854) of *S. oneidensis* MR-1; b) all ORFs of *S. oneidensis* MR-1 with an internal lysine constraint; c) the restricted database (364 proteins) built from stage 1 analysis; and d) the restricted database with an internal lysine constraint.

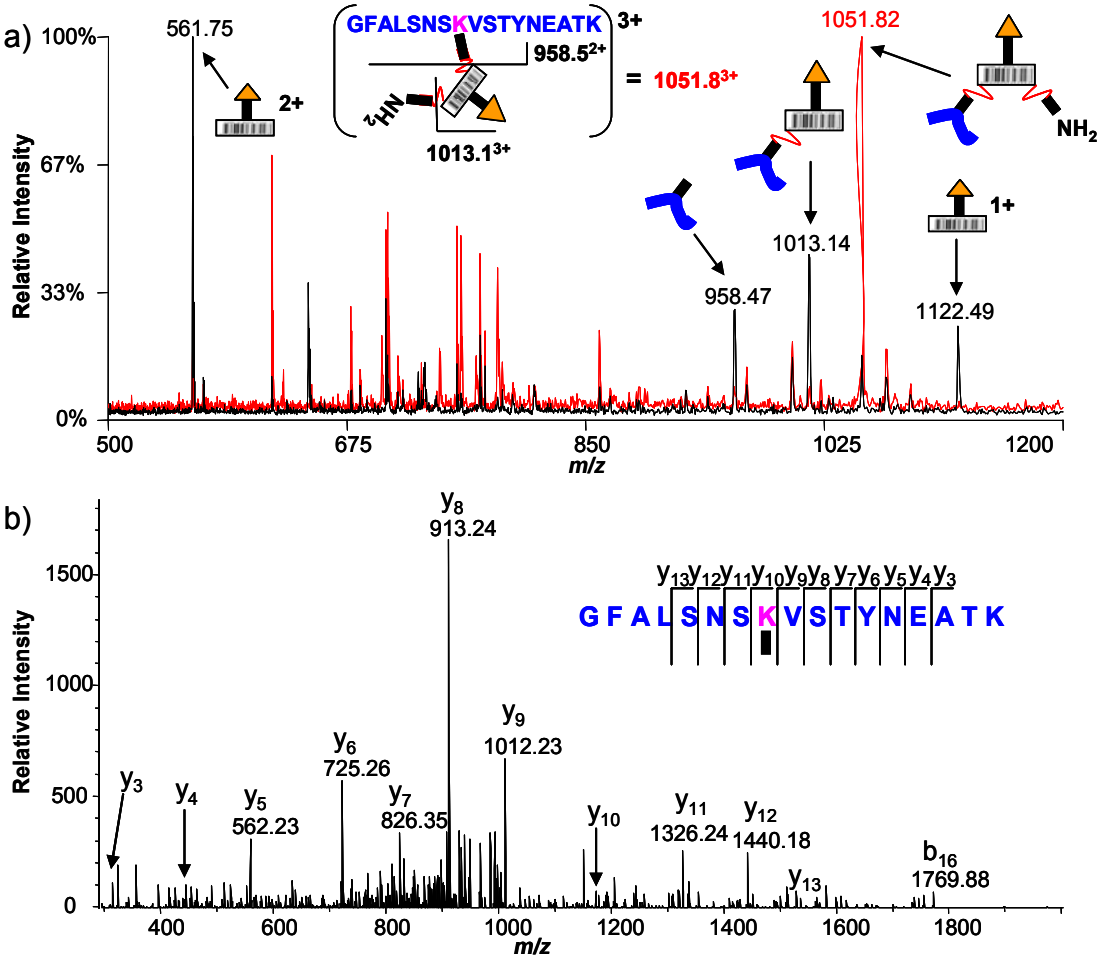



**Figure 6.** Detection of a PIR dead-end labeled peptide from *S. oneidensis*. a) Multiplexed LC/FTICR-MS spectra of a PIR dead-end labeled peptide. Shown are two superimposed spectra from consecutive scans. Red trace is the low-energy scan and black trace is the following PIR activation scan. The precursor ion and its major product ions are assigned as indicated by cartoon symbols. The precursor ion at *m/z* $1051.82^{3+}$ is fragmented to give the reporter ion at *m/z* $561.75^{2+}$ and 1122.49, intact peptide ions at *m/z* $958.47^{2+}$, and a dead-end residual tag at *m/z* 117.06 (not detected in the mass range set for the experiment). b) $MS^3$ sequencing of the released intact peptide at nominal *m/z* $958^{2+}$ by the ion trap mass spectrometer.

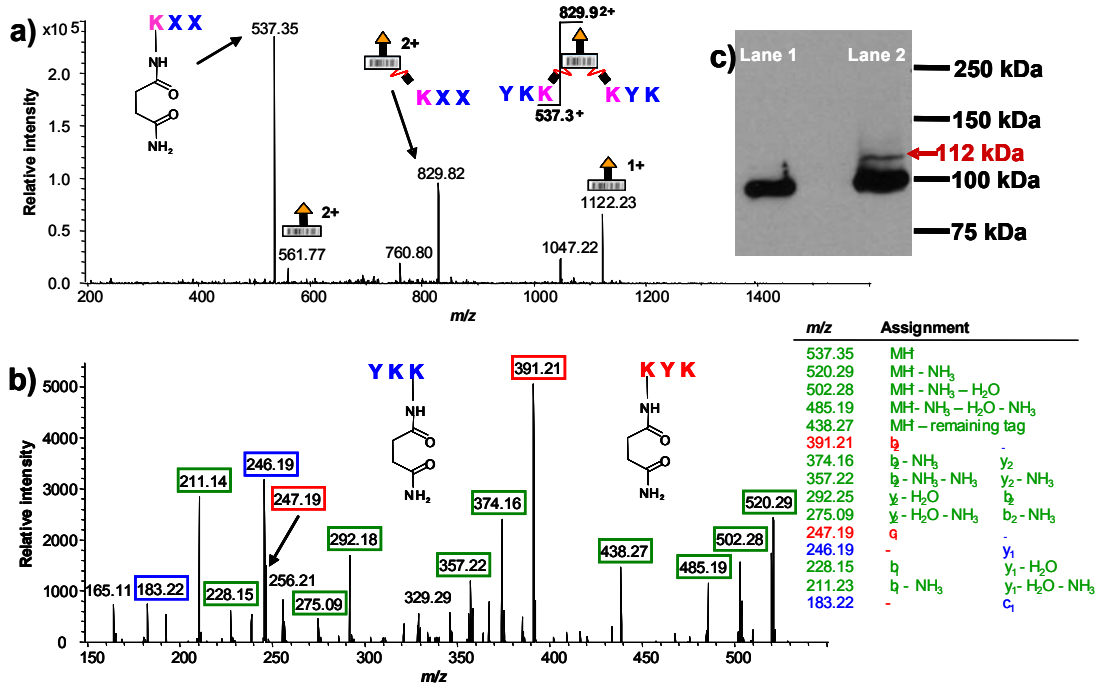
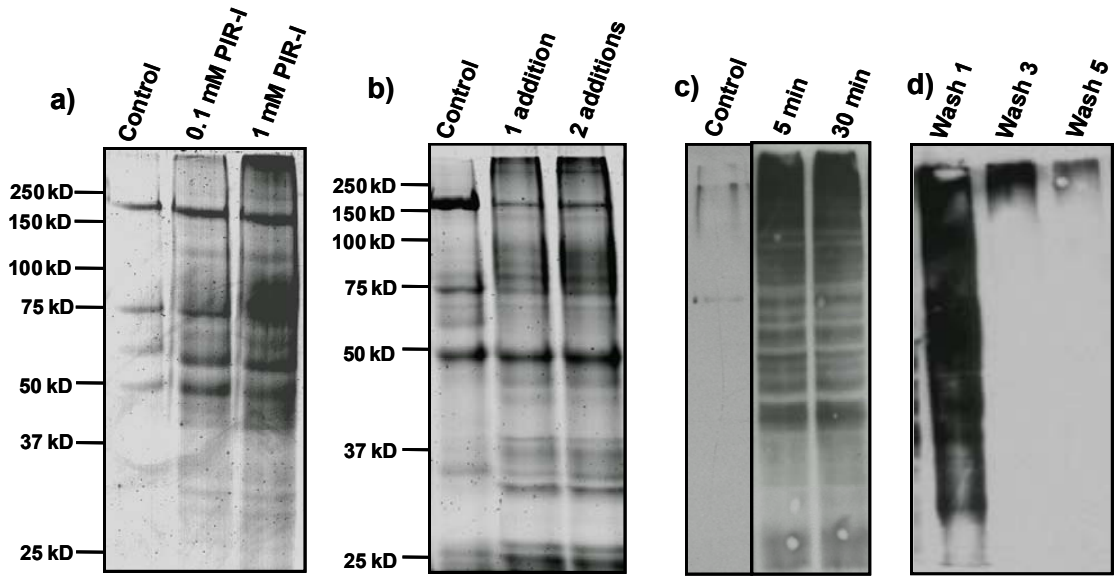



**Figure 7.** Detection of a PIR inter-cross-linked peptide complex from *S. oneidensis*. a) MS/MS spectrum of the isolated precursor ion at nominal *m/z* $732^{3+}$. All major peaks are assigned as indicated by cartoon symbols. The structure of the remaining tag on the peptide is fully drawn. KXX indicates that the peaks could be assigned for both sequence KYK and YKK. *m/z* $760.80^{2+}$ is produced from the loss of a dimethoxyl phenyl group in the reporter ion and *m/z* $1047.22^{2+}$ is produced from the loss of a glycine group in the reporter ion. b) $MS^3$ spectrum of the released peptide chain at nominal *m/z* 537 and assignment table. Ions that can be assigned only to the sequence KYK are boxed in red, those assigned to YKK sequence are in blue, and those assigned to both sequences are in green. c) Anti-SecA Western blot analysis of unlabeled and PIR-labeled *S. oneidensis* cell lysate. Lane 1: negative control (unlabeled) cell lysate. Lane 2: PIR-labeled cell lysate. A PIR inter-cross-linked SecA product is indicated by the band with MW 112 kDa.

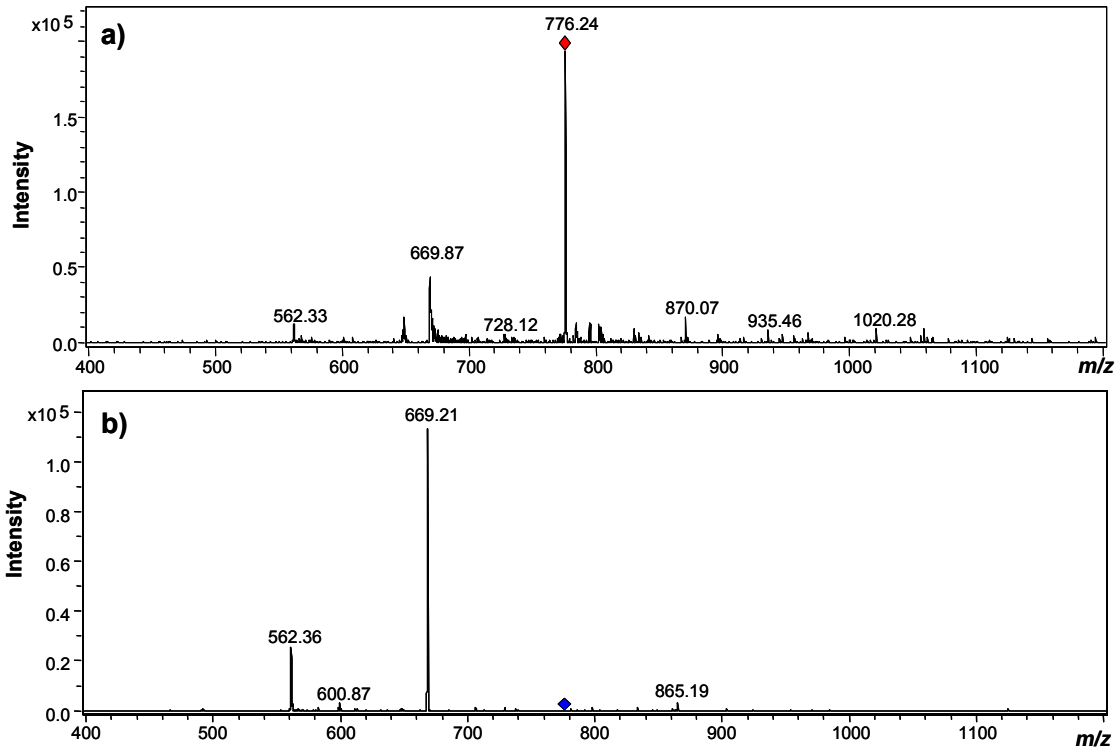
**Table 1.** A summary of identified PIR-labeled peptides and proteins from stage 2 data

| Measured mass | Type[a] | Identified peptide | Error (ppm) | Protein of origin | MW (kDa) |
|---|---|---|---|---|---|
| 503.2826 | DE | SN**K**R | 2.1 | ribosomal protein L2 (RplB) | 30 |
| 1274.6384 | DE | HVGGGH**K**QHYR | 0.3 | | |
| 1373.7815 | DE | HVV**K**VVNSDLHK | 2.7 | | |
| 504.2849 | DE | AM**K**R | 1.4 | ribosomal protein S3 (RpsC) | 26 |
| 1574.9626 | DE | VTIHTARPGVVIG**K**K | 0.4 | | |
| 1922.1023 | DE | AAVS**K**LAGTPAQINIAEIR | 3.9 | | |
| 1815.9068 | DE | GFALSNS**K**VSTYNEATK | 3.7 | decaheme cytochrome c (OmcA) | 79 |
| 1911.9325 | DE | TDNTYPGT**K**VPTSFAWK | 2.2 | | |
| 524.3454 | DE | LH**K**K | 3.9 | decaheme cytochrome c (OmcB) | 71 |
| 2734.4550 | DE | LEIITNVGPNNATLGYSG**K**D-SIFAIK | 0.4 | | |
| 2222.1301 | DE | SQINP**K**FQDLLTADVSFSGR | 1.3 | hypothetical protein (SO0404) | 129 |
| 1091.5402 | DE | M**K**RDTENAK | 0.9 | disulfide interchange protein (DsbA) | 22 |
| 1832.0308 | DE | SAP**K**AAASAAPVVQPLAAGR | 2.2 | dihydrolipoamide succinyltransferase | 43 |
| 1474.7961 | DE | D**K**NPVVFFEP**K**R | 2.0 | (SucB) | 35 |
| 1175.6121 | IA | TDEE**K**SI**K**AR | 2.1 | alpha keto acid dehydrogenase, β subunit | 80 |
| 1083.7086 | IA | II**KK**LSGGIR | 3.9 | isocitrate dehydrogenase (Icd) | 55 |
| 437.2634 | IR | **K**YK | 0.9 | ATP synthase F1, α subunit (AtpA) | 103 |
| | | | | preprotein translocase, SecA subunit | 43 |
| | | | | (SecA) | 46 |
| | | Y**KK** | | acyl-carrier-protein synthase I (FabB) | 9.2 |
| | | | | TPR domain protein | |
| | | | | major outer membrane lipoprotein, putative | |

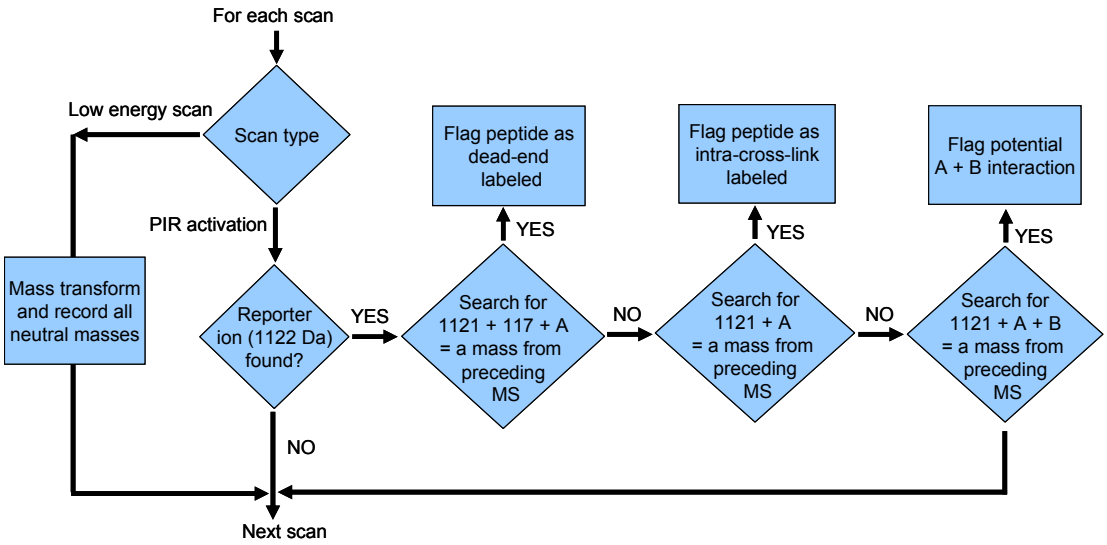
[a]Type of identified cross-linked peptides; DE, IA, and IR denotes dead-end labeled peptide, intra-cross-linked peptide, and inter-cross-linked peptides, respectively.



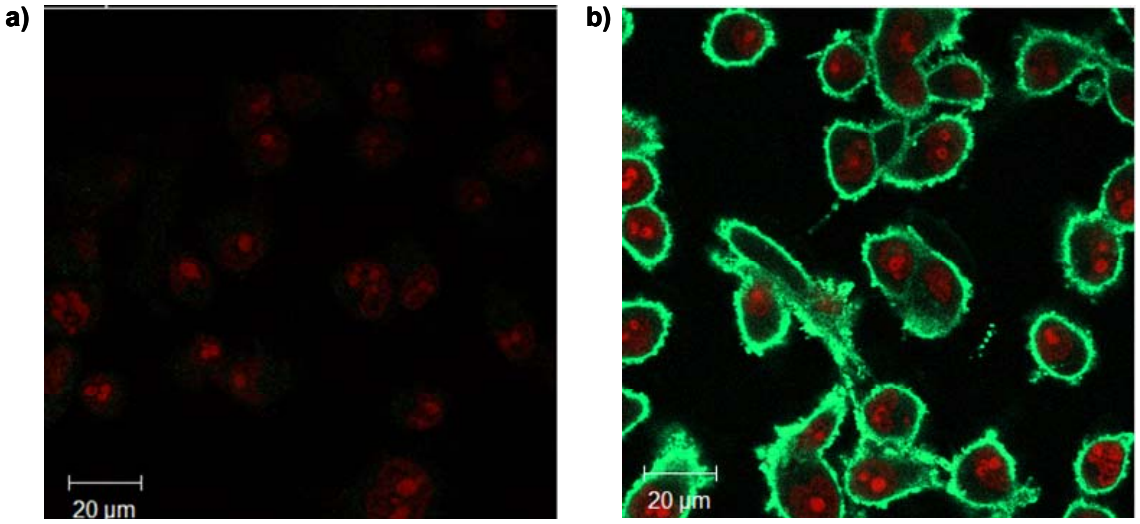

**Supplementary Figure 1.** Optimization of cell labeling experiments with PIR. a) SDS-PAGE of avidin-enriched proteins from control cell lysates (without PIR labeling) and labeled cell lysate using 0.1 mM and 1 mM PIR. b) SDS-PAGE of avidin-enriched proteins from control cell lysates and labeled cell lysates prepared by adding 10 μL, 100 mM PIR to 1 mL cell suspension once and twice. c) Anti-biotin Western blot image of control cell lysate and PIR labeled cell lysate at two reaction times: 5 min and 30 min. d) Anti-biotin Western blot image of washing solutions after cell labeling with PIR.



