Identification of Protein-Protein Interactions and Topologies in Living Cells with Chemical Cross-linking and Mass Spectrometry

Haizhen Zhang§, Xiaoting Tang§, Gerhard R. Munske1, Nikola Tolic2, Gordon A. Anderson2, and James E. Bruce*

1Department of Chemistry, Washington State University, Pullman, Washington 99164
2Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

§Equally contributing authors

*Corresponding author: James E. Bruce

Current Address:
Department of Genome Sciences
University of Washington
Seattle, Washington 98195

(E-mail): jimbruce@u.washington.edu
(Phone): 509-592-7742
(Fax): 206-685-7301

Keywords
Chemical cross-linking, protein-protein interactions, protein topology, protein interaction reporter (PIR), membrane proteins, accurate mass, protein identification, Shewanella oneidensis MR-1.
Abstract

We present results from a novel strategy that enables concurrent identification of protein-protein interactions and topologies in living cells without specific antibodies or genetic manipulations for immuno/affinity purifications. The strategy consists of: (i) chemical cross-linking reaction: intact cell labeling with a novel class of chemical cross-linkers, 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. The primary advantage of the PIR approach and distinction from current technology is that protein interactions together with topologies are detected in native biological systems by stabilizing protein complexes with new covalent bonds while the proteins are present in the original cellular environment. Thus, weak or transient interactions or interactions that require properly folded, localized, or membrane-bound proteins can be labeled and identified through the PIR approach. This strategy was applied to *S. oneidensis* bacterial cells and initial studies resulted in identification of a set of protein-protein interactions and their contact/binding regions. Furthermore, most identified interactions involved membrane proteins, suggesting the PIR approach is particularly suited for studies of membrane protein-protein interactions, an area under-represented with current widely-used approaches.
Introduction

An essential component of the goal to elucidate global biological function is the determination of protein interaction networks. Current approaches for mapping protein-protein interactions include yeast two-hybrid system (YTH) (1), affinity purification procedures based on immunoprecipitation (IP) or single- (e.g., flag-tag (2)) or double- affinity tag (e.g., TAP-tag (3, 4)) followed by protein identification with mass spectrometry, protein microarray technology (5, 6), and computational prediction methods (7, 8). Although all these approaches demonstrate great promise in mapping protein-protein interactions on a proteome-wide level, the resulting large-scale data sets are often associated with high rates of false negatives and false positives (>50%) and poor overlap of data sets among different approaches used for the same system are often observed (9-11). Such observations suggest that no single method is flawless and comprehensive. The strengths and weaknesses of each method have been thoroughly reviewed (12-15). For example, traditional IP-based affinity purification methods require a specific antibody for every protein of interest which is a hindrance for widespread, large-scale application. Tag-based methods overcome this limitation by fusing the bait protein genetically with an affinity tag which is applicable to all proteins. One of the most successful tag-based methods is TAP technology, which fuses two affinity tags to the bait protein and nonspecific binding is significantly reduced with two sequential purification steps (3, 4). While tag-based methods allow bait proteins to be expressed in vivo and interact with native physiological partners, recent studies showed that tagging can also cause overexpression of the bait protein that can result in association with chaperones and improper intercellular localization (16, 17). In addition, tagging one bait protein at a time for large-scale studies can be tedious and costly. Another issue worth noting is that all affinity-based methods require cell lysis prior to purification of the associated complex of the bait protein. During cell lysis, the native cellular system is disturbed and the bait protein is present in the lysis buffer which is very different from the intracellular milieu. As described recently by Berggard et al. (13), the fact that the affinity
between interacting proteins may be substantially different *in vivo* as compared to *in vitro* has not been carefully considered in the literature. We reported the first such comparison of mapping targeted protein interactions using both, intact cells and cell lysates and our results illustrated significantly different protein interaction data, highlighting the importance of identification of protein-protein interactions under native conditions (18).