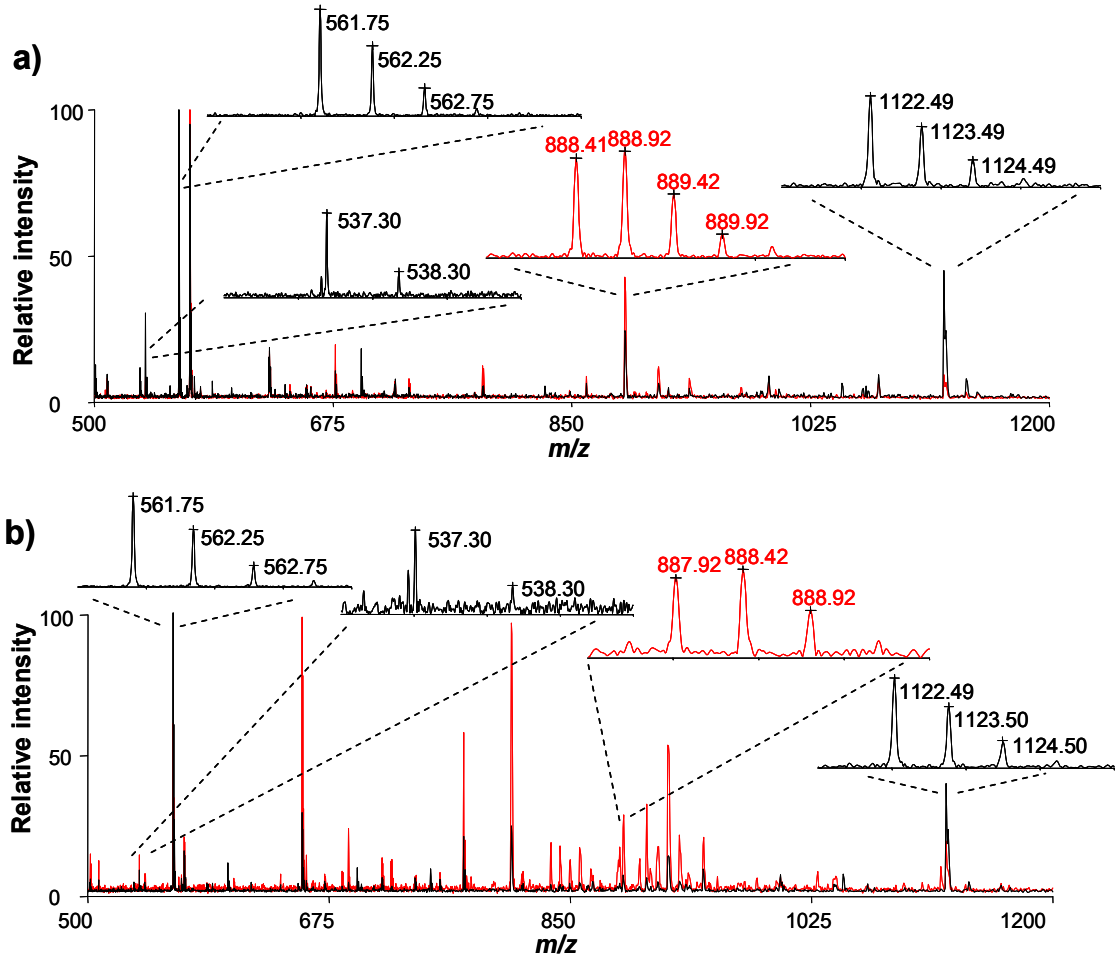

**Supplementary Figure 2**. MS a) and MS/MS b) of PIR compound used for present study. *m/z* 776 corresponds to the expected *m/z* of the $(M+2H)^{2+}$ ions of the PIR. Isolation and activation of these ions results in fragmentation at one and both Rink groups to produce doubly charged ions at *m/z* 669 and 562, respectively.



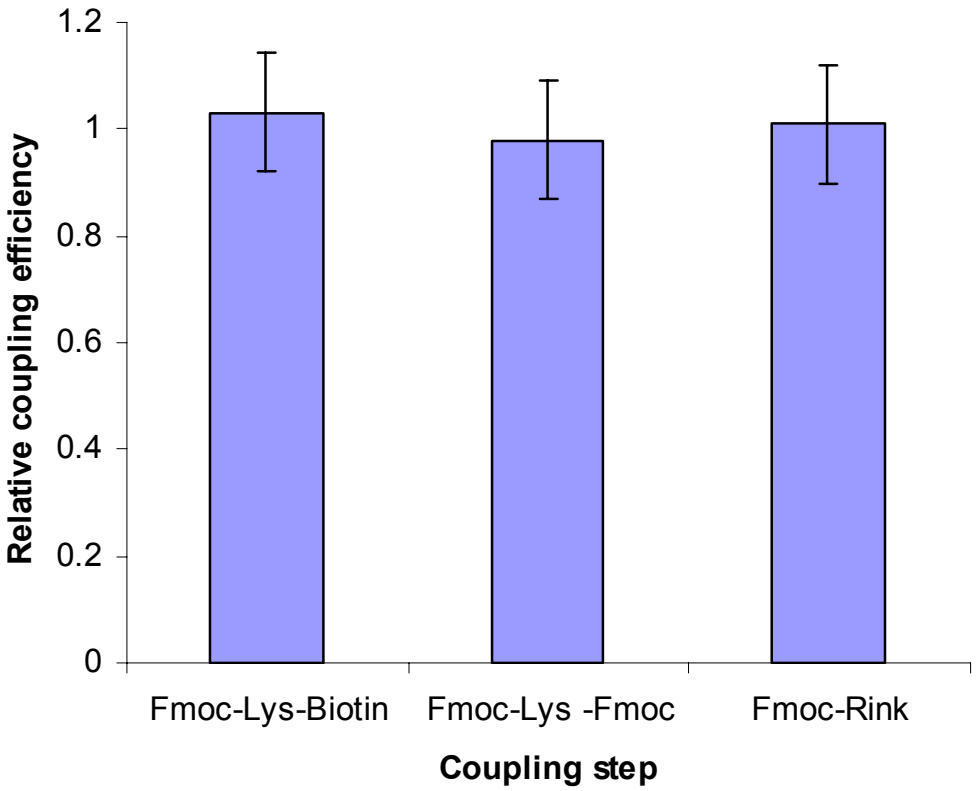

**Supplementary Figure 3.** Schematic diagram of algorithm used by *X-links* to identify PIR-labeled peptides. 1122 is nominal *m/z* of the singly-charged PIR reporter ion. "A" and "B" are masses observed in the PIR activation scan. 117 is the nominal neutral mass of the residual dead-end tag for the current PIR compound.



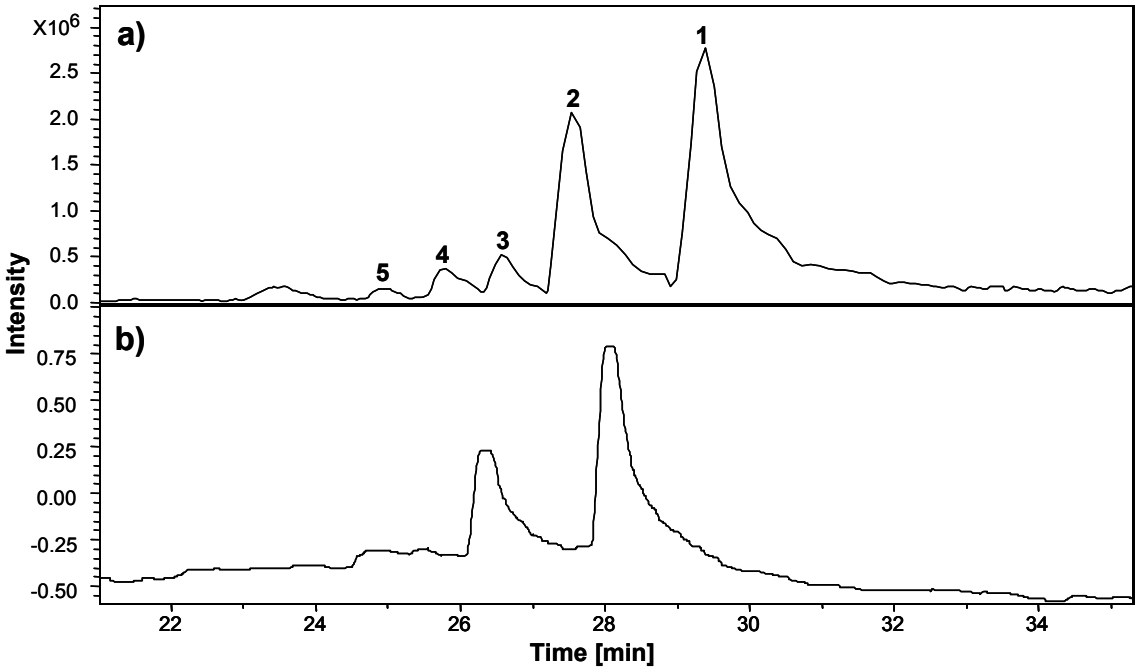

**Supplementary Figure 4.** Confocal fluorescence images of HeLa cells without a) and with b) PIR labeling. Both images are the superposition of fluorescent images acquired from anti-biotin and Alexa Fluor 488 conjugated IgG (green) and nuclear staining with propidium iodide (red). Control cells with no PIR treatment showed no antibiotin response at this level, while treated cells showed significant PIR incorporation of PIR molecules in the cell membranes.



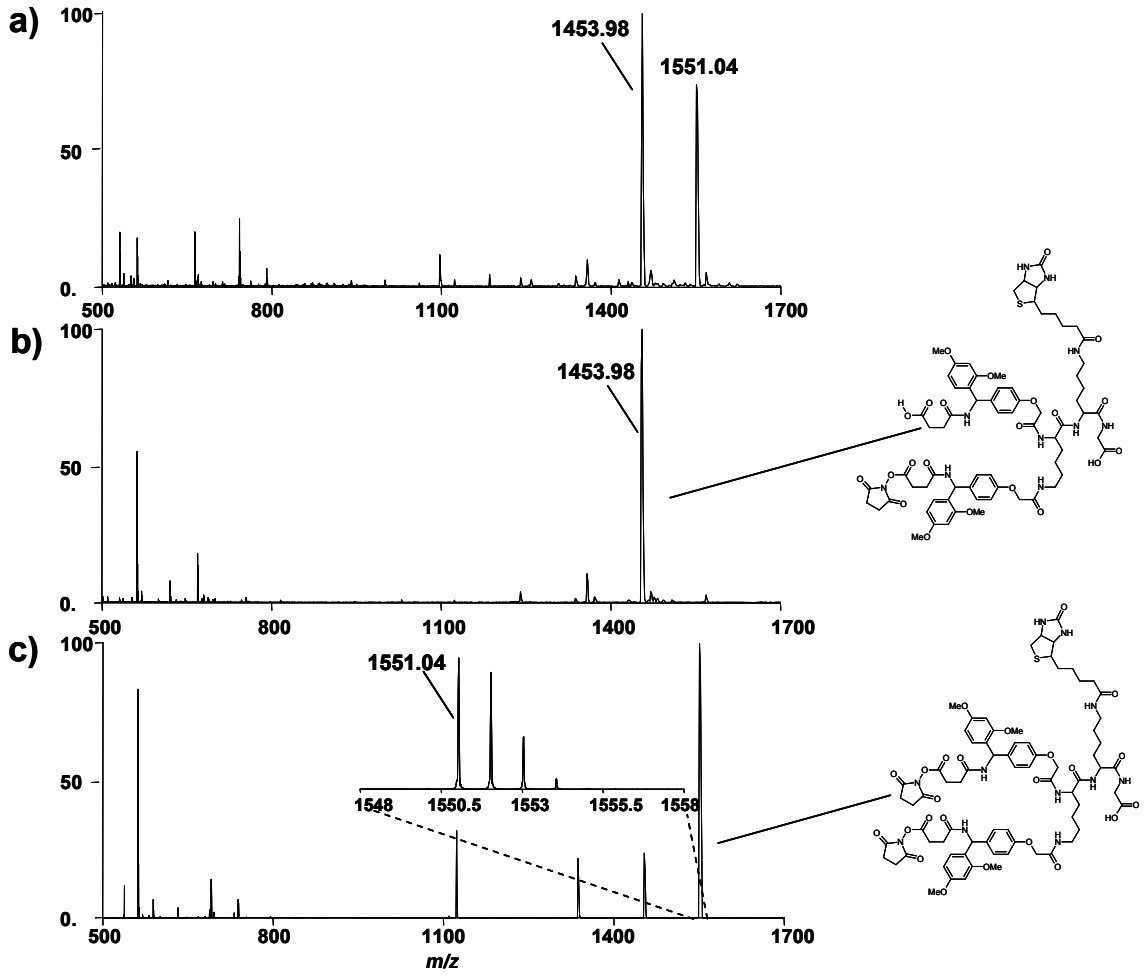

**Supplementary Figure 5.** Multiplexed LC/FTICR-MS spectra of PIR dead-end labeled peptide KYK and YKK in a) water-displacement version (*m/z* $888.4^{2+}$) and b) ammonia-displacement version (*m/z* $887.9^{2+}$). Shown are two superimposed spectra from consecutive scans. Red trace is the MS scan and black trace is the following MS/MS scan. The precursor ion at *m/z* $888.4^{2+}$ or $887.9^{2+}$ was fragmented to generate reporter ion at *m/z* 1122.5 and $561.7^{2+}$ and the intact peptide ion at *m/z* 537.3.



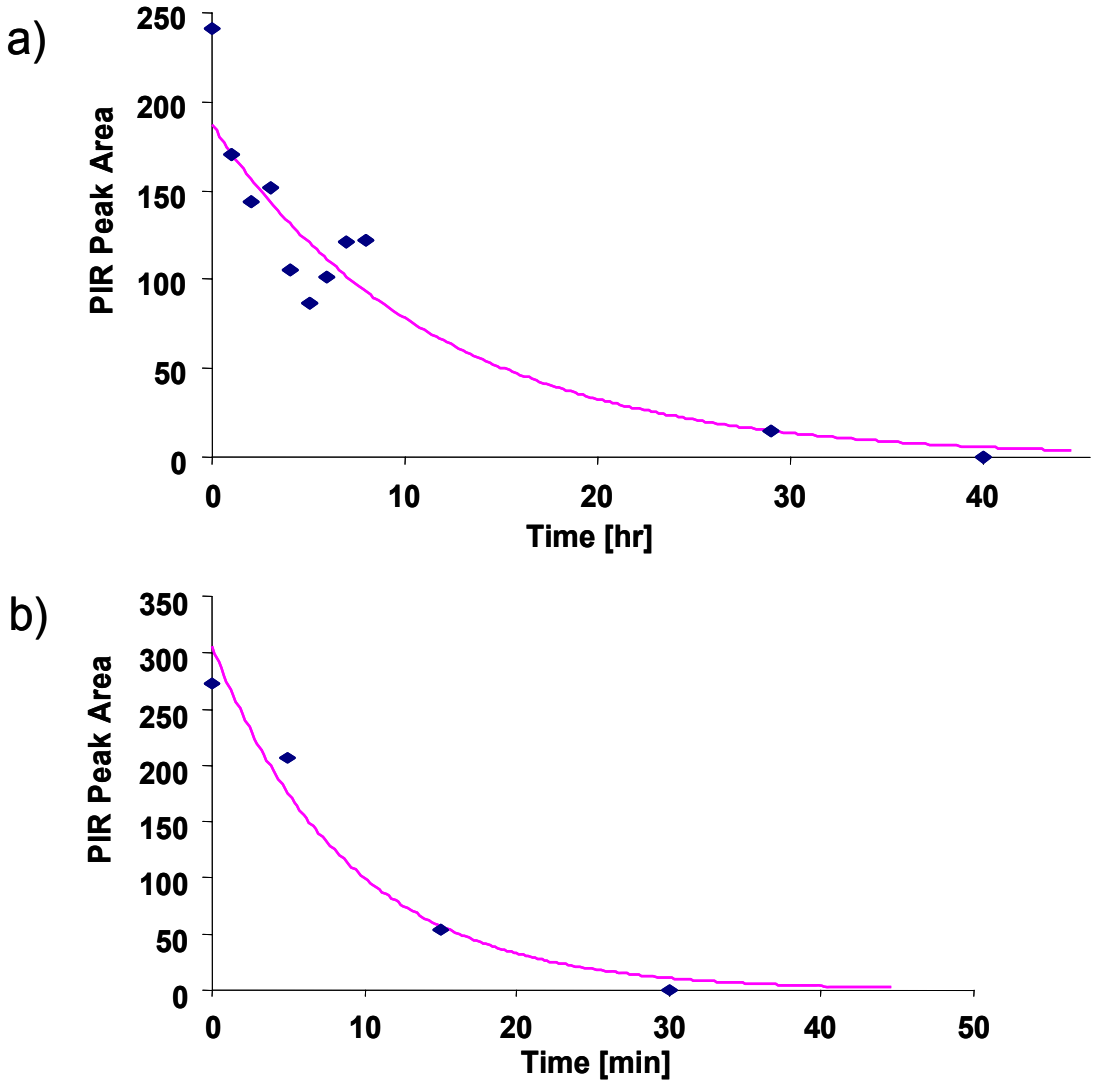

**Supplementary Figure 6.** Measured relative coupling efficiency during PIR synthesis as determined by UV absorbance of released Fmoc at 301 nM. These data are the average of 12 independent synthesis attempts. Error bars indicate +/- one standard deviation of observed efficiency following coupling and deprotection of the Fmoc-Lys-Biotin amino acid. This step resulted in the greatest variability in relative coupling efficiency with a standard deviation of 0.11. The other two steps resulted in standard deviation of 0.05 and 0.07 (Fmoc-Lys-Fmoc and Fmoc-Rink, respectively). Overall, coupling of Fmoc-protected amino acids was achieved with high efficiency. As no Fmoc deprotection step was used in either the succinic acid or NHS ester addition, such monitoring for these synthesis steps was not possible.



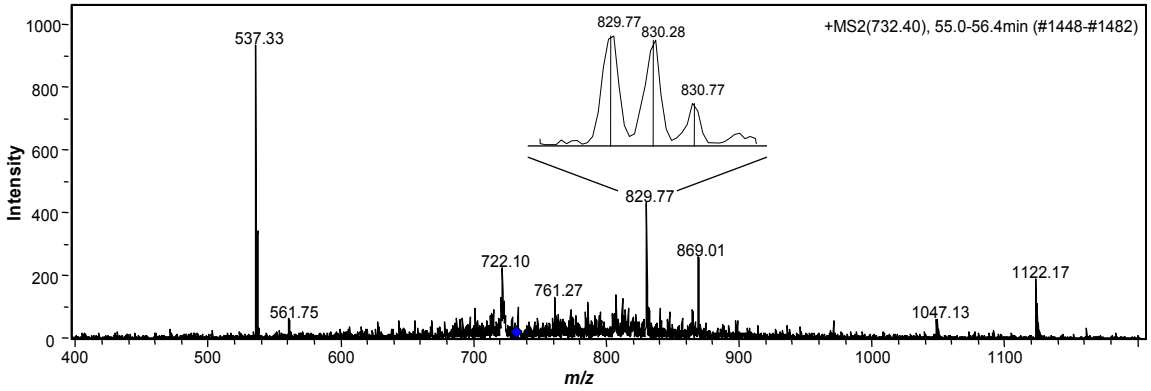

**Supplementary Figure 7.** LC/MS analysis of crude PIR product after cleavage, lyophilization, and re-suspension in DMSO. Solution was diluted in 0.1% TFA prior to injection on column. a) Base peak chromatogram. b) UV absorbance monitored at 280 nm. *1* = expected product with 2 NHS esters, *2* = compound with 1 NHS and 1 OH, *3* = compound with 1 NHS and 1 NH2, *4* = compound with 2 OH groups, *5* = compound with 1 OH and 1 NH2. Note, the delay between the appearance of peaks in the UV absorbance chromatogram and those in the mass spec data (base peak chromatogram) are resultant from the delay that exists between the arrival of sample at the UV detector and the arrival of sample at the ESI tip and mass spectrometer.




**Supplementary Figure 8.** ESI-FTICR mass spectra of a) unpurified PIR, b) collected peak 2 from *Supplementary Figure 7a*, and c) collected peak 1 from *Supplementary Figure 7a*. *m/z* = 1551.04 = $(M+H)^+$ ions of the expected product. *m/z* = 1453.98 = $(M+H)^+$ ions of the compound with only one NHS ester.




**Supplementary Figure 9.** Measured hydrolysis of PIR in 0.1% formic acid solution (pH 3.0) a) and in PBS buffer (pH 7.2) b). Half life for hydrolysis at this pH 3.0 and pH 7.2 is approximately 8 hr and 10 min, respectively.




**Supplementary Figure 10.** LC/MS/MS results from on-cell labeling of *S. oneidensis* carried out with LC-purified PIR. Shown is the MS/MS spectrum of the precursor ion at *m/z* $732^{3+}$ to yield expected reporter ions at *m/z* $561^{2+}$ and 1122, and product ions at *m/z* $830^{2+}$ and 537, demonstrating that the same interlinked product (KYK-YKK) was observed from reactions with purified PIR compound. Inset figure shows resolved charge-state for doubly charged ion at *m/z* 830.

**Supplementary Table 1.** Proteins that were identified from stage 1 mass spectrometric analysis.

| Accession # | Identified proteins | MW (Da) |
|---|---|---|
| SO0011 | DNA gyrase, B subunit (gyrB) | 89931 |
| SO0014 | glycyl-tRNA synthetase, beta subunit (glyS) | 74889 |
| SO0021 | fatty oxidation complex, alpha subunit (fadB) | 76644 |
| SO0052 | protein-export protein SecB (secB) | 17086 |
| SO0097 | urocanate hydratase (hutU) | 60353 |
| SO0098 | histidine ammonia-lyase (hutH) | 54397 |
| SO0102 | formate dehydrogenase, nitrate-inducible, iron-sulfur subunit (fdnH) | 33108 |
| SO0139 | ferritin (ftn) | 20305 |
| SO0152 | conserved hypothetical protein | 69732 |
| SO0161 | hypothetical protein | 3903 |
| SO0162 | phosphoenolpyruvate carboxykinase (ATP) (pckA) | 55848 |
| SO0166 | general secretion pathway protein D (gspD) | 76812 |
| SO0167 | general secretion pathway protein E (gspE) | 57624 |
| SO0217 | translation elongation factor Tu (tufB) | 43320 |
| SO0219 | transcription antitermination protein NusG (nusG) | 20890 |
| SO0220 | ribosomal protein L11 (rplK) | 15098 |
| SO0221 | ribosomal protein L1 (rplA) | 24585 |
| SO0222 | ribosomal protein L10 (rplJ) | 17732 |
| SO0223 | ribosomal protein L7/L12 (rplL) | 12509 |
| SO0224 | DNA-directed RNA polymerase, beta subunit (rpoB) | 150201 |
| SO0225 | DNA-directed RNA polymerase, beta subunit (rpoC) | 155422 |
| SO0227 | ribosomal protein S7 (rpsG) | 17720 |
| SO0228 | translation elongation factor G (fusA-1) | 76901 |
| SO0229 | translation elongation factor Tu (tufA) | 43257 |
| SO0230 | ribosomal protein S10 (rpsJ) | 11761 |
| SO0231 | ribosomal protein L3 (rplC) | 22518 |
| SO0232 | ribosomal protein L4 (rplD) | 21968 |
| SO0233 | ribosomal protein L23 (rplW) | 11061 |
| SO0234 | ribosomal protein L2 (rplB) | 29945 |
| SO0235 | ribosomal protein S19 (rpsS) | 10466 |
| SO0236 | ribosomal protein L22 (rplV) | 12064 |
| SO0237 | ribosomal protein S3 (rpsC) | 25739 |
| SO0238 | ribosomal protein L16 (rplP) | 15386 |
| SO0239 | ribosomal protein L29 (rpmC) | 7166 |
| SO0241 | ribosomal protein L14 (rplN) | 13447 |
| SO0242 | ribosomal protein L24 (rplX) | 11311 |
| SO0243 | ribosomal protein L5 (rplE) | 20219 |

| | | |
|---|---|---|
| SO0244 | ribosomal protein S14 (rpsN) | 11355 |
| SO0245 | ribosomal protein S8 (rpsH) | 14030 |
| SO0246 | ribosomal protein L6 (rplF) | 18911 |
| SO0247 | ribosomal protein L18 (rplR) | 12715 |
| SO0248 | ribosomal protein S5 (rpsE) | 17729 |
| SO0250 | ribosomal protein L15 (rplO) | 15071 |
| SO0253 | ribosomal protein S13 (rpsM) | 13312 |
| SO0255 | ribosomal protein S4 (rpsD) | 23422 |
| SO0256 | DNA-directed RNA polymerase, alpha subunit (rpoA) | 36138 |
| SO0265 | conserved hypothetical protein | 45233 |
| SO0266 | cytochrome c-type biogenesis protein CcmF (ccmF-1) | 72209 |
| SO0267 | thiol:disulfide interchange protein DsbE (dsbE) | 21071 |
| SO0272 | competence/damage-inducible protein CinA (cinA) | 46153 |
| SO0274 | phosphoenolpyruvate carboxylase (ppc) | 99480 |
| SO0285 | type IV pilus biogenesis protein PilQ | 74812 |
| SO0292 | ribulose-phosphate 3-epimerase (rpe) | 24383 |
| SO0314 | ornithine decarboxylase, inducible (speF) | 81049 |
| SO0333 | thiol:disulfide interchange protein DsbA (dsbA) | 22289 |
| SO0335 | conserved hypothetical protein | 8396 |
| SO0358 | endoribonuclease L-PSP, putative | 13700 |
| SO0362 | hypothetical protein | 56090 |
| SO0404 | hypothetical protein | 128699 |
| SO0405 | transcription termination factor Rho (rho) | 47132 |
| SO0406 | thioredoxin 1 (trxA) | 11881 |
| SO0407 | ATP-dependent RNA helicase, DEAD box family | 49232 |
| SO0424 | pyruvate dehydrogenase complex, E1 component, pyruvate dehydrogenase (aceE) | 99325 |
| SO0425 | pyruvate dehydrogenase complex, E2 component, dihydrolipoamide acetyltransferase (aceF) | 69356 |
| SO0426 | pyruvate dehydrogenase complex, E3 component, lipoamide dehydrogenase (lpdA) | 50504 |
| SO0432 | aconitate hydratase 2 (acnB) | 93334 |
| SO0435 | uroporphyrinogen decarboxylase (hemE) | 39211 |
| SO0441 | phosphoribosylamine--glycine ligase (purD) | 45510 |
| SO0467 | DNA helicase II (uvrD) | 81639 |
| SO0483 | formate-dependent nitrite reductase, nrfC protein (nrfC) | 25378 |
| SO0575 | RNA polymerase-associated protein HepA (hepA) | 109194 |
| SO0603 | host factor-I protein (hfq) | 9867 |
| SO0605 | hflK protein (hflK) | 42627 |
| SO0606 | hflC protein (hflC) | 33184 |
| SO0607 | hypothetical protein | 4003 |
| SO0608 | ubiquinol-cytochrome c reductase, iron-sulfur subunit (petA) | 20892 |
| SO0611 | stringent starvation protein a (sspA) | 24065 |
| SO0617 | acetylornithine aminotransferase (argD) | 43187 |

| | | |
|---|---|---|
| SO0624 | catabolite gene activator (crp) | 23691 |
| SO0633 | penicillin-binding protein 1B (mrcB) | 84745 |
| SO0703 | chaperonin GroES (groES) | 10207 |
| SO0704 | chaperonin GroEL (groEL) | 57044 |
| SO0719 | TonB-dependent receptor, putative | 84077 |
| SO0770 | malate dehydrogenase (mdh) | 32117 |
| SO0781 | glycine cleavage system P protein (gcvP) | 104607 |
| SO0783 | hypothetical protein | 76497 |
| SO0807 | hypoxanthine-guanine phosphoribosyltransferase (hpt-1) | 21307 |
| SO0830 | alkaline phosphatase | 48515 |
| SO0840 | acetyl-CoA carboxylase multifunctional enzyme accADC, carboxyl transferase subunit alpha/carboxyl transferase subunit beta/biotin carboxylase | 167518 |
| SO0842 | translation elongation factor G (fusA-2) | 76614 |
| SO0848 | periplasmic nitrate reductase (napA) | 92415 |
| SO0862 | D-3-phosphoglycerate dehydrogenase (serA) | 44286 |
| SO0876 | peptidase B (pepB) | 45192 |
| SO0929 | S-adenosylmethionine synthetase (metK) | 41354 |
| SO0930 | transketolase (tkt) | 71570 |
| SO0932 | phosphoglycerate kinase (pgk) | 40573 |
| SO0933 | fructose-bisphosphate aldolase, class II, Calvin cycle subtype (fba) | 38486 |
| SO0970 | fumarate reductase flavoprotein subunit precursor | 62409 |
| SO0992 | lysyl-tRNA synthetase (lysS) | 57024 |
| SO1016 | NADH dehydrogenase I, G subunit (nuoG) | 100016 |
| SO1065 | FKBP-type peptidyl-prolyl cis-trans isomerase FkpA (fkpA) | 27684 |
| SO1066 | extracellular nuclease | 93635 |
| SO1075 | conserved hypothetical protein | 80624 |
| SO1093 | ISSo7, transposase | 50194 |
| SO1103 | NADH:ubiquinone oxidoreductase, Na translocating, alpha subunit (nqrA-2) | 47771 |
| SO1108 | NADH:ubiquinone oxidoreductase, Na translocating, beta subunit (nqrF-2) | 46060 |
| SO1111 | bacterioferritin subunit 2 (bfr2) | 18373 |
| SO1126 | chaperone protein DnaK (dnaK) | 68819 |
| SO1142 | carbamoyl-phosphate synthase, large subunit (carB) | 117907 |
| SO1149 | conserved hypothetical protein | 44074 |
| SO1164 | D-alanyl-D-alanine carboxypeptidase (dacA-1) | 43082 |
| SO1185 | conserved hypothetical protein TIGR00092 | 39432 |
| SO1197 | cell division protein FtsH (ftsH) | 71182 |
| SO1198 | dihydropteroate synthase (folP) | 29929 |
| SO1201 | preprotein translocase, SecG subunit (secG) | 11554 |
| SO1203 | N utilization substance protein A (nusA) | 55239 |
| SO1204 | translation initiation factor IF-2 (infB) | 96054 |

| | | |
|---|---|---|
| SO1205 | ribosome-binding factor A (rbfA) | 16755 |
| SO1209 | polyribonucleotide nucleotidyltransferase (pnp) | 75784 |
| SO1210 | TPR domain protein | 32135 |
| SO1215 | outer membrane protein OmpK, putative | 31430 |
| SO1217 | deoxyribose-phosphate aldolase (deoC) | 27428 |
| SO1221 | purine nucleoside phosphorylase (deoD-2) | 25586 |
| SO1284 | RNA polymerase sigma-70 factor (rpoD) | 70348 |
| SO1295 | major outer membrane lipoprotein, putative | 9178 |
| SO1300 | glutamate-1-semialdehyde-2,1-aminomutase (hemL) | 46103 |
| SO1315 | tyrosyl-tRNA synthetase (tyrS) | 44000 |
| SO1327 | sensor histidine kinase-related protein | 13238 |
| SO1346 | GTP-binding protein LepA (lepA) | 65930 |
| SO1357 | ribosomal protein S16 (rpsP) | 9103 |
| SO1360 | ribosomal protein L19 (rplS) | 13284 |
| SO1361 | phospho-2-dehydro-3-deoxyheptonate aldolase, tyr-sensitive (aroF) | 39849 |
| SO1367 | chorismate mutase/prephenate dehydratase (pheA) | 74043 |
| SO1398 | hypothetical protein | 3363 |
| SO1429 | anaerobic dimethyl sulfoxide reductase, A subunit (dmaA-1) | 91061 |
| SO1434 | methyl-accepting chemotaxis protein | 59310 |
| SO1480 | GGDEF family protein | 66626 |
| SO1490 | alcohol dehydrogenase II (adhB) | 39988 |
| SO1518 | conserved hypothetical protein | 20059 |
| SO1519 | iron-sulfur cluster-binding protein | 51399 |
| SO1520 | conserved hypothetical protein | 26612 |
| SO1521 | iron-sulfur cluster-binding protein | 102098 |
| SO1524 | heat shock protein GrpE (grpE) | 22688 |
| SO1531 | thiamine biosynthesis protein ThiI (thiI) | 54950 |
| SO1538 | isocitrate dehydrogenase, NAD-dependent | 36002 |
| SO1556 | conserved hypothetical protein | 33629 |
| SO1568 | hypothetical protein | 19474 |
| SO1606 | metallo-beta-lactamase superfamily protein | 72855 |
| SO1629 | ribosomal protein S2 (rpsB) | 26553 |
| SO1630 | translation elongation factor Ts (tsf) | 30395 |
| SO1631 | uridylate kinase (pyrH) | 26235 |
| SO1632 | ribosome recycling factor (frr) | 20574 |
| SO1637 | bacterial surface antigen | 92460 |
| SO1648 | cold shock domain family protein | 8908 |
| SO1677 | acetyl-CoA acetyltransferase (atoB) | 40647 |
| SO1679 | acyl-CoA dehydrogenase family protein | 41946 |
| SO1680 | enoyl-CoA hydratase/isomerase family protein | 27331 |
| SO1682 | 3-hydroxyisobutyrate dehydrogenase (mmsB) | 30799 |
| SO1737 | hypothetical protein | 4402 |
| SO1776 | outer membrane protein precursor MtrB (mtrB) | 77629 |

| | | |
|---|---|---|
| SO1926 | citrate synthase (gltA) | 48049 |
| SO1928 | succinate dehydrogenase, flavoprotein subunit (sdhA) | 64188 |
| SO1930 | 2-oxoglutarate dehydrogenase, E1 component (sucA) | 104798 |
| SO1931 | 2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase (sucB) | 42804 |
| SO1932 | succinyl-CoA synthase, beta subunit (sucC) | 41335 |
| SO1961 | maltose O-acetyltransferase (maa) | 20920 |
| SO1981 | conserved hypothetical protein | 55082 |
| SO2001 | 5-nucleotidase (ushA) | 60953 |
| SO2012 | adenine phosphoribosyltransferase (apt) | 19531 |
| SO2016 | heat shock protein HtpG (htpG) | 71722 |
| SO2018 | adenylate kinase (adk) | 23079 |
| SO2086 | phenylalanyl-tRNA synthetase, beta subunit (pheT) | 86520 |
| SO2087 | integration host factor, alpha subunit (ihfA) | 10970 |
| SO2176 | conserved hypothetical protein | 7860 |
| SO2218 | asparaginyl-tRNA synthetase (asnS) | 52157 |
| SO2264 | cysteine desulfurase (iscS) | 44832 |
| SO1778 | decaheme cytochrome c (omcB) | 71191 |
| SO1779 | decaheme cytochrome c (omcA) | 78567 |
| SO1786 | glutaminyl-tRNA synthetase (glnS) | 64063 |
| SO1793 | trigger factor (tig) | 47567 |
| SO1796 | ATP-dependent protease La (lon) | 87468 |
| SO1797 | DNA-binding protein, HU family | 9440 |
| SO1798 | peptidyl-prolyl cis-trans isomerase D (ppiD) | 67456 |
| SO1824 | conserved hypothetical protein | 24943 |
| SO1825 | MotA/TolQ/ExbB proton channel family protein | 48740 |
| SO1828 | TonB2 protein (tonB2) | 22354 |
| SO1829 | TPR domain protein | 46407 |
| SO1851 | conserved hypothetical protein | 79998 |
| SO1870 | biosynthetic arginine decarboxylase (speA) | 70960 |
| SO1894 | acetyl-CoA carboxylase, biotin carboxylase, putative | 75609 |
| SO1896 | 3-methylcrotonyl CoA carboxylase, beta subunit (pccB-1) | 57949 |
| SO1897 | isovaleryl-CoA dehydrogenase (ivd) | 42048 |
| SO2265 | NifU family protein | 13535 |
| SO2274 | nucleoside diphosphate kinase (ndk) | 15473 |
| SO2299 | threonyl-tRNA synthetase (thrS) | 73673 |
| SO2300 | translation initiation factor IF-3 (infC) | 16458 |
| SO2302 | ribosomal protein L20 (rplT) | 13559 |
| SO2303 | thioredoxin reductase (trxB) | 34063 |
| SO2304 | alanine dehydrogenase, authentic point mutation (ald) | 39156 |
| SO2310 | seryl-tRNA synthetase (serS) | 46854 |
| SO2339 | alpha keto acid dehydrogenase complex, E1 component, alpha subunit | 43528 |
| SO2340 | alpha keto acid dehydrogenase complex, E1 component, beta subunit | 35350 |

| | | |
|---|---|---|
| SO2341 | alpha keto acid dehydrogenase complex, E2 component | 57284 |
| SO2345 | glyceraldehyde 3-phosphate dehydrogenase (gapA-2) | 36458 |
| SO2347 | glyceraldehyde 3-phosphate dehydrogenase (gapA-3) | 52256 |
| SO2348 | hypothetical protein | 4573 |
| SO2363 | cytochrome c oxidase, cbb3-type, subunit II (ccoO) | 23416 |
| SO2395 | acyl-CoA dehydrogenase family protein | 63178 |
| SO2402 | ribosomal protein S1 (rpsA) | 61225 |
| SO2406 | aspartate aminotransferase (aspC-2) | 43322 |
| SO2411 | DNA gyrase, A subunit (gyrA) | 101626 |
| SO2415 | ribonucleoside-diphosphate reductase, alpha subunit (nrdA) | 85658 |
| SO2419 | 2,4-dienoyl-CoA reductase, putative | 73680 |
| SO2427 | TonB-dependent receptor, putative | 91901 |
| SO2433 | aspartyl-tRNA synthetase (aspS) | 66034 |
| SO2469 | conserved hypothetical protein | 97699 |
| SO2477 | alcohol dehydrogenase, iron-containing | 39410 |
| SO2491 | pyruvate kinase II (pykA) | 51154 |
| SO2492 | oxidoreductase, acyl-CoA dehydrogenase family | 83373 |
| SO2495 | Smr domain protein | 21843 |
| SO2506 | excinuclease ABC, B subunit (uvrB) | 76114 |
| SO2592 | dihydroorotate dehydrogenase (pyrD) | 36559 |
| SO2593 | conserved hypothetical protein | 182219 |
| SO2601 | carboxyl-terminal protease | 76395 |
| SO2618 | ATP-binding protein, Mrp/Nbp35 family | 39155 |
| SO2619 | methionyl-tRNA synthetase (metG) | 76238 |
| SO2629 | isocitrate dehydrogenase, NADP-dependent (icd) | 80450 |
| SO2635 | adenylosuccinate lyase (purB) | 50776 |
| SO2636 | conserved hypothetical protein | 42728 |
| SO2638 | leucine dehydrogenase (ldh) | 37058 |
| SO2644 | phosphoenolpyruvate synthase (ppsA) | 86722 |
| SO2649 | cys regulon transcriptional activator (cysB) | 36426 |
| SO2713 | transporter, putative | 24603 |
| SO2730 | peptidase E (pepE) | 25941 |
| SO2746 | conserved hypothetical protein | 27220 |
| SO2756 | antioxidant, AhpC/Tsa family, authentic frameshift | 21827 |
| SO2760 | phosphoribosylformylglycinamidine cyclo-ligase (purM) | 37484 |
| SO2774 | 3-oxoacyl-(acyl-carrier-protein) synthase II (fabF-1) | 42836 |
| SO2775 | acyl carrier protein (acpP) | 8565 |
| SO2776 | 3-oxoacyl-(acyl-carrier-protein) reductase (fabG-1) | 26365 |
| SO2777 | malonyl CoA-acyl carrier protein transacylase (fabD) | 31788 |
| SO2780 | ribosomal protein L32 (rpmF) | 6281 |
| SO2785 | ribonuclease E (rne) | 120421 |
| SO2787 | cold shock domain family protein | 7519 |
| SO2791 | cytidine deaminase (cdd) | 31964 |
| SO2800 | conserved hypothetical protein | 6500 |

| | | |
|---|---|---|
| SO2881 | superoxide dismutase, Fe (sodB) | 21482 |
| SO2903 | cysteine synthase A (cysK) | 34399 |
| SO2907 | TonB-dependent receptor domain protein | 95702 |
| SO2912 | formate acetyltransferase (pflB) | 84806 |
| SO2915 | acetate kinase (ackA) | 43497 |
| SO3023 | tryptophan synthase, beta subunit (trpB) | 42789 |
| SO3064 | amidophosphoribosyltransferase (purF) | 55719 |
| SO3070 | aspartate semialdehyde dehydrogenese (asd) | 37166 |
| SO3072 | 3-oxoacyl-(acyl-carrier-protein) synthase I (fabB) | 43105 |
| SO3099 | long-chain fatty acid transport protein, putative | 47034 |
| SO3110 | protein-export membrane protein SecF (secF-2) | 34073 |
| SO3146 | DNA-binding protein, H-NS family | 14591 |
| SO3154 | prolyl-tRNA synthetase (proS) | 63156 |
| SO3157 | lipoprotein, putative | 26194 |
| SO3175 | asparagine synthetase, glutamine-hydrolyzing (asnB-2) | 73358 |
| SO3178 | hypothetical protein | 37828 |
| SO3185 | polysaccharide biosynthesis protein | 40454 |
| SO3186 | glucose-1-phosphate-thymidylyltransferase (rfbA) | 33792 |
| SO3188 | dTDP-glucose 4,6-dehydratase (rfbB) | 42413 |
| SO3190 | polysaccharide biosynthesis protein | 47203 |
| SO3286 | cytochrome d ubiquinol oxidase, subunit I (cydA) | 58078 |
| SO3287 | phosphoribosylformylglycinamidine synthase (purL) | 140509 |
| SO3293 | inosine-5-monophosphate dehydrogenase (guaB) | 51652 |
| SO3310 | conserved hypothetical protein | 22261 |
| SO3311 | histidyl-tRNA synthetase (hisS) | 47494 |
| SO3312 | 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (ispG) | 40578 |
| SO3343 | conserved hypothetical protein | 20640 |
| SO3403 | ribosomal subunit interface protein (yfiA-1) | 13143 |
| SO3417 | peptidyl-prolyl cis-trans isomerase SlyD (slyD) | 23476 |
| SO3424 | valyl-tRNA synthetase (valS) | 108686 |
| SO3425 | hypothetical protein | 5942 |
| SO3426 | carbon storage regulator (csrA) | 7121 |
| SO3428 | alanyl-tRNA synthetase (alaS) | 94833 |
| SO3430 | recA protein (recA) | 38049 |
| SO3440 | enolase (eno) | 42124 |
| SO3441 | CTP synthase (pyrG) | 60111 |
| SO3455 | GTP pyrophosphokinase (relA) | 83311 |
| SO3466 | riboflavin synthase, beta subunit (ribH) | 16679 |
| SO3467 | 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II, putative (ribBA) | 40171 |
| SO3471 | serine hydroxymethyltransferase (glyA) | 45254 |
| SO3532 | isoleucyl-tRNA synthetase (ileS) | 105725 |
| SO3537 | ribosomal protein S20 (rpsT) | 9751 |
| SO3545 | OmpA family protein | 40157 |