Another challenge that affinity-based methods face is related to the inherent difficulty involved in maintaining the integrity of native protein complexes while removing the nonspecific bindings during washing steps. Most transient and weak protein-protein interactions may not survive through harsh washing steps, which is particularly true for interactions involving membrane proteins. For example, a high level of detergent normally required for maintaining the solubility of membrane proteins can also disturb non-covalent associations (15, 19). Chemical cross-linking can be used to stabilize and freeze protein-protein interactions by forming covalent bonds with proteins while proteins are present in the native cellular environment (15, 20, 21). The cross-linked protein complexes can remain intact during cell lysis and stringent washes. Therefore, cross-linking strategies have been successfully combined with affinity-based methods for studies in protein-protein interactions. *In vivo* cross-linking applications coupled with IP (22-27) and TAP-tag (28, 29) techniques have been extensively reported and reviewed (15, 20, 21, 30-32).

Another important feature of chemical cross-linking methods is the potential for mapping topology of proteins and protein complexes (see reviews (15, 20, 21, 30)). If cross-linked residues/peptides can be identified, this information can yield clues about the contact/binding interfaces among protein complexes. While *in vivo* cross-linking coupled with affinity purification can readily allow identification of interacting protein partners for a particular protein of interest with the detection of higher bands in gels or Western blot images, identification of cross-linked peptides/residues is not trivial even for purified protein complexes available in large quantity. Improved cross-linkers such as chemically-cleavable cross-linkers (DSP and DTSP)
(33), isotope-encoded cross-linkers (34), and cross-linkers with affinity tags (35, 36) have been employed to aid in assignment of cross-linked residues in mapping topological structures of purified proteins and protein complexes. Factors that have slowed or inhibited the advancement of cross-linking strategies for large-scale *in vivo* applications stem from the complex nature inherent with the cross-linking approach. Cross-linking reactions are often carried out with large excessive amount of cross-linkers to increase reaction rate and product yield, thus resulting in highly complex mixtures. Proteolysis performed prior to MS analysis further increases the degree of complexity and the desired cross-linked peptides are only a tiny fraction (<0.1%) of total theoretical peptide combinations (37). In addition, 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 peptide/protein identification. The Aebersold group recently reported breakthrough results illustrating identification of cross-linked peptides directly from complex *E. coli* cell lysates with advanced informatics software development. Their efforts resulted in identification of 3 interactions from homo-oligomers and 2 interactions from hetero-oligomers (37), which suggests strongly growing interest in cross-linking methods for protein interaction identification and as yet, untapped potential in this challenging research area.

To help overcome the limitations of current cross-linking methods, we have pursued a chemistry-based strategy by developing a novel class of cross-linkers called Protein Interaction Reporters (PIRs). The spacer chain 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. 1a). Affinity-based purification is employed to enrich PIR-labeled native protein complexes; however, no genetic manipulation is required as used in Flag-tag and TAP-tag methods nor is any antibody required for each protein of interest, since the affinity moiety is part of the PIR cross-linkers. This feature conceivably allows further extension of the PIR strategy to native biological systems other than cells, such as tissues and biological
fluids. When combined with a novel two-stage mass spectrometric analysis strategy, the PIR approach allows large-scale mapping of protein-protein interactions and their sites of interaction in native biological systems. Here, we report the initial application of the PIR strategy to the microbial system, S. oneidensis MR-1 cells and the results of identification of a set of protein-protein interactions including information on contact/binding interfaces from living cells. These results represent the first such successful application of a chemical cross-linking strategy to map interactions and interaction sites from native living cells in a large-scale, non-targeted way. The tunable chemistry of PIR strategy will ultimately allow multiplexing an array of PIR cross-linkers with different structures and reporters for more comprehensive mapping of protein-protein interactions from native biological systems.

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 buffers and solutions was 18-MΩ deionized water produced with a Barnstead Nanopure Water System.