| | | |
|---|---|---|
| SO3547 | glucose-6-phosphate isomerase (pgi) | 59624 |
| SO3649 | GTP-binding protein, GTP1/Obg family | 42842 |
| SO3652 | ribosomal protein L21 (rplU) | 11338 |
| SO3669 | heme transport protein (hugA) | 76324 |
| SO3715 | oxygen-insensitive NAD(P)H nitroreductase | 24169 |
| SO3733 | hypothetical protein | 62759 |
| SO3765 | conserved hypothetical protein | 25374 |
| SO3772 | conserved hypothetical protein | 20142 |
| SO3780 | ABC transporter, ATP-binding protein CydD (cydD) | 70376 |
| SO3800 | serine protease, subtilase family | 126262 |
| SO3803 | hypoxanthine phosphoribosyltransferase (hpt-2) | 19928 |
| SO3837 | ribose-phosphate pyrophosphokinase (prsA) | 33964 |
| SO3842 | conserved hypothetical protein | 34092 |
| SO3844 | peptidase, M13 family | 77158 |
| SO3863 | molybdenum ABC transporter, periplasmic molybdenum-binding protein (modA) | 27321 |
| SO3870 | disulfide bond formation protein | 23879 |
| SO3896 | outer membrane porin, putative | 39874 |
| SO3915 | conserved hypothetical protein | 16501 |
| SO3927 | ribosomal protein L9 (rplI) | 15652 |
| SO3928 | ribosomal protein S18 (rpsR) | 8840 |
| SO3930 | ribosomal protein S6 (rpsF) | 14993 |
| SO3937 | adenylosuccinate synthetase (purA) | 46897 |
| SO3939 | ribosomal protein S9 (rpsI) | 14527 |
| SO3940 | ribosomal protein L13 (rplM) | 15733 |
| SO3942 | serine protease, HtrA/DegQ/DegS family | 46522 |
| SO3952 | mce-related protein | 16919 |
| SO3967 | molybdenum ABC transporter, periplasmic molybdenum-binding protein, putative | 27080 |
| SO3988 | aerobic respiration control protein ArcA (arcA) | 27203 |
| SO3991 | fructose-1,6-bisphosphatase (fbp) | 35279 |
| SO4002 | sensory transduction histidine kinase | 186636 |
| SO4013 | hypothetical protein | 68497 |
| SO4034 | ATP-dependent RNA helicase DeaD (deaD) | 69339 |
| SO4098 | rod shape-determining protein MreB (mreB) | 37249 |
| SO4105 | MSHA pilin protein MshA (mshA) | 17645 |
| SO4120 | ribosomal protein L31 (rpmE) | 7607 |
| SO4133 | uridine phosphorylase (udp) | 26835 |
| SO4163 | heat shock protein HslVU, ATPase subunit HslU (hslU) | 49769 |
| SO4190 | inorganic pyrophosphatase, manganese-dependent (ppaC) | 33428 |
| SO4202 | Sec-independent protein translocase protein TatA (tatA) | 9333 |
| SO4208 | delta-aminolevulinic acid dehydratase (hemB-2) | 37121 |
| SO4211 | preprotein translocase, SecA subunit (secA) | 102559 |
| SO4215 | cell division protein FtsZ (ftsZ) | 40910 |
| SO4235 | 3-isopropylmalate dehydrogenase (leuB) | 39343 |

| | | |
|---|---|---|
| SO4240 | hypothetical protein | 43510 |
| SO4264 | type I restriction-modification system, S subunit (hsdS-2) | 55104 |
| SO4313 | porphobilinogen deaminase (hemC) | 33443 |
| SO4317 | RTX toxin, putative | 284925 |
| SO4320 | agglutination protein (aggA) | 52311 |
| SO4321 | OmpA family protein | 22162 |
| SO4408 | virulence regulator BipA (bipA) | 65940 |
| SO4410 | glutamine synthetase, type I (glnA) | 51741 |
| SO4507 | TorA specific chaperone, putative | 25332 |
| SO4513 | formate dehydrogenase, alpha subunit | 105376 |
| SO4514 | formate dehydrogenase, iron-sulfur subunit (fdhB-2) | 20478 |
| SO4557 | methyl-accepting chemotaxis protein | 67264 |
| SO4602 | glycerol-3-phosphate acyltransferase (plsB) | 91504 |
| SO4623 | DNA-binding response regulator | 25497 |
| SO4669 | DNA polymerase I (polA) | 101845 |
| SO4674 | 2-amino-3-ketobutyrate coenzyme A ligase (kbl) | 42878 |
| SO4680 | conserved hypothetical protein | 41126 |
| SO4681 | glycosyl transferase, group 1 family protein | 45630 |
| SO4719 | conserved hypothetical protein | 29348 |
| SO4741 | glucosamine--fructose-6-phosphate aminotransferase (isomerizing) (glmS) | 66687 |
| SO4747 | ATP synthase F1, beta subunit (atpD) | 49742 |
| SO4748 | ATP synthase F1, gamma subunit (atpG) | 31495 |
| SO4749 | ATP synthase F1, alpha subunit (atpA) | 55114 |
| SOA0003 | type II restriction endonuclease, putative | 39357 |
| SOA0025 | ISSo1, transposase OrfA | 11170 |
| SOA0051 | hypothetical protein | 13943 |
| SOA0106 | methyl-accepting chemotaxis protein | 59632 |
| SOA0110 | lipoprotein, putative | 145837 |
| SOA0112 | lipoprotein, putative | 145527 |
| SOA0115 | lipoprotein, putative | 145527 |
| SOA0135 | hypothetical protein | 63737 |
| SOA0161 | zinc-binding dehydrogenase | 40145 |

## CHAPTER 5

## Characterizing Electrospray Ionization Using Atmospheric Pressure Ion Mobility Spectrometry


### Abstract

Reduced flow-rate electrospray ionization has been proven to provide improved sensitivity, less background noise, and improved limits of detections for ESI-MS analysis. Miniaturizing the ESI source from conventional electrospray to micro-electrospray and further down to nano-electrospray has resulted in higher and higher sensitivity. However, when effects of flow rate were investigated for atmospheric pressure ESI-IMS using a nanospray emitter, a striking opposite result was observed. The general tendency we observed in ESI-IMS was that higher flow rate offered higher ion signal intensity throughout a variety of conditions investigated. Thus further efforts were undertaken to rationalize these contradictory results. Our study revealed that unlike in ESI-MS, decreased flow rate increases both ionization efficiency and transmission efficiency thus improves ion signal; whereas in ESI-IMS, ion transfer is constant despite flow rate decreases thus ion signal decreases. Since ion transfer is constant in atmospheric pressure ESI-IMS, ionization efficiency can be studied independently, which otherwise is not possible in ESI-MS where both ionization efficiency and transmission efficiency vary as conditions alter. In this report, we present a systematic study on signal intensity and ionization efficiency at various experimental conditions using ESI-IMS and

demonstrated the ionization efficiency as a function of flow rate, analyte concentration, and solvent composition.

## Introduction

Electrospray ionization (ESI), as an atmospheric ionization technique to generate ions by spraying analyte solution at high voltage, was first introduced by Dole *et al* in 1968.[1] Fenn *et al* further demonstrated the applications ESI in mass spectrometry (MS) techniques in the early 1980s[2, 3]and received the Nobel prize in chemistry in 2002 due to this invention. ESI has become one of the most important and powerful ionization techniques for MS because of its effectiveness in detecting large biomolecules[4] and ease-of-use for interfacing liquid-based separation techniques such as liquid chromatography (LC)[5] and capillary electrophoresis (CE).[6,7]

A conventional ESI source typically employs flow rates in the range of 4-200 μL/min through a capillary that has 50-200 μm inner diameter (i.d.).[5] A micro-electrospray source makes use of a tapered emitter with a small terminal orifice (usually < 20 μm i.d.) and much lower flow rates in the range of 0.1-1 μL/min. The use of low flow rate in ESI-MS has proven advantageous in detection of biological molecules since it showed significantly improved sensitivity.[8,9] At nanoliter per min flow rates, the attomole limit of detection of proteins is possible.[10] With these successes, the edge of envelope was further pushed by Wilm and Mann[11,12] who introduced nano-electrospray (or nanospray) ionization with further decreased flow rate. Without the use of solvent pump, this static nanospray ion source allows the liquid to flow through the spray emitter by capillary action when high voltage is applied. The measured flow rate is as low as 20-

40 nL/min through a very small tip that has ~1 μm i.d.. It is believed that the lower flow rate provides better ionization efficiency because it produces smaller initial charged droplets which require less solvent evaporation and fewer droplet fission events prior to eventual ion release in the gas phase.[13,14] In addition, the closer alignment at the interface between the emitter tip and the MS inlet aperture or capillary and the elimination of the need for sheath flow and sheath gas with the use of nanospray source have been proven to improve ion transmission efficiency.[15] As a result, the overall detection efficiency hence sensitivity is significantly enhanced.

A few years after the development of ESI, Dole and co-worker demonstrated the detection of lysozyme with three broad peaks by coupling ESI for ion mobility spectrometry (IMS).[16] Similar ESI-IMS spectra of lysozyme and cytochrome *c* were reported by Smith and co-workers in 1991.[17] We further advanced ESI-IMS in analysis of large molecule with improved sensitivity and resolution by incorporating counter flow heated drift gas and water-cooled ESI source.[18] Very recently, Creaser and coworkers reported the first application of nano-ESI in an atmospheric pressure ion mobility spectrometer.[19, 20] However unlike nano-ESI-MS which shows great sensitivity with use of low μM to nM concentration of analyte solution, in their studies instead, significantly higher concentration of solution (> 5 mM) were used to obtain signals for peptides in ESI-IMS. No systematic study of various flow rates ranging from nanoliters per min to microliters per min for ESI-IMS has been reported yet. To this end, we took further efforts to look into the rationale behind the opposite flow rate effects observed in ESI-IMS. We designed and implemented a target Faraday detector to characterize ion transmission in ESI-IMS. With constant ion transmission and minimal loss through the

drift space, we were able to characterize electrospray ionization efficiency under various experimental parameters. In this report, the unique atmospheric pressure ion detection property offered by IMS was utilized for the first time to evaluate and compare ionization efficiency for electrospray ion source.

## Experimental Section

**Chemicals and Reagents**. Methanol (MeOH) used in this study was HPLC grade and obtained from J. T. Baker (Phillipsburgh, NJ). Acetic acid (HOAc) and hydrofluoric acid (HF) were obtained from Sigma (St. Louis, MO). Water was 18-MΩ deionized, prepared by Barnstead Nanopure Water Systems. Cocaine, amphetamine, and caffeine were purchased from RBI (Natick, MA). L-serine was purchased from Sigma (St. Louis, MO).

**Ion Mobility Spectrometer**. The atmospheric pressure ion mobility spectrometer used for this research was constructed at Washington State University, Pullman, WA. The instrument configuration is shown in Figure 1. The drift tube was build upon a standard stacked ring configuration by assembling repeating units. Each repeating unit, composed of a conductive stainless steel ring with a dimension of 50 mm (o.d.) × 48 mm (i.d.) × 3 mm (width) and an insulating ceramic ring with a dimension of 60 mm (o.d.) × 50 mm (i.d.) × 4.5 mm (width), is stacked sequentially in an alumina tube. The stainless steel rings, also called guard rings, were connected via a series of half (desolvation region) and one (drift region) megaohm resistors (Caddock Electronics Inc., Riverside, CA). The ceramic ring served to isolate the guard rings from the alumina tube as well as from each

other. The current instrument consists of 42 repeating units with a total length of 34 cm, but the length of the drift tube can be easily shortened or lengthened by removing or adding a number of rings. A Bradbury-Nielsen gate ring divided the drift tube into a 7.5-cm long desolvation region and a 26.5-cm long drift region. A 10 kV voltage was normally applied to the first ring electrode and the last ring electrode voltage was adjusted by variable resistor to be ~200 V referenced to ground. The drift voltage was dropped gradually across the drift tube via the resistor chain to form an electric field of 157 V/cm in the desolvation region and 333 V/cm in the drift region. The lower electric field in the desolvation region allowed solvated ions to spend more time in the heated drift gas to get more efficient desolvation prior to injection to drift region. The Bradbury-Nielsen gate was made of electrically isolated alternating parallel Alloy 46 wires (76 μm in diameter) (California Fine Wire Co., Grover Beach, CA) spaced 0.65 mm apart. When the potentials applied on the alternating wires were the same as the reference potential, the gate was "open" to allow ions to pass through; while the potentials on the adjacent wires were offset ±50 V with the respect to the reference potential, an electric field of ~1500 V/cm was created orthogonal to the drift field and the gate was "closed" to shut off ion transmission.

An aperture grid ring, made in a similar manner as the Bradbury-Nielsen gate ring with the exception that adjacent wires were in common, was placed right in front of the terminal Faraday plate (60 mm in diameter) with ~0.5 mm spacing. The function of the aperture grid was to shield the incoming ion cloud from the detector prior to its arrival and reduce peak broadening. The counter flow drift gas was introduced to the drift tube near the tube terminus through a hollow drift ring which had 8 radially distributed

apertures. The drift tube oven was constructed from two pieces of 20 cm long aluminum cylinder with 3 heating cartridges (Heatcon, Seattle, WA) embedded inside each cylinder. The electronics system used for the instrument included a high voltage power supply for the ESI emitter, a high voltage power supply for the drift tube, a two-channel temperature controller for the drift tube oven, a gate controller, and a current amplifier for signal amplification. The data acquisition software was programmed by LabVIEW 6.1 (National Instruments, Austin, TX) and communicated with the gate controller and current amplifier through an interface board (PCI-6030E) (National Instruments, Austin, TX) installed in a Dell PC computer.

**Target Faraday Detector**. In order to investigate ion beam's distribution at the drift tube exit, we designed and built a target-like, segmented, Faraday plate detector (Figure 1), which was composed of a set of concentric copper rings mounted on an insulating support. Each copper ring, measured ~2 mm wide and separated from the adjacent rings with a space of ~0.2 mm, was attached to an individual lead that was wired to the current amplifier separately through a switch box. Ion signal can be either measured individually for each ring or measured with any combined number of rings with all other rings grounded. The charge density (coulomb/$mm^2$) was calculated by dividing the peak area with the surface area of the respective ring.

**Nano-ESI Source**. The nanospray emitter was made by etching fused silica capillary (360 μm o.d. × 20 μm i.d., Polymicro Technologies, Phoenix, AZ) in 49% HF solution. The wall thickness of the emitter terminus was thinned down to 0.1 μm or less after 30

min to 1 hr etching. The emitter was then cut to be ~5 cm long. The relatively large tip aperture (20 μm i.d.) was chosen for this research because it allowed the liquid to be delivered through the emitter at relatively wide range of flow rates. Normally flow rate was varied from 0.1 μL/min up to 10 μL/min using the solution delivered from the same syringe (250 μL Hamilton gastight syringe, Reno, NV) for the entire set of the experiments, which ensured consistency throughout. Flow rates over 12 μL/min often caused backpressure overtime and eventually broke weak connections between the syringe needle and the injection port. The electrical contact was applied through a stainless steel zero dead volume union (Upchurch Scientific, Oak Harbor, WA) that connected the emitter with the fused silica capillary transfer line (360 μm o.d. × 75 μm i.d., ~30 cm long).

**Operating Conditions**. Electrospray ionization was normally induced by applying a potential difference of 2.5-3.5 kV between the nanospray emitter and the target focus screen (16-mesh stainless steel), which also served as the first ring electrode for the drift tube. 10 kV was normally applied to the first ring electrode and 13 kV was used for the nanospray emitter. It should be noted that the effective drift voltage across the drift region was the potential difference between the gate ring and the aperture grid ring (typically ~8.6 kV) but not total applied voltage. The drift tube was normally heated at 200-250$^{o}$C. $N_2$ was used as counter flow drift gas throughout the experiments, which was injected to the drift tube at a constant flow rate of 500 mL/min to sweep out neutral interferences during spectral acquisitions. All experiments were carried out at atmospheric pressure which typically ranged from 690 to 710 Torr in Pullman, WA. The

solvent used for most experiments was a mixture 50% MeOH/45% $H_2O$/5% HOAc unless otherwise specified. A syringe pump (KD Scientific, Holliston, MA) was used to deliver solution to the spray emitter at the desired flow rate. The charged droplets emitting from ionization source were further desolvated by heated drift gas while migrating slowly in the desolvation region and desolvated ions were then injected into the drift region by opening the ion gate for a brief 0.2 ms pulse width at a frequency of 20 Hz. The target Faraday detector was only used to acquire spectra for charge and current distribution study. All the other experiments used the regular Faraday plate. The current amplifier gain was usually set at $10^9$ V/amp. Each IMS spectrum was data averaged from 500 gate pulses. 3-5 IMS spectra were acquired consecutively to obtain one averaged data point (peak intensity).

**Results and Discussion**

It is well accepted that the sensitivity of ESI-MS measurements improves inversely with liquid flow rate. Miniaturizing ionization source from conventional ESI to nano ESI has resulted in signal improvements by orders of magnitude. Nanospray ionization using small tip diameter in a few μm and low flow rate in nanoliters per minute has been the method of choice in most biological mass spectrometry laboratories because of limited sample quantity. However, when nanospray was applied to a standalone atmospheric pressure IMS instrument, an opposite flow-rate effect was observed, i.e., ion signal intensity steadily increased with increased flow rate. Thus further experiments were carried out in this study to look into the rationale behind these contradictory observations.

**Initial Observations of Flow-rate Effects**. A nanospray emitter that was normally used for nano-ESI-MS applications in our laboratory[21,22] was employed for all the ESI-IMS experiments in this study. The spray emitter was made by etching fused silica capillary (360 μm o.d. × 20 μm i.d.) with HF solution until thin wall (50-80 nm thickness)[8, 23] was formed at the tip terminus. The extremely thin terminal wall thickness helps generate high electric field for stable electrospray.[24] Initially it was expected that better sensitivity should be achieved with nanospray than microspray (50 μm tip i.d., 2-5 μL/min flow rate) that was routinely used for ESI-IMS in our lab.[25,26] However, ion signals proportionally increased with increased flow rates when 100 ppb lutidine samples were electrosprayed into the IMS. Thus further investigations were corroborated by using a variety of compounds at different concentrations. Consistent results were observed throughout all the experiments. Figure 2 shows a few selected examples including 50 μM cocaine, 100 μM L-serine, 10 μM amphetamine, and 10 μM caffeine. These observations seemed to indicate ESI-IMS response is mass-sensitive rather than concentration-sensitive, which is exactly opposite to the common findings in ESI-MS. Figure 2b shows the full ESI-IMS spectra of 100 μM L-serine with flow rates ranging from 0.1 μL/min to 7.5 μL/min. As shown in the figure, the peak intensity for L-serine with drift time of ~23.5 ms increased steadily with flow rate. However, signal intensities for the two solvent ion peaks exhibited the opposite trend, i.e., higher response was obtained at lower flow rate. In addition, their drift times shifted dramatically to longer times (the first solvent peak went from the drift time of 16.4 ms to 19.1 ms and the second solvent peak migrated from 20.4 ms to 22.2 ms when flow rate changed from 0.1 μL/min to 7.5 μL/min). These

observations for the solvent peaks indicated that at higher flow rate larger solvent cluster ions were created and the solvent components ($H_2O$, MeOH, and HOAc) were ionized less sufficiently, which conformed to the results one would expect from ESI-MS experiment. Furthermore, when a closer look was taken on the L-serine peak, it was found the drift time of L-serine was slightly shifted to longer time as flow rate increased, from 23.4 ms at 0.1 μL/min to 23.6 ms at 7.5 μL/min. Hill *et al* reported previously that the reduced mobility values for some analytes by using ESI-IMS are lower than those by $^{63}$Ni-IMS and suggested this is the consequence of incomplete desolvation of the analyte ions.[18] In analogy, we can deduce that larger cluster solvent ions with longer drift time observed in our experiments were the result of less efficient desolvation at higher flow rate. However, the higher response obtained for the analyte ions at higher flow rates had not been addressed in previous literature.

**Ion Transmission in ESI-IMS.** Since the flow-rate induced opposite effects on ion signals in ESI-IMS and ESI-MS, we speculated this difference was derived primarily from the different ion transmission processes inherent with these two techniques (Figure 3). In ESI-IMS, ions are created, transferred, and detected in the atmospheric pressure regime throughout. In ESI-MS, on the other hand, ions are formed in atmospheric pressure, then transported through various stages of vacuum regions with gradual reduction in pressure by differential pumping, and finally detected by MS analyzers at high-vacuum. In general the sensitivity or overall detection efficiency for ESI-MS is determined by two major factors, i.e., ionization efficiency and ion transmission efficiency. Nanospray becomes the method of choice due to both its improved

ionization efficiency and ion transmission efficiency. It is believed that the small-sized initial droplet emanating from small orifice tip at low flow rate experiences less evaporation and fission processes prior to eventual ion formation in gas phase. Wilm and Mann[12] reported that at 1 μM concentration and sufficiently low flow rate (< 40 nL/min), nanospray can produce very small-sized droplet (< 200 nm in diameter) which on average contains only one analyte molecule. Thus, ionization efficiency can approach 100% if every analyte molecule is dispersed in very small charged droplet. With 100% ionization efficiency, the remaining determent for sensitivity of ESI-MS is ion transmission efficiency. Most ESI-MS instruments employ a very small orifice as the MS inlet using either a small pinhole (~50-200 μm in diameter) or a capillary tube (~0.5 mm i.d.) to achieve the requisite pressure drop between the atmospheric ion source region and the first vacuum stage in MS. However, charged droplets exiting the electrospray Taylor cone are expanding into a plume which is significantly larger than the MS inlet orifice. Consequently only a small portion of total ions can be transported into the MS inlet by dynamic gas flow and potential difference imposed between the spray emitter and the MS inlet (Figure 3). Efforts to increase the size of MS entrance with the use of multicapillary inlet[27] or flared inlet tube[28] have been reported to obtain better sampling efficiency. However, there is an upper limit for the MS inlet size since higher pump speed is needed to maintain the low pressure as MS inlet size increases. In this regard, nano-ESI is advantageous over conventional ESI in terms of ion transmission efficiency as well because closer proximity and in-axis alignment between the nanospray emitter and the MS inlet allows higher transfer efficiency.[15]

When nanospray is applied to ESI-IMS, the ionization process is similar to that in ESI-MS, however, the ion transfer process is totally different in a couple of respects. First, without the vacuum restriction, the IMS instrument operated at atmospheric pressure can have the entrance aperture as large as possible to accept 100% of incoming ion plume (Figure 3). The initial charged droplets or ion plume are drawn into the drift tube by potential difference applied between the spray tip and the IMS inlet without any gas dynamic disturbance. Second, the Faraday plate detector with the acceptance dimension covering the entire IMS exit aperture allows capturing all arrival ions if no ions are lost during their transmission through the drift space. And finally, ion transfer is constant regardless of flow rate variation. However, we know that ion packets injected to the drift tube experience further spatial broadening in the drift region due to thermal diffusion, Coulomb expansion, and electric field inhomogeneity.[25, 29] If the width of the ion packets exceeds the inner diameter of the drift tube prior to its arrival at the IMS terminus, then ions are lost and ion transmission efficiency is reduced. The transmission efficiency through the IMS space alone was assumed to be 100% when Tang et al demonstrated near-perfect ion transfer through an ion funnel interface for ESI-IMS-MS.[30] Sysoev *el al* also concluded ion loss during transmission through IMS tube was minimal when they obtained similar signals with and without a modular IMS inserted between ESI-MS.[31] Gillig *et al* simulated ion trajectories inside the drift tube and showed that ions were transferred without any loss through the drift tube when they reported an ion focusing guide for MALDI-IMS-MS.[32] However to our knowledge, no direct empirical data has been reported to demonstrate ion transmission is close to 100% in IMS tube.

To picture the incoming ion beam at the IMS terminus, we designed and incorporated a target-like segmented Faraday plate detector (Figure 1) to detect arrival ion distribution at the drift tube terminus. This target detector was composed of 12 concentric conductive copper rings (2 mm width for each ring) spaced ~0.1 mm apart. The diameter of each ring increased radially as it was further away from the center, e.g., ring 1 in the center has an o.d. of 5 mm and ring 12 in the most edge has an o.d. of 49 mm. Each ring was wired to the current amplifier individually through a switch box. The ion current or signal can be measured for each individual ring or any number of combined rings. Because the non-conductive resin support that target detector was mounted to is not heat-resistant, the current distribution experiment was performed at room temperature (18$^{o}$C) using 50 μM cocaine with following parameters: 500 mL/min counter flow $N_2$ drift gas, 2.5 μL/min infusion flow rate, 15 kV and 11.5 kV applied to the nanospray emitter and the first drift ring, respectively. For IMS spectrum acquisition, a combination of 2 adjacent rings was used to ensure good quality for each spectrum with the exception of ring 11 and ring 12 where no ion signal was deteced. Figure 4a shows an example of IMS spectrum collected with a combination of ring 5 and 6. Two peaks were detected for each spectrum. The first peak is the solvent peak with the drift time of ~40.4 ms; and the second peak is cocaine with the drift time of ~53.1 ms. The charge density was calculated by dividing the analyte peak area ($Q = It$, where $Q$ is charge in coulomb, $I$ is current in ampere, and $t$ is time in second) over the ring area (mm$^2$), then plotted against the radius of each ring (Figure 4b). As shown in the figure, the charge density started to decay when ring diameter reached over 34 mm, and no ion signal was observed on ring 11 and ring 12 which covered the diameter of 45 mm and beyond.

Approximately 90% of the total charge was found within 33 mm diameter, suggesting that the arrival ion beam circumference was smaller than the inner diameter (48 mm) of the drift tube. Thus it can be concluded that no ions were lost to the inner wall of the drift tube since, if any signal was observed in ring 12, one would expect that the ion beam grew bigger and exceeded the size of drift tube at certain point during its transmission. To further confirm this conclusion, the total current on each ring was also measured, and current density was calculated and plotted versus its ring radius (Figure 4c). Similarly, most current was found within 36 mm diameter and current rapidly decayed to zero at ring 11. Siems et al[29] reported the edge effect due to inhomogeneous electric field in IMS could cause ion losses at the drift wall and speculated that the edge effect was the reason that no ion signals are observed at the edge of detector. However in our experiments, since a majority of current or ion signal was detected within 33 mm circumference, which is much smaller than the drift tube inner diameter (48 mm), we speculate that ion loss in the drift space is minimal or negligible.

**Electrospray Ionization Efficiency** Higher flow rate tolerance in ESI-IMS than that in ESI-MS was reported previously.[33,34] In order to explain why flow rate responses behave differently in ESI-IMS and ESI-MS, we need to consider the overall ion detection efficiency (*DE%*) which ultimately determines the sensitivity of the detection method. Overall ion detection efficiency, defined as total number of ions recorded at the detector ($n$) divided by the total number of analyte molecules introduced ($n'$), is dependent on two factors: ionization efficiency (*IE%*) and ion transmission efficiency (*TE%*).[35,36] Therefore we have $DE\% = (n/n') 100\% = IE\% \cdot TE\%$. The bottleneck that limits the ion

transmission efficiency due to the small orifice inlet employed in ESI-MS is not a problem for atmospheric pressure ESI-IMS. Ion transfer at the IMS inlet is close to 100% due to the large entrance orifice and further ion transfer in the drift tube showed minimal loss as discussed in the Target detector study. Importantly, ion transfer in ESI-IMS is constant regardless of flow rate variations and ion loss is negligible as compared to that in ESI-MS. As a consequence, the overall ion detection efficiency in ESI-IMS should be only affected by one variable, ionization efficiency. The overall ion detection efficiency can be calculated from experimental measurements, thus ionization efficiency can be obtained.

To calculate overall ionization efficiency, first we need to calculate the total number of ions detected by Faraday plate. We know the total charge ($Q$) in coulomb of all ions is given by

$$Q = I\,t \qquad \textit{Equation 1}$$

where $I$ is the current in ampere and $t$ is time in second. $Q$ can also be measured from IMS spectrum since $I\,t = Peak\ Area$. In addition, according to Coulomb's law,

$$Q = n\,z\,e \qquad \textit{Equation 2}$$

where $n$ is the number of ions, $z$ is the charge state of the ion (e.g., $z = 1$ for proton), and $e$ is the charge on one electron which is $1.602 \times 10^{-19}$ coulomb. Thus the total number of ions detected ($n$) can be calculated by

$$n = \frac{Peak\ Area}{ze} \qquad \textit{Equation 3}$$

On the other hand, the total number of analyte molecules introduced (*n'*) at the ion source to the drift tube at every gate pulse is equal to

$$n' = cft_g N_{avg} \qquad \textit{Equation 4}$$

in which $c$ is the analyte concentration in molar, $f$ is flow rate in liter/second, $t_g$ is gate pulse width in second, and $N_{avg}$ is Avogadro's constant ($6.0221415 \times 10^{23}$ $mol^{-1}$). Thus the overall detection efficiency (*DE%*) is given by

$$DE\% = \frac{Peak\ Area}{zecft_g N_{avg}} \times 100\% \qquad \textit{Equation 5}$$

*TE%* is constant in ESI-IMS and we assume *TE%* is 100% here since ion loss is minimal as compared to that in ESI-MS, thus *DE% = IE%*. In this regard, atmospheric IMS provides a good tool for characterizing electrospray ionization.

**Effects of Analyte Concentration** With the equation for calculating ionization efficiency, we further evaluated effects of flow rate on ionization efficiency by varying analyte concentration. Cocaine was used as analyte for this study with the concentration varying from 1 μM to 200 μM in the solution of 50% MeOH, 45% $H_2O$, and 5% HOAc. At each concentration, the IMS spectra were acquired with flow rate variation from 0.05 μL/min to 10 μL/min. All the experiments were performed using the same operating parameters: 13.5 kV ESI voltage, 10 kV drift voltage, 8.8 kV gate voltage, 500 mL/min $N_2$ drift gas flow, and 225°C oven temperature. Figure 5a shows plots of cocaine peak intensities extracted from IMS spectra as a function of flow rate for different cocaine concentration. When concentration was below 50 μM, the response of cocaine was steadily increased as flow rate increased. At the concentration of 100 μM and 200 μM, the ion signals started to plateau when flow rate was over 2.5 μL/min and 5 μL/min,

respectively, indicating the total number of ions detected at the Faraday detector reached the maximum. These results were consistent with our initial observations described earlier, i.e., higher flow rate produced higher response. Next we calculated the ionization efficiency (*IE%*) using *Equation 5* and plotted it as a function of flow rate at different concentrations (Figure 5b). In this case $z = 1$ since cocaine is detected as a singly charged ion.[34] As predicted from any ESI-MS experience, the ionization efficiency increased when flow rate decreased. This finding has been known for many years in the field of mass spectrometry and the theoretical reasoning has also been extensively characterized, nonetheless, this data presented the important empirical proof to confirm that ionization efficiency is higher with lower flow rate. For the first time, the electropray ionization efficiency can be characterized independently without consideration of the other variable, transmission efficiency, with the use of atmospheric pressure IMS. In addition, Figure 5b also shows the ionization efficiency was lower when the analyte concentration was higher. This is due to the fact that at higher concentration, the competition among the analyte molecules for available charges on the droplets increases, resulting in fewer analyte molecules ionized.

Regardless of improved ionization efficiency obtained at lower flow rate and lower concentration, in ESI-IMS ion response signal increased with increasing flow, opposite to the trend observed in ESI-MS. We know ion signal intensity is directly related to the total number of ions captured ($n$) by detector, which is proportional to both *DE%* and total number of analyte molecules injected ($n'$) at ion source.

$$n = n'DE\% = cft_g N_{avg} DE\% \qquad \textit{Equation 6}$$

Thus we calculated *n* in the unit of femtomole (fmol) using *Equation 6* and plotted it versus flow rate (Figure 5c). These curves in Figure 5c showed basically the same trend and shape as those in Figure 5a, which explained why the increased ion response was observed at higher flow rate. Although the ionization efficiency decreases, the absolute amount of analyte (*n'*) introduced increases at higher flow rate and dominate *Equation 6*, thus contributing to higher signal response. When the flow rate further increases, ionization efficiency further decreases and becomes the dominating factor, ion response will plateau and eventually decrease. We did not carry out any experiments with flow rate over 10 μL/min due to the backpressure restriction from the spray emitter. However, the observation that signal response decreased at substantially high flow rate with 50 μm i.d. emitter was reported previously.[33]

**Effects of Solvent Composition** A composition of 1:1 ratio of water and organic solvent (methanol or acetonitrile) with 0.1-5% of ion paring agent such as acetic acid, formic acid, or trifluoroacetic acid is often used as electrospray solution for most analytes. Organic modifier has lower surface tension than water which helps disperse charged droplets into smaller droplets and thus promote ionization efficiency. In addition, organic solvent helps improve desolvation due to its higher volatility. To test the influence of solvent composition on ionization efficiency, 50 μM cocaine solutions prepared in 5 different solvent buffers with varying MeOH composition (0%, 25%, 50%, 75%, 95%) and 5% HOAc in water were used. The following conditions were applied for this set of experiments: 13 kV ESI, 10 kV drift voltage, 8.8 kV gate voltage, 500 mL/min $N_2$ drift gas, and 225$^{o}$C oven temperature. The cocaine peak intensity was plotted against flow

rate and the results were shown in Figure 6a. The general trend of increasing response with increasing flow rate was observed in this case as well. Moreover, the ion signals generally appeared to be higher with higher percent of MeOH, with the exception of 95% MeOH. We further calculated ionization efficiency for each data point in Figure 6a and plotted versus flow rate (Figure 6b). Once again we observed the general trend that lower flow rate had better ionization efficiency and the inclusion of organic solvent helped improve ionization efficiency. However, in case of 95% MeOH and 5% HOAc with no $H_2O$, the ion signal and ionization efficiency started to reduce. When comparing the individual IMS spectra with different flow rates and different MeOH compositions, two basic tendencies appeared obvious. First the solvent cluster peaks shifted to higher drift time and decreased in peak intensity with higher flow rate (data not shown in this case, but similar results were shown in Figure 2b). This indicated that the solvent cluster was growing bigger leading to slower mobility and was ionized less efficiently as a result of increased flow rate. Second, with more MeOH and less $H_2O$ included in the solvent, less and weaker solvent cluster peaks were observed (Figure 6c). No solvent peaks were observed in case of 95% MeOH throughout all flow rate variations. All these observations indicated that a certain amount of $H_2O$ is necessary for electrospray ionization to obtain good signal. Water tends to have clusters more easily than MeOH, due probably to the strong hydrogen bonding existent among water molecules. This clustering equilibrium between $H_2O$ and analyte molecules in the gas phase and continuous consumption of analyte-$H_2O$ clusters as a result of complete desolvation of analyte will push the chain reactions forward to create more "naked" analyte ions. This could explain why we observed signal decrease when no $H_2O$ was included in case of

95% MeOH. Thus taken together, we think although the inclusion of certain amount of MeOH is necessary to help better ionization because of its lower surface tension and higher volatility, certain amount of $H_2O$ should also be included to improve ionization.

**Conclusions**

This report has presented the first comprehensive study on effects of flow rate ranging from nanoliter per min to microliter per min for ESI-IMS. The general tendencies to obtain higher signals from higher flow rate appeared to be surprising at first glance. However, after thorough and careful analysis, these opposite effects were attributed to the different ion transmission mechanisms which exist in ESI-IMS and ESI-MS. Our data showed that the general recognition of improved ionization efficiency at reduced flow rate for ESI-MS still held true in case of ESI-IMS despite ion signal loss was observed. Based on our ionization efficiency analysis, now we are able to understand why no substantial sensitivity improvement was observed for nano-ESI-IMS (use of >5mM analyte concentration)[19] as compared to nano-ESI-MS which shows at least two orders of magnitude signal enhancement over conventional ESI-MS.[12] Thus it seems necessary to further improve ionization efficiency for IMS in order to improve overall detection sensitivity. However, the inherent limitation of IMS is attributed to its low duty cycle (<1%). Improving duty cycle is the fundamental solution for improving sensitivity of IMS as demonstrated recently by using Hadamard transform[37] modes of operation which resulted in significant signal gains. On the other hand, it is believed the limiting factor for the sensitivity of ESI-MS is the ion transfer process from atmospheric region to low

vacuum region in MS.[38] However, the efficiency of the ion transfer process for ESI-MS is still unclear because the efficiency of the ionization process is also unclear. Based on our measurements of ionization efficiency, the ion transfer efficiency for ESI-MS can be estimated. The ionization efficiency was assumed near 100% for spraying 1 μM analyte at 20 nL/min (one molecule per charged droplet)[12], while at 2 μL/min, the ionization efficiency is estimated to be ~ 2% according to our data, thus if assuming the signal gain is ~100 fold with the use of nanospray as opposed to conventional spray, the ion transfer efficiency is ~200 times higher in nanospray than that in conventional spray. The overall ion transfer efficiency for conventional spray was reported to be ~0.001% (at detector),[38] thus for nanospray the overall ion transfer efficiency is ~0.2%. Therefore, there is still a room for improving ion transfer efficiency which will ultimately contribute to overall sensitivity improvement for ESI-MS.

Our study also presented the new perspective of IMS that has not been employed previously. The unique constant ion transmission property and atmospheric ion detection property allow IMS to characterize ionization process for different analytes at various flow rates, concentration, and solvent composition as demonstrated in this study. However, IMS certainly can be used to characterize and evaluate any type of atmospheric pressure ion sources besides ESI, such as atmospheric pressure MALDI,[39] desorption electrospray ionization (DESI),[40] or any new modifications made to the ion sources, e.g. multisprayer emitter.[41] We believe the knowledge gained from IMS ionization characterization could further provide the insight into new method development and improvement and eventually help enhance the sensitivity for both IMS and MS techniques.

**Acknowledgements**

This research was supported by the grants from NIH, Grant No. R21 DK070274, the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924, and National Science Foundation, Grant No. DBI-0352451.

**References**

(1) Dole, M.; Mach, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. P.; Alice, M. B. *J Chem Phys* **1968**, *49*, 2240-2249.

(2) Yamashita, M.; Fenn, J. B. *J Phys Chem* **1984**, *88*, 4451-4459.

(3) Yamashita, M.; Fenn, J. B. *J Chem Phys* **1984**, *88*, 4671-4675.

(4) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(5) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal Chem* **1985**, *57*, 675-679.

(6) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *J Chromatogr* **1990**, *516*, 157-165.

(7) Hofstadler, S. A.; Swanek, F. D.; Gale, D. C.; Ewing, A. G.; Smith, R. D. *Anal Chem* **1995**, *67*, 1477-1480.

(8) Gale, D. C.; Smith, R. D. *Rapid Commun Mass Spectrom* **1993**, *7*, 1017-1021.

(9) Andren, P. E.; Emmett, M. R.; Caprioli, R. M. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 867-869.

(10) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199-1202.

(11) Wilm, M.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167-180.

(12) Wilm, M.; Mann, M. *Anal Chem* **1996**, *68*, 1-8.

(13) Juraschek, R.; Dulcks, T.; Karas, M. *J Am Soc Mass Spectrom* **1999**, *10*, 300-308.

(14) Schmidt, A.; Karas, M.; Dulcks, T. *J Am Soc Mass Spectrom* **2003**, *14*, 492-500.

(15) Smith, R. D.; Shen, Y.; Tang, K. *Acc Chem Res* **2004**, *37*, 269-278.

(16) Gieniec, M. L.; Cox, J., Jr.; Teer, D.; Dole, M., Dallas, TX, June 4-9 1972.

(17) Smith, R. D.; Loo, J. A.; Ogorzalek, R. R.; Busman, M.; Udseth, H. R. *Mass Spectrometry Reviews* **1991**, *10*, 359-452.

(18) Wittmer, D.; Chen, Y. H.; Luckenbill, B. K.; Hill, H. H. *Anal Chem* **1994**, *66*, 2348-2355.

(19) Bramwell, C. J.; Colgrave, M. L.; Creaser, C. S.; Dennis, R. *Analyst* **2002**, *127*, 1467-1470.

(20) Colgrave, M. L.; Bramwell, C. J.; Creaser, C. S. *International Journal of Mass Spectrometry* **2003**, *229*, 209-216.

(21) Wu, S.; Tang, X. T.; Siems, W. F.; Bruce, J. E. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, *822*, 98-111.

(22) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal Chem* **2005**, *77*, 311-318.

(23) Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal Chem* **1995**, *67*, 3802-3805.

(24) Chowdhury, S. K.; Chait, B. T. *Anal Chem* **1991**, *63*, 1660-1664.

(25) Wu, C.; Siems, W. F.; Asbury, G. R.; Hill, H. H. *Anal. Chem.* **1998**, *70*, 4929-4938.

(26) Clowers, B. H.; Hill, H. H., Jr. *Anal Chem* **2005**, *77*, 5877-5885.

(27) Kim, T.; Udseth, H. R.; Smith, R. D. *Anal Chem* **2000**, *72*, 5014-5019.

(28) Wu, S.; Zhang, K.; Kaiser, N. K.; Bruce, J. E.; Prior, D. C.; Anderson, G. A. *J Am Soc Mass Spectrom* **2006**.

(29) Siems, W. F.; Wu, C.; Tarver, E. E.; Hill, H. H., Jr.; Larsen, P. R.; McMinn, D. G. *Anal Chem* **1994**, *66*, 4195-4201.

(30) Tang, K.; Shvartsburg, A. A.; Lee, H. N.; Prior, D. C.; Buschbach, M. A.; Li, F.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D. *Anal Chem* **2005**, *77*, 3330-3339.

(31) Sysoev, A.; Adamov, A.; Viidanoja, J.; Ketola, R. A.; Kostiainen, R.; Kotiaho, T. *Rapid Commun Mass Spectrom* **2004**, *18*, 3131-3139.

(32) Gillig, K. J.; Ruotolo, B. T.; Stone, E. G.; Russell, D. H. *International Journal of Mass Spectrometry* **2004**, *239*, 43-49.

(33) Asbury, G. R.; Hill, H. H., Jr. *Int. J. Ion Mob. Spectrom.* **1999**, *2*, 1-8.

(34) Wu, C.; Siems, W. F.; Hill, H. H., Jr. *Anal Chem* **2000**, *72*, 396-403.

(35) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal Chem* **1990**, *62*, 882-899.

(36) Cech, N. B.; Enke, C. G. *Mass Spectrometry Reviews* **2001**, *20*, 362-387.

(37) Clowers, B. H.; Siems, W. F.; Hill, H. H.; Massick, S. M. *Anal Chem* **2006**, *78*, 44-51.

(38) Zook, D. R.; Bruins, A. P. *International Journal of Mass Spectrometry and Ion Processes* **1997**, *162*, 129-147.

(39) Laiko, V. V.; Moyer, S. C.; Cotter, R. J. *Anal Chem* **2000**, *72*, 5239-5243.

(40) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471-473.

(41) Tang, K.; Lin, Y.; Matson, D. W.; Kim, T.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 1658-1663.