PIR synthesis. The PIR cross-linkers were synthesized using a 431A peptide synthesizer (Applied Biosystems, Foster City, CA). 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 (for Rink-based PIR) or proline and aspartic acid (for DP-based PIR) were coupled to lysine in forms of Fmoc-amino acid. Carboxyl groups were then introduced by coupling the primary amines of the Rink groups or aspartic acid with succinyl
anhydride. Subsequently, the two carboxyl groups were activated by forming the esters with N-hydroxysuccinimide (NHS). The crude 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. The crude product was purified using reversed phase HPLC and the final product had a purity of about 90%.

**Two-stage analysis.** To fully utilize the advanced features of PIRs, a two-stage mass spectrometric approach was developed (Fig. 2). Stage 1 analysis was carried out with avidin capture of *in vivo* labeled proteins, followed by digestion and shotgun 2D-LC/MS/MS for protein identification, as described previously(38).

Stage 2 analysis was carried out by labeling 0.2 gram cell pellets suspended in 4 mL PBS buffer (150 mM sodium phosphate, 100 mM NaCl, pH 7.5), and the cross-linkers were added to the suspended cell pellets to produce a final concentration of 1 mM. The cross-linking reaction was carried out at 4°C for 1 h and quenched by 1 M ammonium bicarbonate. Extensive washing steps with PBS after cross-linking reactions were used to eliminate most non-specific contamination as reported previously (38). Then cells were lysed in 2% NP-40 PBS solution by sonication for 2 min. The cell lysates were centrifuged at 15,000 × g at 4°C for 45 min. The pellets were discarded and supernatants were collected in four clean tubes with a total amount of approximately 4 mL. Protein concentration of the supernatants was determined to be approximately 5 mg/mL by a Bradford Assay (Bio-Rad, Hercules, CA). The PIR labeled proteins were enriched by adding a 50 μL slurry of monomeric avidin beads (Pierce, Rockford, IL) to each tube and the reaction was incubated for 2 h with gentle mixing at room temperature. Then, the avidin beads were washed three times with 500 μL 100 mM NH₄HCO₃ (pH 7.8). After the last wash, the avidin beads were suspended in 100 μL solution of 8M urea and 100 mM NH₄HCO₃ (pH 7.8). Both enriched PIR-labeled proteins and avidin were reduced directly on the
beads with 10 mM DTT at 56 °C for 1 h and alkylated with 20 mM iodoacetamide for 1 h at room temperature. After reduction and alkylation, additional 10 mM DTT was added to neutralize the unreacted iodoacetamide for another 1 h at room temperature. The bead suspension was then diluted 4 times with 100 mM NH₄HCO₃ (pH 7.8) buffer and 10 µg trypsin (Promega, Madison, WI) was added for overnight digestion at 37°C. After digestion, 1 mM phenylmethanesulfonyl fluoride (PMSF) was added to quench the trypsin activity. A 50 µL slurry of monomeric avidin beads were added to the digestion mixture and gently mixed for 2 h at room temperature. After beads were washed three times with 100 mM NH₄HCO₃ (pH 7.8), PIR labeled peptides were eluted from avidin beads with 200 µL 70% acetonitrile and 0.5% TFA solution four times. Finally the volume of the combined eluent was reduced to ~30 µL using speed vacuum and ~10 µL was used for multiplexed LC/FTICR-MS analysis.

**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, the reversed phase nano-column and LC gradient methods were the same as those used with the ion trap mass spectrometer as described previously(39). Ions formed by nano-ESI source entered the instrument through a glass capillary and then passed through a hexapole followed by a quadrupole and then a second hexapole, which was used for ion accumulation and PIR activation. After the second hexapole, ions were guided through a series of ion transfer optics and enter the ICR cell where ions were 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, -22 v, which efficiently fragmented the labile bonds in PIR molecules but not peptide bonds, was applied during alternating scans throughout the LC elution. External calibration of the instrument was performed to a mass accuracy of less than 1 ppm with direct infusion of BSA
tryptic peptides immediately prior to 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* (40). Lock-mass calibration using reporter ions was performed to correct all the neutral masses produced in .PEK file to further improve MMA. All possible tryptic peptides of restricted protein database were produced by *X-links* with the constraint of at least one internal lysine in the sequence (40). 10 ppm error tolerance was used in *X-links* searches for both cross-link type identifications (using the relationship between the precursor ion in the MS scan and the released peptide and reporter ions in the subsequent multiplexed MSMS scan) and cross-linked peptide sequence identifications (using the measured accurate mass of the released peptide ions in the multiplexed MSMS scans). MS$^2$ and MS$^3$ validation experiments were performed using a quadrupole ion trap mass spectrometer (Esquire HCT, Bruker Daltonics, Billerica, MA) as described previously (39). GPMAW (version 6.21) was used to calculate MS/MS fragmentation. The whole genome sequence of *S. oneidensis* MR-1 was completed in 2001 and the database containing the entire genome (4,854 ORFs) was downloaded in July 2006 from The Institute for Genomic Research (TIGR) (http://www.tigr.org/).