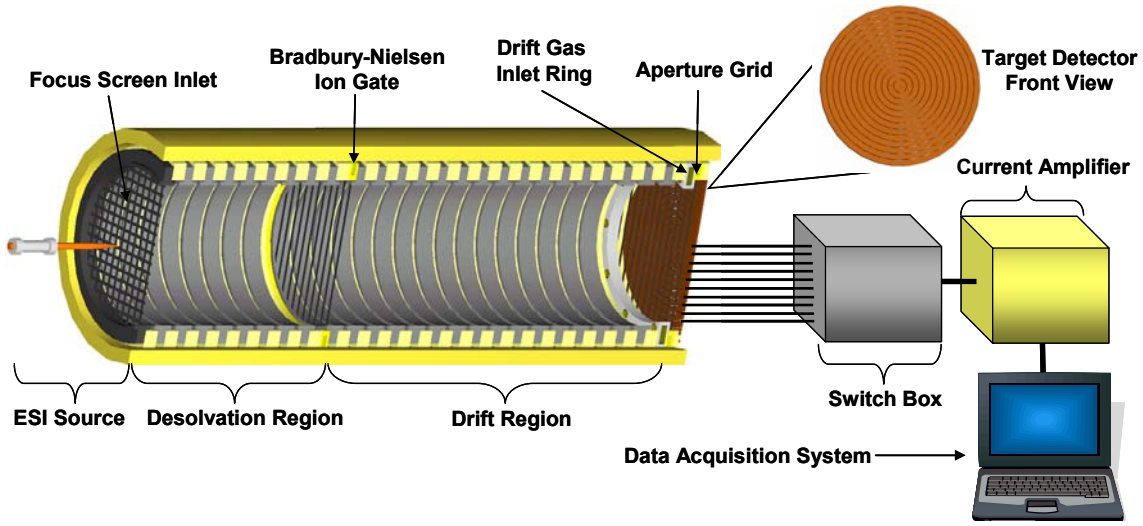



**Figure 1.** Schematic of the atmospheric pressure ESI-IMS instrument with target Faraday collector.

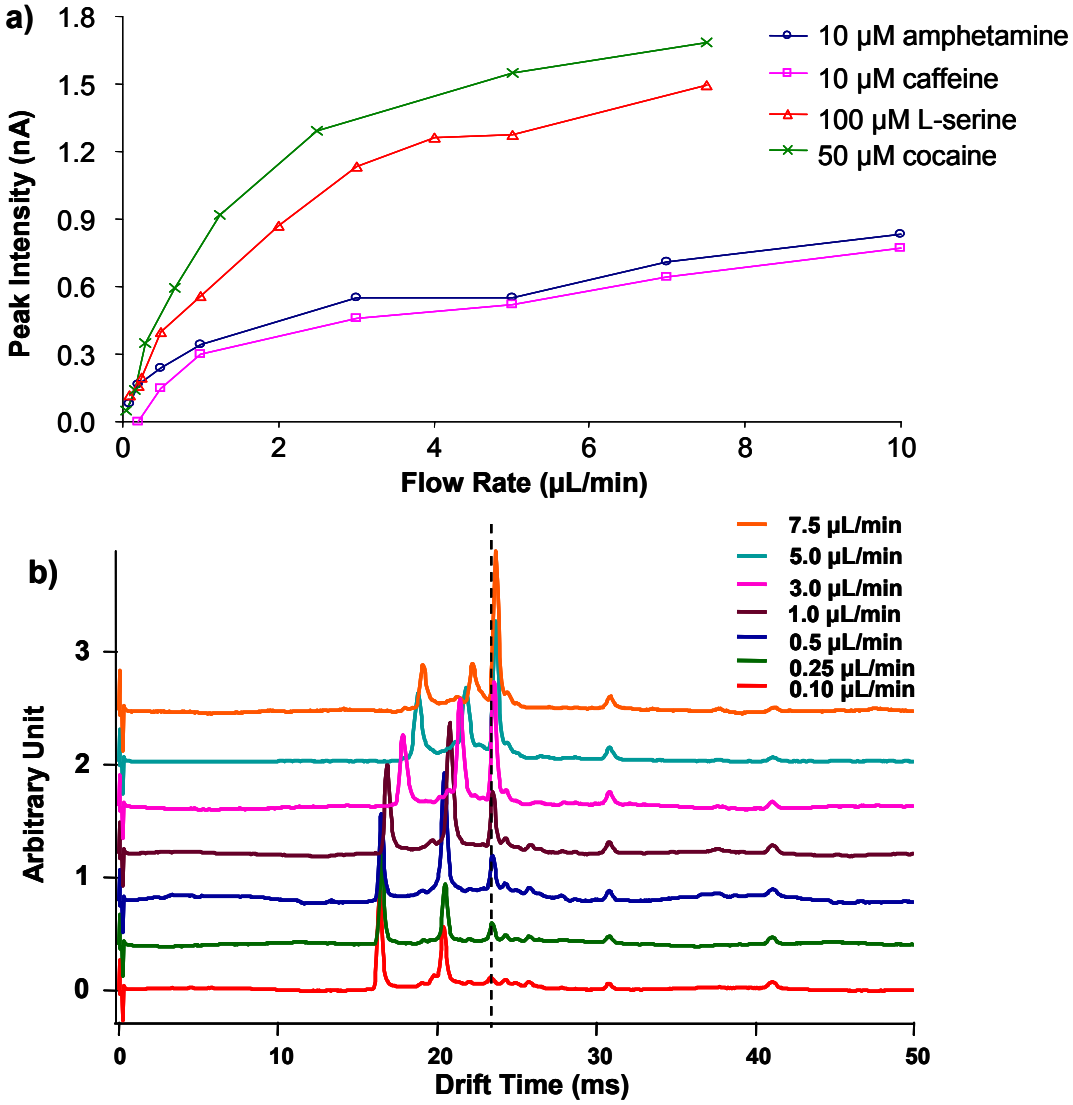



**Figure 2.** Observation of increased ion signal response with increased flow rate in ESI-IMS. a) Plots of ion signal intensity as a function of flow rate for 4 different compounds including 10 µM amphetamine, 10 µM caffeine, 100 µM L-serine, and 50 µM cocaine. b) Overlay of 7 IMS spectra of 100 µM serine at various flow rates.

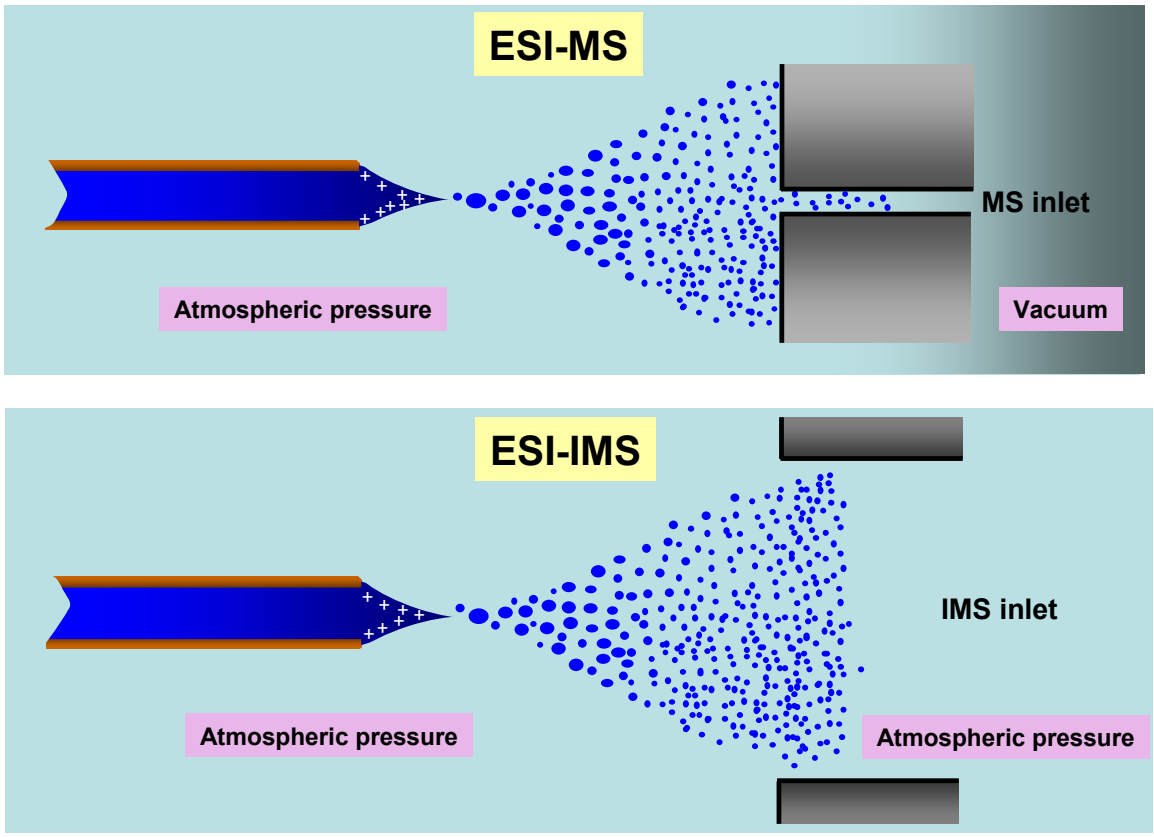



**Figure 3.** Schematic illustration showing different ion transfer process at the interface of ion source and instrument inlet for ESI-MS and ESI-IMS, respectively.

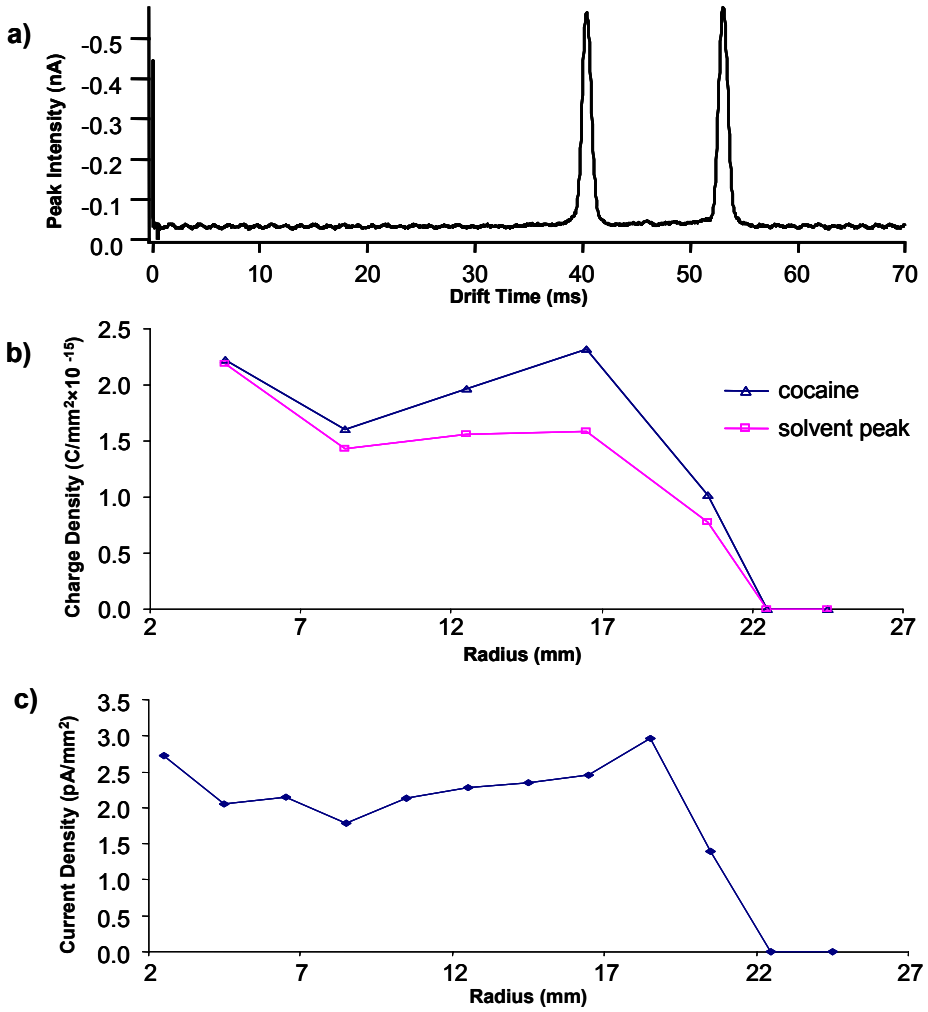



**Figure 4.** Arrival ion beam charge and current distribution study by a target Faraday detector. a) ESI-IMS spectrum of 50 μM cocaine. The spectrum was acquired using combined ring 5 and ring 6 in the target detector at room temperature. The peak at drift time of ~40.4 ms is solvent ion peak and the peak at drift time of ~53.1 ms is cocaine ion peak. b) A plot of charge density (coulomb/mm$^2$ × 10$^{-15}$) as a function of the ring radius (mm) for cocaine ions and solvent ions arrived at the IMS exit. c) A plot of current density (pA/mm$^2$) distribution at the IMS terminus versus the ring radius (mm).

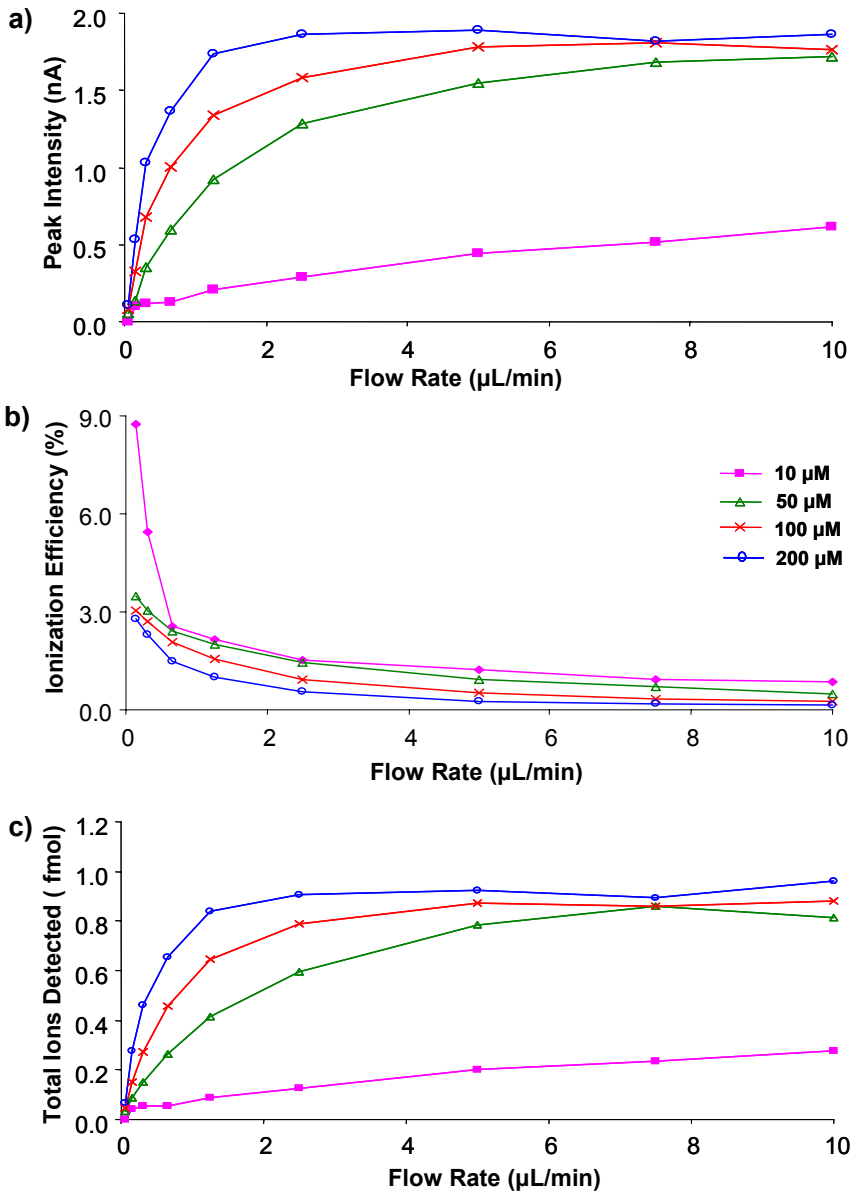



**Figure 5.** Effects of analyte concentration. a) Plots of cocaine ion signal intensity as a function of flow rate at 4 different concentrations. b) Plots of ionization efficiency versus flow rate for cocaine at 4 different concentrations. c) Calculated total amount of ions recorded at the detector (in fmol) as a function of flow rate. All figure legends are shown in b).

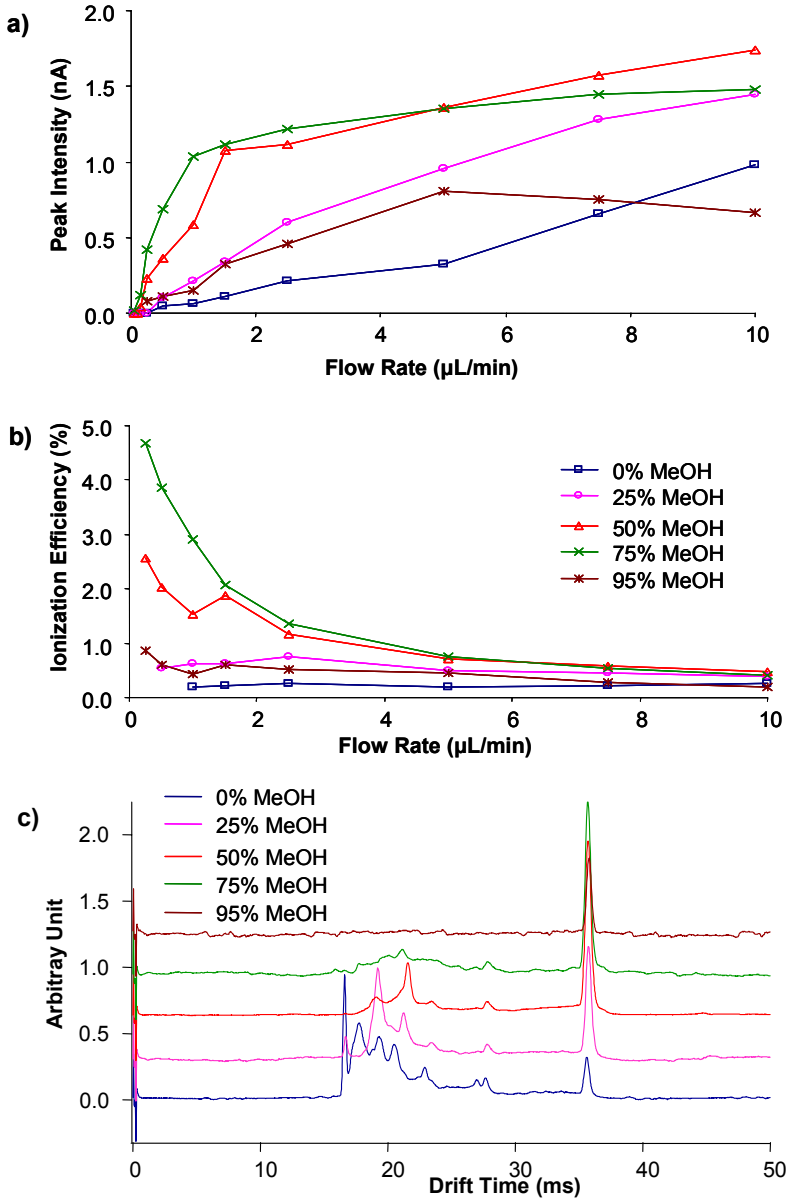



**Figure 6.** Effects of solvent composition. a) Plots of cocaine ion signal intensity as a function of flow rate for 5 different solvents with MeOH percent varying from 0% to 95%. b) Plots of cocaine ionization efficiency versus flow rate for 5 different solvents. Both a) and b) figure legends are shown in b). c) ESI-IMS spectra of 50 µM cocaine solution in 5 different solvents.

## CHAPTER 6

# A Hybrid Instrument That Combines Atmospheric Pressure Ion Mobility Spectrometry with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry


## Abstract

In this paper, we describe our initial experience with construction of a hybrid instrument that combines an atmospheric pressure ion mobility spectrometer (AP-IMS) with a Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) and present the preliminary results obtained from analysis of peptide mixtures. The IMS was coupled to the commercial FTICR-MS instrument through a flared inlet capillary interface without any other modification. Dual-gate ion filtration was adapted to allow concurrent measurement of both mobility and *m/z* values. The feasibility of mobility separation was demonstrated with baseline separation of peptide bradykinin and angiotensin II and their measured reduced mobility constants which were consistent to those previously reported. Furthermore, the unique size-to-charge separation mechanism of IMS that allows isomer separation was explored and demonstrated with the partial separation of two isomeric phosphopeptides.


## Introduction

Since its inception in 1970,[1,2] ion mobility spectrometry (IMS), a rapidly evolving technique, has been combined with a variety of other analytical technologies and employed for detection of a wide range of analytes varying from small organic compounds[3-6] to large biological molecules.[7-13] IMS separates ions in gas phase on the basis of their differential mobility under a uniform weak electric field. In its early developments, IMS is termed as plasma chromatography[1,2] because of its gas phase separation properties as analogy to condensed phase separation techniques such as gas chromatography (GC) and liquid chromatography (LC). As a matter of fact, IMS shares similarities to both chromatography and mass spectrometry (MS) but works in principle different from both, thus it can be coupled to either GC[14] and LC[15] as a detector or MS as a separation device.

While the IMS with a Faraday plate detector offers rapid and sensitive detection in real-time with low-cost and field-deployable benefits, coupling IMS with MS has recently attracted more interest due to additional qualitative and specific fragmentation information that can be gained with MS. One important feature that distinguishes IMS from MS is that IMS separates ions based on their size-to-charge ratio ($\Omega/z$) whereas MS measures ions based on their mass-to-charge ratio ($m/z$). Thus combining IMS with MS produces more comprehensive information than is possible with either technology alone. In addition, as a separation technique, IMS is more advantageous for analysis of biological samples owing to its orders of magnitude faster separation (in ms scale) than condensed-phase chromatography.

IMS can be operated at either atmospheric pressure or low vacuum pressure (1-10 Torr) and both types of IMS have been coupled with various types of mass

spectrometers[16] such as quadrupole MS,[3,17-21] TOF-MS,[7,8,22,23] ion trap MS,[12,24-26] and FTICR-MS.[27] To synchronize the acquisition cycle of IMS and MS, most two-dimensional IMS-MS measurements are achieved by either acquiring IMS spectra for a single *m/z* value or scanning *m/z* window for a selected drift time window, with exception of TOF-MS which is able to measure a wide window of *m/z* simultaneously in μs scale. IMS-TOF-MS combination allows hundreds of MS spectra recorded for elution of a selected IMS peak, which was called a "nested' technique.[7,28] However, given the relatively slow data acquisition nature of most other MS instrument, Clemmer[24] and Creaser[25] have employed a reversed sequence in the hybrid IMS-MS design. In their setup, quadrupole ion trap was used as an ion storage and concentration device to collect ions from continuous ionization source (ESI) and then pulses of ions were injected to drift cell for mobility measurement. Russell[27] also employed a reversed sequence in their hybrid FTICR-MS-IMS instrument with ICR cell placed prior to drift cell. In this design, ICR cell was used as ion-molecule reaction study device and TOF-MS was the actual mass analyzer for IMS. In all injection-type of MS-IMS hybrids, only the low-pressure IMS can be used due to the dramatic pressure difference between the MS region and IMS region. In addition, this type of hybrid construction requires major instrumental modifications thus imposing significant technological challenges, such as differential pumping and aperture alignment.[27] Conversely, atmospheric pressure IMS can be readily coupled with atmospheric ionization source and interfaced in front of an MS with minimal or no modifications to the instrument. Recently Clowers and Hill[26] have implemented a union of an atmospheric pressure IMS with quadrupole ion trap MS without any modifications. In this research, the hybrid instrument of IMS-FTICR-MS

adapts the dual-gate IMS design reported by Clowers et al,[26] but with a flared inlet capillary reported by Bruce[29] to enhance ion transfer efficiency at the interface. This report will highlight the construction of this hybrid instrument and present the preliminary results of separation of standard peptide mixtures and isomeric phosphopeptides.

## Experimental

A schematic diagram of the hybrid nano-ESI-AP-IMS-FTICR-MS instrument is presented in Fig. 1. The basic components of this hybrid instrument consist of (i) nano-ESI source; (ii) AP-IMS drift tube; (iii) flared inlet capillary interface; and (iv) FTICR-MS. Each module is described in detail below.

### Nano-ESI Source

The nanospray emitter was made by etching fused silica capillary (360 μm o.d. × 20 μm i.d., Polymicro Technologies, Phoenix, AZ) in 49% HF solution. The wall thickness of the emitter terminus was thinned down to 0.1 μm or less after 30 min to 1 hr etching. The emitter was then cut to be ~5 cm long. The electrical contact was applied through a stainless steel zero dead volume union (Upchurch Scientific, Oak Harbor, WA) that connected the emitter with the fused silica capillary transfer line (360 μm o.d. × 75 μm i.d., ~30 cm long).

### Atmospheric Pressure Ion Mobility Spectrometer

The atmospheric pressure ion mobility spectrometer (AP-IMS) used for this research was constructed at Washington State University, Pullman, WA. The drift tube was build upon a standard stacked ring configuration by assembling repeating units. Each repeating unit, composed of a conductive stainless steel ring with a dimension of 50 mm (o.d.) × 48 mm (i.d.) × 3 mm (width) and an insulating ceramic ring with a dimension of 60 mm (o.d.) × 50 mm (i.d.) × 4.5 mm (width), is stacked sequentially in an alumina tube. The stainless steel rings, also called guard rings, were connected via a series of half (desolvation region) and one (drift region) megaohm resistors (Caddock Electronics Inc., Riverside, CA). The ceramic ring served to isolate the guard rings from the alumina tube as well as from each other. The current instrument consists of 42 repeating units with a total length of 34 cm, but the length of the drift tube can be easily shortened or lengthened by removing or adding a number of rings. Two Bradbury-Nielsen gate rings divided the drift tube into three regions: 7.5-cm long desolvation region, 26.5-cm long drift region and ~ 1 cm long interface region. A 10 kV voltage was normally applied to the first ring electrode and the last ring electrode voltage was adjusted by variable resistor to be ~200 V referenced to ground. The drift voltage was dropped gradually across the drift tube via the resistor chain to form an electric field of 157 V/cm in the desolvation region and 333 V/cm in the drift region. The lower electric field in the desolvation region allowed solvated ions to spend more time in the heated drift gas to get more efficient desolvation prior to injection to drift region. The two Bradbury-Nielsen gates were made of electrically isolated alternating parallel Alloy 46 wires (76 μm in diameter) (California Fine Wire Co., Grover Beach, CA) spaced 0.65 mm apart. When the potentials, applied on alternating wires, were the same as the reference potential, the gate was “open” to

allow ions to pass through; while the potentials on the adjacent wires were offset ±50 V with the respect to the reference potential, an electric field of ~1500 V/cm was created orthogonal to the drift field and the gate was "closed" to shut off ion transmission.

The drift tube oven was constructed from two pieces of 20-cm long aluminum cylinder with 3 heating cartridges (Heatcon, Seattle, WA) embedded inside each cylinder. The drift tube was normally heated at 150$^{o}$C. All mobility experiments were carried out at atmospheric pressure which typically ranged from 690 to 710 Torr in Pullman, WA. The electronics system used for the IMS tube included a high voltage power supply for the ESI emitter, a high voltage power supply for the drift tube, a two-channel temperature controller for the drift tube oven, and two gate controllers.

**Flared Inlet Capillary Interface**

Bruce recently reported that a 2-5 fold signal gain can be achieved by replacing the original glass capillary with a flared inlet metal capillary in an FTICR-MS instrument.[29] This interface was adapted for the current hybrid instrument without any further modification. In particular, the flared inlet tube used in this research had a dimension of 0.015" i.d. and 0.03" o.d. at the unflared end and was flared with a 45 degree angle from the center line to ~2.2 mm o.d. at other end. The length of the tube is 6". The inlet tube was housed into a 1/4'' stainless steel tube which was interchangeable with the original glass capillary tube.

**FTICR-MS**

The mass spectrometer used for the hybrid instrumentation is a Bruker Apex-Q 7T FTICR mass spectrometer (Bruker Daltonics, Billerica, MA). This instrument consists of a Q-interface: a front hexapole (h1), a mass-selective quadrupole filter (Q), and a second hexapole (h2), and other associated ion transfer optics prior to ICR cell. The h2 can be effectively used as both an ion storage trap and a collision cell. Ions formed by nano-ESI source entered the instrument through the flared inlet capillary and then passed through h1 and Q followed by accumulation in h2 for duration of 1-2 s. Ions were then injected into the ICR cell through an electrostatic ion guide. *Xmass* 7.0.6 software program was used to acquire all mass spectra. All datasets were acquired with 256K points.

**Instrument Operating Modes**

*Continuous mode.* In this mode, both ion gates were open and ions formed at ESI source were continuously transferred to the MS instrument. This mode was normally used for tuning the instrument for optimal signal.

*Selected mobility monitoring.* In this mode of operation, the first ion gate was set at certain pulse width, e.g., 0.5 ms; while the second gate was open only in the user-defined scan window at selected delay times, e.g., 25-35 ms, which means only those ions with drift time between 25 and 35 ms would pass gate 2 and enter the mass spectrometer. This mode of operation helped user to narrow down the mobility scanning window for the analyte ions of interest in a short period of time.

*Mobility scanning dual-gate mode.* Mobility scanning mode sets both gates at certain pulse width (e.g., 0.5 ms) with a delay applied to gate 2. The scanning of this delay time with defined step increment allows acquiring the mobility spectrum for the analyte ions.

The delay time scan window (start of the delay - end of the delay) can be approximately determined using selected mobility monitoring mode as described above. For example, if the scan window is 25-30 ms and the step increment is 0.1 ms, 50 data points will be obtained for the mobility spectrum.

*Acquisition timing sequence.* To allow concurrent mobility and *m/z* measurement for the hybrid IMS-FTICR-MS instrument, IMS dual-gate selection events will be "nested" under each MS acquisition cycle. Fig. 2 shows an example of typical acquisition timing sequence. The rising edge of the h2 TTL pulse triggers the timing sequence of dual-gate pulsing with defined delays. We normally set h2 accumulation time at 2 s. Thus, for a delay time of 40 ms, we are able to repeat IMS injections 50 times for each MS spectrum acquisition. It should be noted that the maximum of IMS injection pulses per MS cycle is governed by the end time of the scan window. To enhance signal intensity, we typically set data averaging of 50-100 for each data point.

**Data Analysis**

IMS control software was programmed by LabVIEW 6.1 (National Instruments, Austin, TX) and communicated with the gate controller and TTL trigger through an interface board (PCI-6601) (National Instruments, Austin, TX) installed in a Dell PC computer. More details were described previously.[26] FTICR-MS was operated in chromatography mode during mobility scanning dual-gate mode and the recorded chromatogram was reconstructed to get selected-ion chromatogram using the program *ICR-2LS* (version 2.30).

## Chemicals and Reagents

Acetonitrile used in this study was HPLC grade and obtained from J. T. Baker (Phillipsburgh, NJ). Formic acid and hydrofluoric acid (HF) were obtained from Sigma (St. Louis, MO). Water was 18-MΩ deionized, prepared by Barnstead Nanopure Water Systems. Standard peptides were purchased from Sigma (St. Louis, MO) and used without further purification. Phosphopeptides were made in the peptide synthesis facility in Washington State University. The sequence and *m/z* information of all the peptides used in this research is presented in Table 1. Typically 100 μM solution prepared in 50% acetonitile and 0.1% formic acid was used for acquiring mobility spectrum.

## Results and Discussion

### Optimizing Ion Signal with Continuous Mode

The interface of most hybrid atmospheric pressure IMS-MS instrument is an unmodified MS inlet, which is either a small pinhole aperture (50-200 um i.d.) or an inlet capillary (0.5 mm i.d. by 25 cm long). Ion loss at the interface of IMS-MS can be more than 99%,[30] which has been a bottleneck for wider application of this technique. Smith[30] recently reported using ion funnel interface at both front ESI-IMS and rear IMS-QTOF region to achieve nearly 100% ion transmission efficiency. However, the rf ion focusing capability of ion funnel works only at relatively low pressure (1- 10 Torr) regime.[31] Therefore, ion funnel focusing is not suitable for atmospheric pressure IMS-MS. However, Bruce[29] recently implemented a flared inlet capillary for FTICR-MS

instrument and demonstrated that the increased-size entrance aperture offered by flared inlet receives more incoming ions thus improves ion transfer efficiency. We adapted this flared inlet interface when coupling AP-IMS with FTICR-MS.

With the IMS operating in the continuous mode, we were able to optimize the position of the flared tube entrance relative to the IMS exit. The optimal signal was normally achieved when the distance between the second gate and flared tube entrance was ~1 cm. With 10 kV applied at the first ring electrode in IMS tube, the voltage at gate 2 was ~ 660 V and the voltage on the last ring electrode located right after gate 2 was ~200 V. The voltage on the flared inlet capillary could be varied from 40 V to 150 V without dramatic changing signal intensity. In addition, we found that ion signals could be further increased ~20% if we applied 800-1000 V to the IMS exit cover, which was made of aluminum and placed right after the last ring electrode.

The flared inlet tube was heated at 200$^{o}$C for all experiments. No significant change on ion signals was observed whether the flared inlet tube was heated or not. We further investigated the influence of the drift tube temperature using a peptide mixture of bradykinin, neurotensin, and angiotensin II. Interestingly, the heating temperature of IMS tube seemed able to govern charge state selection. As shown in Fig. 3, the charge state of all peptides consistently shifted to lower values as the heating temperature rose. At 250$^{o}$C, even the singly charged neurotensin was observed. Li and Cole[32] reported that charge state can be selectively controlled with varying the orifice diameter of the nanospray tip, i.e., smaller tip orifice favors highly charged ion formation. For a given precursor molecule, ions of higher charge state are more prone to fragmentation than those of lower charge state.[33,34] For the current purpose of mobility measurement, we

chose 150°C heating temperature for further experiments since the ions of higher charge state have higher mobility and resolution.

Nano-ESI has become the technique of choice in most biological mass spectrometry laboratories because of its improved ionization efficiency, ion transfer efficiency, and sensitivity. In a separate manuscript that we submitted to the journal *analytical chemistry*, we described a recent discovery that with standalone IMS the ion signal increases with increased flow rate which was controversial to previous observations in nano-ESI-MS. Here, with the atmospheric pressure IMS placed in between nano-ESI and MS, we further investigated the flow rate effects with use of a nanospray tip. Interestingly, the flow rate of 1-2 μL/min seemed to offer the optimal signal with heating temperature at 150°C (Fig. 4). Therefore, the following experiments were performed with flow rate of 1 μL/min.

**Separation of Peptide Mixtures**

With the instrument conditions optimized, we evaluated the utility of this hybrid instrument with analysis of peptide mixtures. The initial attempt was to separate three standard peptide ions, i.e., doubly charged bradykinin (*m/z* 530), doubly charged angiotensin II (*m/z* 523), and triply charged neurotensin (*m/z* 558), with dual-gate filtration experiments. The selected mobility scanning mode was utilized to predict the scan window, which was determined to be 46-51 ms. Then the mobility separation experiments were set up using the mobility scanning dual-gate mode with the following parameters: 2 s h2 accumulation time, 39 dual-gate pulses per h2 accumulation cycle, 100 data averages, 0.2 ms step increment, 12.5 kV nanospray voltage, 10 kV applied to the

first ring electrode, 150°C drift tube heating temperature (measured drift space temperature was 125°C), and air used as drift gas. Fig. 5 shows the reconstructed mobility spectra of these three ions with gate pulse width at 0.5 ms a) and 0.4 ms b). It should be noted the same pulse width was applied to both gates. As shown in Fig. 5, bradykinin was base-line resolved from neurotensin and angiotensin II; neurotensin and angiotensin II were partially resolved; and 0.4 ms pulse width showed slightly better resolution. Further decrease in gate pulse width resulted in no mobility signal. Given the nature of dual-gate mobility experiments, the effective drift length ($L$) and drift voltage ($V$) should be the distance and potential difference between the first gate and the second gate, since once ions exit the second gate, the drift time measurement is completed. We calculated the reduced mobility constants ($K_0$) for these three ions using the equation

$$K_0 = \frac{L^2}{Vt_d}\frac{273}{T}\frac{P}{760}$$

with $L$ ~26.5 cm, $V$ 8.3 kV, $T$ (temperature) 398 K, and $P$ (pressure) 700 Torr. The measured $K_0$ values for all three ions are listed in Table 1. The measured $K_0$ values for doubly charged bradykinin and angiotensin II ions in this study were 1.12 and 1.08, respectively, which were in close agreement with those reported in the literature.[15]

One primary advantage of IMS is its capability of separating isomeric compounds, which is unattainable for FTICR-MS in spite of its superior resolving power. Additional stage of MS experiments is necessary for mass spectrometry to distinguish these isomeric compounds. However, LC/MS/MS experiments will not be able to resolve the differences and make correct assignments of fragment ions if two isomers coelute, which is unfortunately true in most cases because of high structural similarities in

isomers. Thus, we took further effort to look into the isomer separation capability of this hybrid instrument. The isomers we chose for this experiment were two peptides with the same sequence but different phosphorylation sites (see Table 1), one type of post translational modifications with enormous biological interests. The mobility scanning dual-gate mode was applied to acquire the mobility spectra for each peptide individually and then a mixture of these two peptides. The experimental parameters for this set of experiments included 2 s h2 accumulation time, 47 dual-gate pulses per h2 cycle, 100 data averages, 37-42 ms scan window, 0.5 ms gate pulse width, 13 kV nanospray voltage, 11.8 kV first ring electrode voltage, drift temperature 125$^{o}$C, and air as drift gas. The effective drift voltage was 9.7 kV. The resulting mobility spectra for each individual peptide and a mixture of two peptides are presented in Fig. 6, which shows partial separation of the two isomeric phosphopeptides. The calculated $K_0$ values for these two peptides are provided in Table 1. Due to the limited instrument sensitivity, further decrease in gate pulse width resulted in no mobility signal. Therefore, additional improvement of ion transfer efficiency and duty cycle will be essential for achieving base-line separation of these two isomeric phosphopeptides with the use of 0.2 ms or lower gate pulse width.

**Conclusions**

In this report, we presented the first hyphenation of atmospheric pressure IMS with FTICR-MS which allowed two-dimensional mobility and high mass resolution *m/z* measurements without major instrument modification. We found the charge states of peptide ions were tunable with the control of drift tube heating temperature, which could

be beneficial when selection of a specific charge state is needed. Nanospray ionization was employed, however, as contrast to nano-ESI-MS, the optimal signal was obtained with relatively high flow rate, i.e., 1 μL/min. The mobility separation feasibility of the hybrid instrument was demonstrated with standard peptide mixtures. In addition, the capability of separating isomeric phosphopeptides with IMS was demonstrated for the first time. Due to the limited instrument sensitivity, relatively large gate pulse width (0.5 ms) was used. Further improving the instrument sensitivity will allow using smaller gate pulse width thus increase the resolution. Finally, though a flared inlet capillary was employed as interface for signal improvement, further enhancement in ion transfer and duty cycle will be critical for applying this technique to analysis of biological samples with low quantity.

**Acknowledgements**

This research was supported by the grants from NIH, Grant No. R21 DK070274, the Office of Science (BER), National Science Foundation, Grant No. DBI-0352451, and U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

**References**

1. Karasek FW, *Res Dev*, 1970, **21**, 25.
2. Karasek FW, *Res Dev*, 1970, **21**, 34.
3. Karasek FW, Hill HH, Jr., and Kim SH, *J Chromatogr*, 1976, **117**, 327.

4. Lawrence AH and Nanji AA, *Biomed Environ Mass Spectrom*, 1988, **16**, 345.

5. Wu C, Siems WF, and Hill HH, Jr., *Anal Chem*, 2000, **72**, 396.

6. Beegle LW, Kanik I, Matz L, and Hill HH, Jr., *Anal Chem*, 2001, **73**, 3028.

7. Hoaglund CS, Valentine SJ, Sporleder CR, Reilly JP, and Clemmer DE, *Anal Chem*, 1998, **70**, 2236.

8. Henderson SC, Valentine SJ, Counterman AE, and Clemmer DE, *Anal Chem*, 1999, **71**, 291.

9. Matz LM, Asbury GR, and Hill HH, Jr., *Rapid Commun Mass Spectrom*, 2002, **16**, 670.

10. Koomen JM, Ruotolo BT, Gillig KJ, McLean JA, Russell DH, Kang M, Dunbar KR, Fuhrer K, Gonin M, and Schultz JA, *Anal Bioanal Chem*, 2002, **373**, 612.

11. Leavell MD, Gaucher SP, Leary JA, Taraszka JA, and Clemmer DE, *J Am Soc Mass Spectrom*, 2002, **13**, 284.

12. Sowell RA, Koeniger SL, Valentine SJ, Moon MH, and Clemmer DE, *J Am Soc Mass Spectrom*, 2004, **15**, 1341.

13. Liu X, Plasencia M, Ragg S, Valentine SJ, and Clemmer DE, *Brief Funct Genomic Proteomic*, 2004, **3**, 177.

14. Baim MA and Hill HH, *Anal Chem*, 1982, **54**, 38.

15. Chen YH, Hill HH, and Wittmer DP, *J Microcol Sep*, 1994, **6**, 515.

16. Collins DC and Lee ML, *Anal Bioanal Chem*, 2002, **372**, 66.

17. Hill HH, Jr., Siems WF, St Louis RH, and McMinn DG, *Anal Chem*, 1990, **62**, 1201A.

18. Liu Y and Clemmer DE, *Anal Chem*, 1997, **69**, 2504.

19. Dugourd PH, Hudgins RR, Clemmer DE, and Jarrold MF, *Rev Sci Instrum*, 1997, **68**, 1122.

20. Wu C, "Development and application of electrospray ionization-high resolution lon mobility spectrometry-mass spectrometry" *Ph.D. Dissertation,* Washington State University, 1997.