**RESULTS**

**PIR Structure**

The PIR strategy is enabled by incorporating two labile bonds in the spacer chain of the cross-linker that can be cleaved with high specificity in the mass spectrometer resulting in release of a mass-encoded reporter ion (**Fig. 1a**). 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 released in the subsequent high-energy scan with accurate mass and/or additional MSMS analysis. The conceptual mathematical relationships that exist between PIR-labeled precursors, released peptides, and reporters are illustrated in Fig. 1b and these relationships can be used to facilitate differentiation of cross-link types and identification of cross-linked peptides with informatics software tools. The tunable chemistry of PIR structure allows cross-linking a variety of different proteins by incorporation of different labile bonds, affinity tags, and reactive groups. For the current study, we employed a biotin group as the affinity tag, N-hydroxysuccinimide (NHS) esters as the reactive group, and the Rink group (Fig. 1c) and Asp-Pro (DP) bond (Fig. 1d) as low-energy MSMS cleavable bonds.

The relatively large size of PIR cross-linkers has been a concern for obtaining useful structural constraint information (21). However, our previous studies using the model RNase S noncovalent complex showed that the distance (14Å) between two cross-linked residues was much smaller than the maximum length (43Å) of PIR (39). Furthermore, Rinner et al. recently reported identification of the same inter-cross-link in RNase A using the smaller cross-linker DSS (37). These results suggest that in solution, the PIR is constrained to give rise to a shorter distance between the two reactive groups than the fully extended length, which agrees with the simulated estimates of the realistic lengths of 32 commercial cross-linkers in solution reported by Green et al. (41) Finally, it is also worth considering that within the cellular environment, many potential labeling sites that would allow identification of interactions span a wide range of distances. Therefore, flexibility in the cross-linker structure that can allow labeling sites over a range of distances is most critical to enable identification of larger numbers of interactions. Once identified as an interaction that exists in cells, these complexes can be studied for more detailed structural information using a wide variety of conventional molecular biology, analytical techniques, and computational methods.

The NHS was employed as reactive group in PIR cross-linkers. The NHS is the most
frequently-used group in protein labeling reagents (21) such as iTRAQ (42), one of the most widely-used protein quantitation reagents, and chemical cross-linkers such as DSS, DSP, BS$_3$, etc. (Pierce). The reason for this is multiple-fold. First, the most common targets for labeling proteins are the primary amine groups, which are present on the large majority of proteins due to the high occurrence of lysine side chain ε-amines. Second, NHS forms stable amide linkage to the primary amines in proteins at physiological condition (pH 7.0-7.5), which is important for in vivo labeling. Finally, the NHS ester reaction has very rapid reaction rates with primary amines (43). Our studies (38) and those of others (23, 25) showed that a 5-min reaction time was sufficient for completion of labeling. The fast reaction rate is critical for in vivo experiments since it can quickly “snap-shot” or “freeze” the protein complexes with stable covalent bonds before living cells are extensively perturbed.

Two-stage analysis

To fully utilize the advanced features of the PIR strategy, a novel two-stage mass spectrometric approach was developed (Fig. 2). Intact cells were labeled with PIR compounds under near-physiological conditions. After labeling, proteins were extracted and PIR-labeled proteins were enriched by avidin capture. Then the enriched PIR-labeled proteins were divided into two parts for the subsequent stage 1 and stage 2 analyses. Stage 1 analysis was carried out with digestion of PIR-labeled proteins followed by protein identification with shotgun 2D-LC/MS/MS (44). The goal of stage 1 analysis is to construct a candidate protein pool as complete as possible, which provides a “lookup” table for stage 2 analyses. We previously employed the Rink-based PIR compound as a general biotinylation reagent for labeling intact cells (38) and the ~400 identified labeled proteins were used to constitute the restricted database for the current study. Stage 2 analysis was performed by further digestion of enriched PIR-labeled proteins on avidin beads, followed by another affinity capture of PIR-labeled peptides. These peptides were 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 following PIR activation scan. Finally we employed an informatics software tool, X-links (40), to identify cross-links in two steps: step 1 involved characterization of the precursor-product relationships and cross-link types based on the equations depicted in Fig. 1b; and step 2 involved assignment of cross-linked peptide sequences/proteins of origin with accurate peptide mass-based protein identification strategy.

Fig. 3 shows an example dataset from a multiplexed LC/FTICR-MS analysis of stage 2 sample, including the base peak chromatogram (BPC) and the extracted ion chromatogram (EIC) for the expected reporter ion. The inset in Fig. 3a shows the overlaid EICs for an identified PIR-labeled precursor ion (red), the released intact peptide ion (blue), and the reporter ion (yellow). The observed complementarities of the peak intensity patterns and the high degree of overlap of the LC elution profiles between the red and blue traces strongly suggest a precursor-product relationship. This information can supplement multiplexed LC/MS analysis to help define precursor-product relationships and cross-link types and can be used to further constrain our search results.

Since mass measurement accuracy (MMA) is the most important factor in both identification of cross-link types and cross-linked peptide sequences, we routinely calibrated the FTICR-MS instrument externally to ~1 ppm MMA with BSA tryptic peptide mixture immediately prior to LC/FTICR-MS analysis. However, space charge effects and variation of ion populations(45) can severely reduce MMA for LC/FTICR-MS experiments. Introduction of a standard mass used as internal lock mass calibrant via dual spray has been employed to improve MMA for LC/FTICR-MS analysis(46). The PIR strategy provides an inherent lock mass, i.e., the known mass of reporter ion. Thus, we calibrated LC/FTICR-MS dataset using reporter ion as lock mass and MMA for reporter ions was corrected to ~3 ppm (Supplementary Fig. 1). However, the observed MMA for detected peptide ions can be worse than the lock mass since our approach works with all detected ions, irrespective of their detected S/N. Low S/N peaks are
more prone to error due to the effect of noise on the extraction of the detected peak centroid. Thus, to be conservative, 10 ppm was chosen as mass tolerance limit for all X-links searches in the current report. Future work will implement automatic gain control as is employed by LTQ-FT instrument (47) to further improve MMA.

**Identification of relationships/cross-link types**

The description of the program X-links and the estimated false discovery rate are reported elsewhere (40). Briefly, the program first searches the observed neutral masses for relationships between precursors and released peptide and reporter species to assign dead-end, intra- and inter-cross-links using the equations defined in Fig. 1b. Second, PIR-labeled peptide sequences and proteins of origin are identified using the measured accurate peptide masses to search against the restricted protein database. Fig. 4 shows an example relationship identification resultant from X-links searches. Fig. 4a illustrates a plot of overlaid EICs of an inter-cross-linked precursor (m/z 9785+), released peptide 1 (m/z 12432+), released peptide 2 (m/z 1281+), and reporter (m/z 1122+) ions, respectively. The precursor ion was observed with an alternating intensity pattern opposite to that of the reporter and released peptide ions. Fig. 4b shows superimposed spectra extracted from a low-energy precursor ion scan (black) and subsequent high-energy product ion scan (red) with corresponding precursor and product ions labeled schematically. To confirm the putative X-links assignment, we repeatedly performed LC/MS/MS with a quadrupole ion trap mass spectrometer, isolated the precursor ions (m/z 9785+) and fragmented them with CID. All intense fragment peaks observed in the MSMS spectrum were accounted for the fragmentation of one or two PIR labile bonds (Fig. 5a), validating the relationship identified with X-links and further illustrating the high cleavage specificity of PIR bonds. More validation examples of relationship identifications are provided in Supplementary Fig. 2, 3, and 4.
Identification of cross-linked peptide sequences/protein of origins