21. Wu C, Siems WF, Asbury GR, and Hill HH, *Anal Chem*, 1998, **70**, 4929.

22. Steiner WE, Clowers BH, Haigh PE, and Hill HH, *Anal Chem*, 2003, **75**, 6068.

23. Srebalus B, Hilderbrand AE, Valentine SJ, and Clemmer DE, *Anal Chem*, 2002, **74**, 26.

24. Liu Y, Valentine SJ, Counterman AE, Hoaglund CS, and Clemmer DE, *Anal Chem*, 1997, **69**, 728A.

25. Creaser CS, Benyezzar M, Griffiths JR, and Stygall JW, *Anal Chem*, 2000, **72**, 2724.

26. Clowers BH and Hill HH, Jr., *Anal Chem*, 2005, **77**, 5877.

27. Bluhm BK, Gillig KJ, and Russell DH, *Rev Sci Instrum*, 2000, **71**, 4078.

28. Hoaglund-Hyzer CS, Lee YJ, Counterman AE, and Clemmer DE, *Anal Chem*, 2002, **74**, 992.

29. Wu S, Zhang K, Kaiser NK, Bruce JE, Prior DC, and Anderson GA, *J Am Soc Mass Spectrom*, 2006, **17**, 772.

30. Tang K, Shvartsburg AA, Lee HN, Prior DC, Buschbach MA, Li F, Tolmachev AV, Anderson GA, and Smith RD, *Anal Chem*, 2005, **77**, 3330.

31. Shaffer SA, Prior DC, Anderson GA, Udseth HR, and Smith RD, *Anal Chem*, 1998, **70**, 4111.

32. Li Y and Cole RB, *Anal Chem*, 2003, **75**, 5739.

33. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, and Udseth HR, *Anal Chem*, 1990, **62**, 882.

34. McLuckey SA, Glish GL, and Van Berkel GJ, *Anal Chem*, 1991, **63**, 1971.

**Table 1.** Information of the five peptides used in this research

| **Peptide Name** | **Peptide Sequence** | ***m/z*** **$(M+H)^+$** | ***m/z*** **$(M+2H)^{2+}$** | ***m/z*** **$(M+3H)^{3+}$** | **$K_0$ ($cm^2$/V s)[d] (Measured)** | **$K_0$ ($cm^2$/V s)[e] (literature)** |
|---|---|---|---|---|---|---|
| Bradykinin | RPPGFSPFR | 1060.57 | <u>530.79</u> | 354.19 | 1.12 | 1.14 |
| Neurotensin[a] | pELYENKPRRPYIL | 1672.92 | 836.96 | <u>558.31</u> | 1.09 | - |
| Angiotensin II | DRVYIHPF | 1046.54 | <u>523.77</u> | 262.39 | 1.08 | 1.10 |
| S3P[b] | YLpSRSGR | 918.42 | <u>459.71</u> | 306.81 | 1.16 | - |
| S5P[c] | YLSRpSGR | 918.42 | <u>459.71</u> | 306.81 | 1.18 | - |

[a] The N-terminus of the neurotensin sequence has a pyroglutamate modification. [b] S3P peptide has a phosphorylation modification at serine 3. [c] S5P peptide has a phosphorylation modification at serine 5. [d] The measured $K_0$ is for the ions with *m/z* underlined. [d] The literature $K_0$ was for the ions with *m/z* underlined. "-" means no $K_0$ values can be found in previous literature.

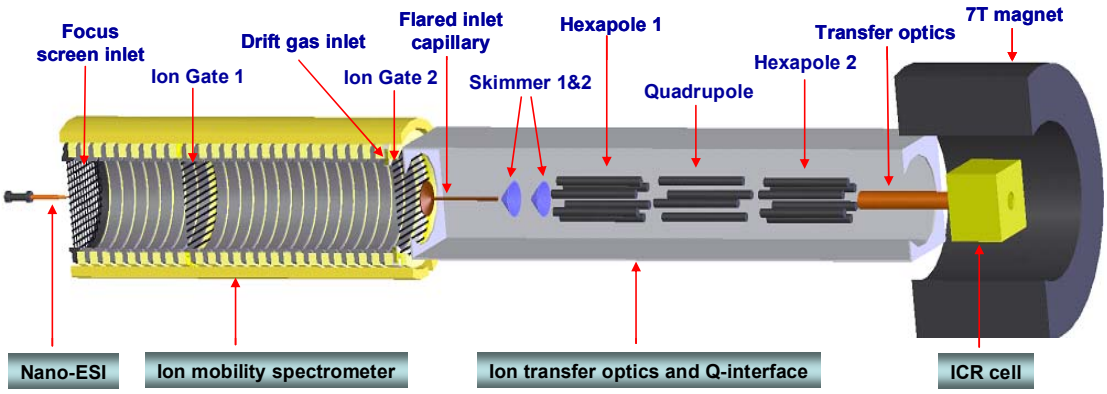



**Figure 1.** Schematic of the electrospray ionization atmospheric ion mobility spectrometer hyphenated with Fourier transform ion cyclotron resonance mass spectrometer (ESI-AP-IMS-FTICR-MS) instrument.

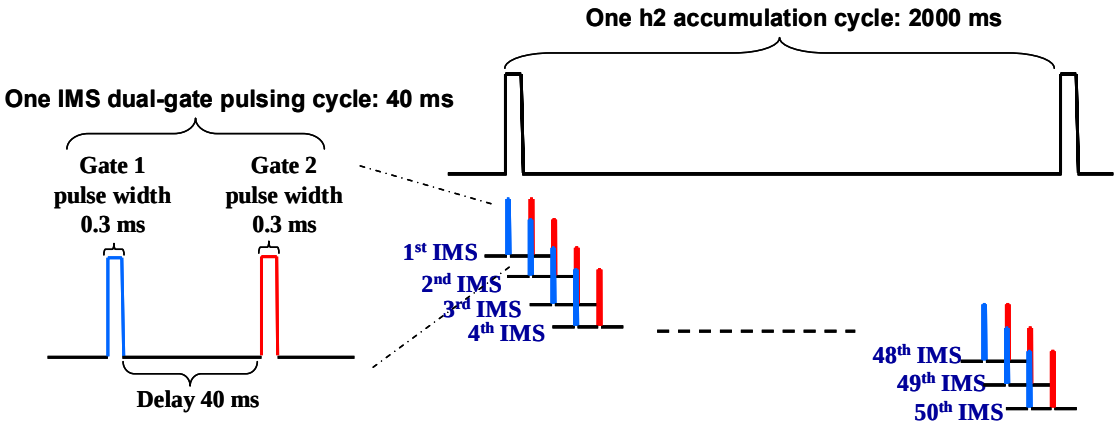



**Figure 2.** Illustration of timing sequence for data acquisition. Each MS injection cycle embraces a number of repeated IMS dual gate pulsing cycles. For example, if h2 accumulation time is 2000 ms, gate 2 delay is 40 ms, then the dual-gate filtration events can be repeated 50 times. The maximum number of IMS pulsing repeats is determined by the end of the delay time in the mobility scanning dual-gate mode.

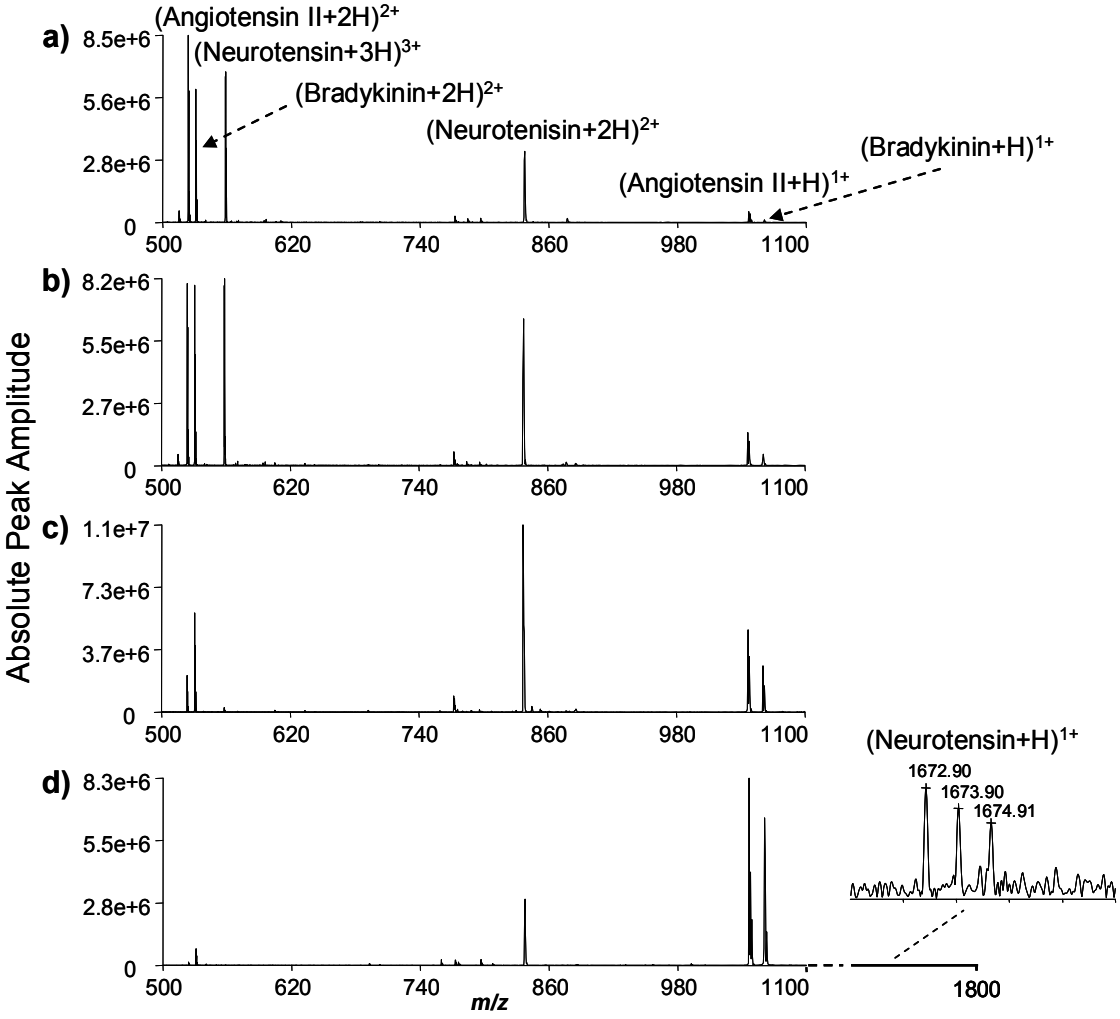



**Figure 3.** Charge state selection observed with varying the drift tube heating temperature. Data were collected using continuous mode ESI-AP-IMS-FTICR-MS analysis of a peptide mixture containing angiotensis II, bradykinin, and neurotensin with the drift tube heating temperature set at a) 100°C, b) 150°C, c) 200°C, and d) 250°C. Lower charge state of the ions is favored at higher heating temperature for all three peptides. The inset figure in d) shows that at 250°C, a weak signal of singly charged neurotensin was observed.

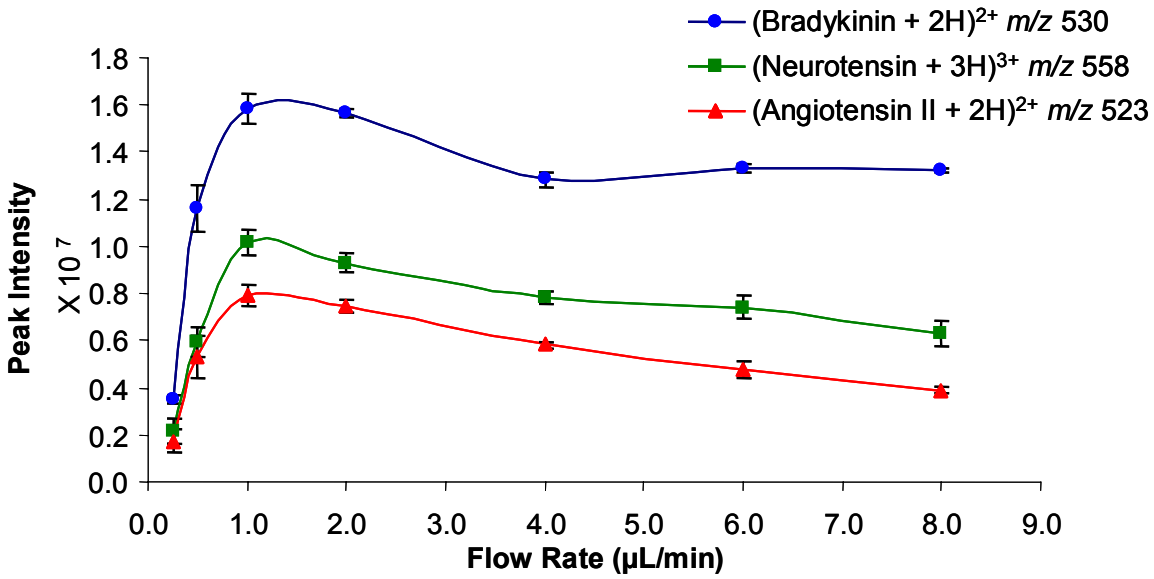



**Figure 4.** Measured peak intensities for doubly-charged bradykinin at *m/z* 530, triply-charged neurotensin at *m/z* 558, and doubly charged angiotensin II at *m/z* 523 as a function of infusion flow rates. Each FTICR-MS spectra were collected with the use of continuous IMS mode, 256 K data points and 20 data averages. Each data point is the average of 3 measurements of peak intensities extracted from ICR spectra.

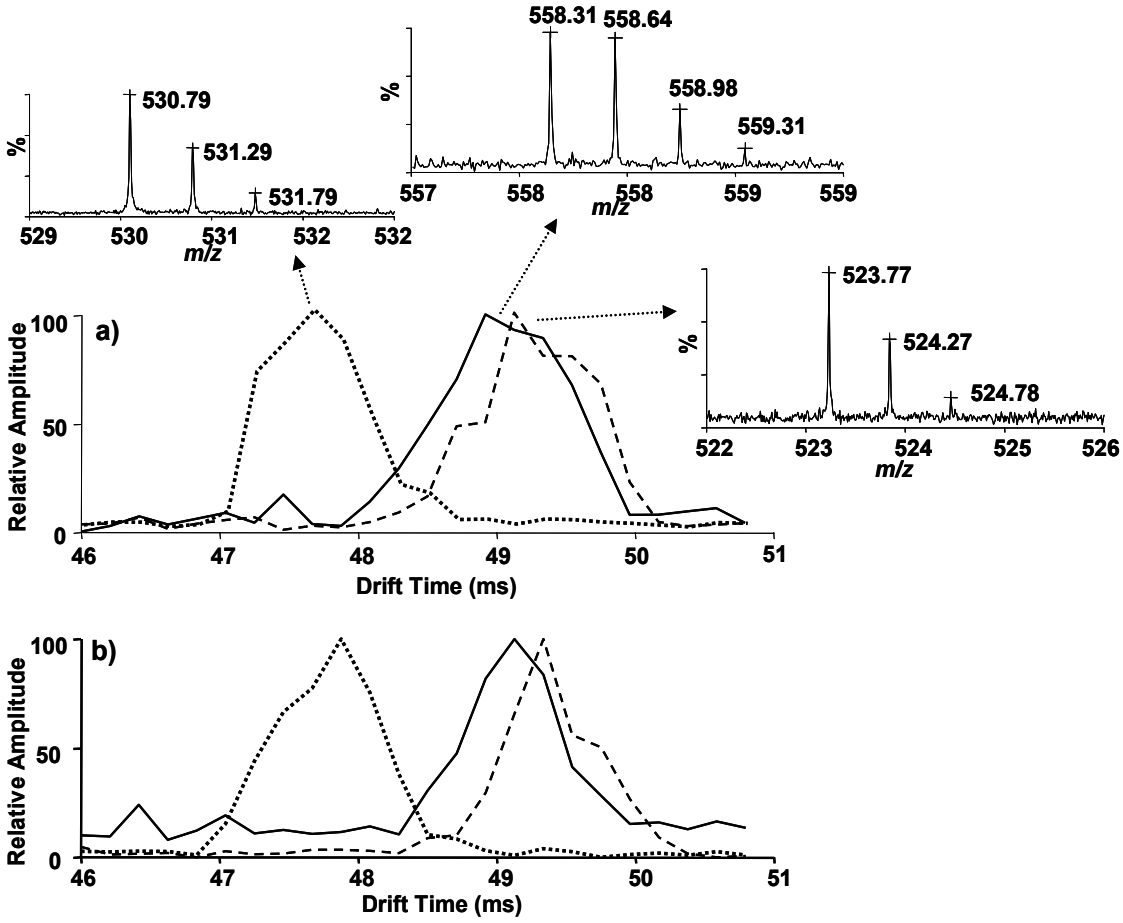



**Figure 5.** ESI-AP-IMS-FTICR-MS analysis of standard peptide mixtures using mobility scanning dual-gate mode with a) 0.5 ms and b) 0.4 ms gate pulse width. Dotted line represents the reconstructed selected ion mobility spectrum of $(\text{bradykinin}+2\text{H})^{2+}$ at *m/z* 530.8; solid line represents the reconstructed mobility spectrum of $(\text{neurotensin}+3\text{H})^{3+}$ at *m/z* 558.7; and dashed line represents the reconstructed mobility spectrum of $(\text{angiotensin II}+2\text{H})^{2+}$ at *m/z* 523.8. Inset figures show examples of dual-gate filtered FTICR-MS spectra for bradykinin, neurotensin, and angiotensin II.

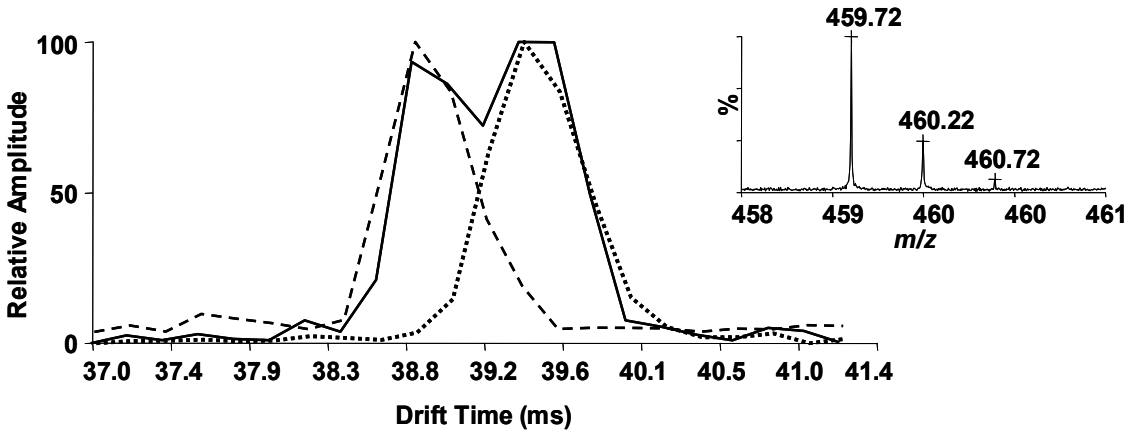



**Figure 6.** ESI-AP-IMS-FTICR-MS analysis of isomeric phosphopeptides with gate pulse width at 0.5 ms. Shown is the reconstructed selected ion mobility spectrum of doubly charged ion at *m/z* 459.7. Dashed line represents the mobility spectrum collected with S3P alone; dotted line represents the spectrum collected with S5P alone; and solid line represents the mobility spectrum collected with a mixture of S3P and S5P. Inset figure shows an example FTICR-MS spectrum of doubly charged peptide ion at *m/z* 459.7 after dual-gate selection.

# CHAPTER 7

# CONCLUSIONS

## Overall Conclusions

In an effort to study biological systems at global level, we have developed new analytical strategies and technologies to address the challenges associated with complexity and dynamics of systems biology research. We have designed and developed novel chemical cross-linkers strategies, called protein interaction reporters (PIRs), with MS cleavable bonds, mass-encoded reporter tag, and affinity tags to facilitate identification of protein-protein interactions. The utility of PIR strategy was first validated with a known model protein complex system, ribonuclease S and our study demonstrated that complexity of cross-linking approach can be significantly reduced with aid of reporter ions and the unique relationships existent in PIR-labeled peptides and their products. Furthermore we applied PIR strategy to study protein-protein interactions in *S. oneidensis* MR-1 bacterial cells at proteome-wide level with additional two-stage mass spectrometric strategies. Our results presented the first demonstration of identification of protein-protein interactions as well as the sites of interactions in living cells, which set important landmark step for future large-scale mapping protein-protein interaction networks and interaction topologies at systems-level. These PIR cross-linkers were also employed to profile membrane proteins and cell surface proteins at proteome-wide level. We demonstrated for the first time that hydrophobic affinity chemical probes can preferentially label

membrane and membrane-associated proteins. Among the 40% cell envelope proteins we identified from *S. oneidensis* cells, many are known as important surface proteins participating in electron transfer and metal reduction process. In addition, we have attempted to develop instrumental strategies for the ultimate goal of systems biology. Our study demonstrated for the first time that standalone IMS instruments was capable of evaluating ionization efficiency for various experimental conditions and presented the first empirical data that ionization efficiency increases as flow rate of electrospray decreases. Furthermore, we implemented the first hybrid instrument that combined atmospheric pressure ion mobility spectrometry with FTICR-MS. We demonstrated the mobility separation of isomeric phosphopeptides for the first time with this hybrid instruments. The biological systems is far from being fully understood, however, we have made important first step toward our goals and our initial efforts in developing new analytical tools for systems biology have presented promising preliminary results.