Peptide sequences and proteins of origin were identified using accurate peptide masses by searching against the restricted protein database. Proteins identified in the restricted database represent only that fraction of PIR-labeled proteins with highest abundance and PIR-reactivity/accessibility due to the limited dynamic range of detection. However, it should be noted that stage 2 samples were similarly constrained and thus, the detected peptides were likely from the most abundant and PIR-reactive proteins present in the stage 1 database. Statistically, stage 1 analysis expected a much greater probability of identifying proteins because any of the many tryptic peptides of a given protein can be used for identification. Stage 2 identification required analysis of the specific PIR-labeled peptides. Thus, even though the stage 1 database was likely not comprehensive, it was much more so than the stage 2 datasets and therefore, most PIR-labeled peptides were expected to originate from proteins identified in stage 1 studies.

Table 1 shows the summary of identified inter-cross-linked peptides using Rink-based PIR. A selected list of dead-end labeled peptides and intra-cross-linked peptides is shown in Supplementary Table 1 and 2, respectively. It is worth noting that most identified proteins are membrane proteins, which is consistent with our previous reports that PIR molecules preferentially label membrane proteins (38). Our results also show that several identified proteins were labeled with multiple peptide sequences indicating their high abundance or accessibility. For example, SO1778 (MtrC), SO0404 (hypothetical protein), and SO1779 (OmcA) were identified by 4, 4, and 7 different peptide sequences, respectively. On the other hand, the same peptide sequence can be labeled multiple times, which was observed previously in our efforts (39) and those of others (35, 36) to cross-link purified protein complexes. For example, FNPALKSELTYLYSNFFYDAK, a peptide from SO0404, was detected in two different inter-cross-links and one dead-end labeled peptide, suggesting this residue/region is highly exposed and accessible. In addition, several identified inter-cross-linked peptides in Table 1
were resultant from the same protein, e.g., OmcA, MtrC, SO0404, and SO4410 (glutamine synthetase, glnA), which may be produced by intra-cross-link within the same protein sequence or inter-cross-link from subunits of homo-oligomers. It is worth noting that the current mass tolerance (10 ppm) used in this report is not sufficient to uniquely identify all the cross-linked peptides, as demonstrated in the previous simulation study (40). Many PIR labeled peptide masses resulted in more than one hit (Supplementary Table 3). Further improvement of MMA or performing another MSMS on released peptides will help increase the ability to produce unambiguous protein interaction identification.

All the inter-cross-linked peptide masses listed in Table 1 were also used to search against unrestricted database with a 10 ppm constraint (Supplementary Table 4). In agreement with the previous simulation, most peptide masses generated multiple protein hits at 10 ppm MMA. At this mass tolerance, a restricted database and/or MS$^3$ analyses are required to further differentiate between multiple possibilities.

Although the accurate mass-based strategy for protein identification is highly reliable (40, 48), validation of a few examples was necessary for the initial application of PIR technology. This required MSMS of isolated PIR-labeled precursor and another isolation and fragmentation of the released peptide, thus MS$^3$. Fig. 5b illustrates an example of MS$^3$ validation of inter-cross-linked peptides found in X-links analysis and identified by accurate mass to be peptides from the proteins MtrC and SO0404. These data were obtained by isolating the inter-cross-linked precursor ion ($m/z$ 978$^{5+}$), activating the PIR labile bonds, releasing the peptides, isolating the intact peptide ions at $m/z$ 1243$^{2+}$ and 1280$^+$, respectively, and then fragmenting each. It is worth noting that two MS$^3$ mass spectra in Figure 5b did not generate comprehensive fragment ions with decent S/N due to the dominant neutral loss of the precursor ions and sensitivity limitations of MS$^3$ analysis. To further validate the peptide sequence identification by accurate mass and MS$^3$, a medium-sized peptide, LVYTTTKDLK with the expected tag on internal lysine was synthesized and the singly charged precursor ion was
subjected to tandem mass spectrometry (Supplementary Experimental and Supplementary Figure 8). These experiments resulted in spectra that showed the same neutral loss and backbone fragmentation peaks (MH⁺-H₂O, y7, y4 and y4-NH₃) as were observed with the endogenous PIR-labeled peptide shown in Figure 5b, supporting the fragment assignments. These results confirmed the accurate mass-based identification and MS³ validation analysis. Furthermore, this approach with synthetic peptides represents an additional, general and useful approach that can be employed to help confirm all identified PIR-labeled proteins and peptides. Additional MS³ sequence validation examples are included in Supplementary Fig. 2, 4 and 5.

It should be noted that MS³ analysis is manually obtained during repeated LC separation of stage 2 samples and thus both throughput and sensitivity of current validation experiments are low. Only those precursors with relatively high peak intensities in MS and MS² scans can generate sufficient signals in MS³. On the other hand, these results illustrate the benefits using high throughput multiplexed LC/MS and accurate mass-based protein identification, since none of the identified products need ever be among the most abundant species to be identified by accurate mass so long as they are detected.

DISCUSSION

The primary advantage of PIR technology that distinguishes it from other available methods is the capability of concurrent mapping protein-protein interactions and their binding interfaces in native biological systems. Two essential components enable PIR applications at the systems-level: novel PIR cross-linker structures and novel two-stage mass spectrometric analysis. The first application of Rink-based PIR with the use of S. oneidensis cells at proteome-wide scale resulted in identification of over 30 different inter-cross-linked peptides, 24 of which with unambiguous identification were included in Table 1. 17 peptide pairs were derived from inter-protein cross-links while 7 were resultant from either intra-protein cross-links or inter-subunit cross-links of homo-dimers. Importantly, while our initial PIR applications do not
show hundreds to thousands of interactions one might detect with other leading methods, such as YTH and TAP-tag technology, these results are distinguished by the fact that the obtained information is relevant to interactions detected in the native cellular environment and by the fact that the regions of close proximity of the interacting proteins can be determined. This constitutes information not readily achieved by any of the leading strategies. Furthermore, it is important to consider that such chemistry-based strategies are eminently extendable to include an array of chemical structures that can label and identify a wider range of protein interactions in cells. For example, the initial application of the more hydrophilic DP-based PIR molecules generated more than 10 inter-cross-linked peptides, most of which were different from those identified with Rink-based PIR. An example of identified inter-cross-link between ribosomal protein L10 (SO0222, rplJ) and ribosomal protein L7/L12 (SO0223, rplL) using DP-based PIR is shown in Supplementary Fig. 6. PIR cross-linkers with different structures and different physical properties can be used to simultaneously label and identify interactions among many different classes of proteins. As a consequence, the PIR strategy is scalable to enable multiplexed application of an array of PIR compounds, with variable physical properties e.g., hydrophobic character, charges, reactivity, etc., and independent reporter masses to enable even larger numbers of protein interactions to be identified from cells in a single experiment.

Among the identified cross-linked interactions, most involved outer membrane proteins. These results suggest that the current Rink-based PIR is particularly suitable for studies of membrane protein interactions, which is challenging with currently available techniques. In addition, the PIR approach allowed us to recapitulate some previously known interactions. For example, both in vitro (49) and in vivo (18, 50) studies showed that outer membrane cytochromes, OmcA and MtrC form a complex to facilitate electron transport in S. oneidensis. The identification of inter-cross-linked peptides (precursor mass 3793.8837 in Table 1) between OmcA and MtrC further confirmed the direct interaction of these two proteins (Supplementary
Fig. 7) and provides insight into the regions of protein sequences that are in close proximity in cells. Furthermore, outer membrane proteins SO0404, SO2001 (5-nucleotidase ushA), and SOA0110 (lipoprotein, putative) identified in the previous targeted OmcA-MtrC interaction network study using IP methods (18) were also identified in the present results in the inter-cross-links (2 from MtrC-SO0404, 1 from OmcA-SO2001, and 1 from SO0404-SOA0110). In the previous report, the OmcA-MtrC interaction network was mapped from protein identification of a gel band that appeared above 250 kDa. This high MW gel band was the result of cross-linking protein complexes of 3 or more components, and did not directly identify connectivity among interacting proteins. Identification of inter-cross-linked residues/peptides produced direct evidence of pair-wise protein-protein interactions and enables one to pinpoint contact regions of the interactions.

Taken together, our initial results using PIR technology on living cells of \textit{S. oneidensis} have demonstrated the capability for probing protein-protein interactions and protein topology concurrently without the need of specific antibodies and genetic manipulations of the system. This strategy can be further used as an \textit{in vivo} tool to investigate, snapshot, and compare potential binding sites of proteins and protein complexes in cells at different growth phases, culture conditions, cell cycles, external stresses, etc. Future work will include further improvement of the throughput, efficiency, and MMA for stage 2 analysis, and application of a library of PIRs with different reporter ions and isotopic labeling, which will ultimately allow large-scale quantitative profiling of protein-protein interactions and protein topologies in living cells.

**ACKNOWLEDGEMENTS**

This research was supported by NIH-NCRR, Grant No. 1 R01 RR023334-01A1 and the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.
REFERENCES

1. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340, Page.
2. 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., and Tyers, M. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415, Page.
3. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17, Page.
4. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24, Page.
5. Zhu, H., and Snyder, M. (2003) Protein chip technology. *Curr Opin Chem Biol* 7, Page.
6. Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., Walter, J. C., and LaBaer, J. (2004) Self-assembling protein microarrays. *Science* 305, Page.
7. Schachter, V. (2002) Bioinformatics of large-scale protein interaction networks. *Biotechniques* Suppl, Page.
8. Aloy, P., and Russell, R. B. (2006) Structural systems biology: modelling protein interactions. *Nat Rev Mol Cell Biol* 7, Page.
9. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, Page.
10. Lu, L., Arakaki, A. K., Lu, H., and Skolnick, J. (2003) Multimeric threading-based prediction of protein-protein interactions on a genomic scale: application to the Saccharomyces cerevisiae proteome. * Genome Res* 13, Page.
11. Aloy, P., and Russell, R. B. (2004) Ten thousand interactions for the molecular biologist. *Nat Biotechnol* 22, Page.
12. Ramachandran, N., Larson, D. N., Stark, P. R., Hainsworth, E., and LaBaer, J. (2005) Emerging tools for real-time label-free detection of interactions on functional protein microarrays. *Febs J* 272, Page.
13. Berggard, T., Linse, S., and James, P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics* 7, Page.
14. Kocher, T., and Superti-Furga, G. (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 4, Page.
15. Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 8, Page.
16. Bouwmeester, T., Bauch, A., Ruffner, H., Angrard, P. O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A. M., Schirle, M., Schlegl, J., Schwab, M., Stein, M. A., Bauer, A., Casari, G., Drewes, G., Gavin, A. C., Jackson, D. B., Jobery, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004) A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* 6, Page.
17. Gingras, A. C., Aebersold, R., and Raught, B. (2005) Advances in protein complex analysis using mass spectrometry. *J Physiol* 563, Page.
18. Zhang, H., Tang, X., Munske, G. R., Zakharova, N., Yang, L., Zheng, C., Wolff, M. A., Tolic, N., Anderson, G. A., Shi, L., Marshall, M. J., Fredrickson, J. K., and Bruce, J. E. (2008) In Vivo Identification of the Outer Membrane Protein OmcA-MtrC Interaction Network in Shewanella oneidensis MR-1 Cells Using Novel Hydrophobic Chemical Cross-Linkers. *J Proteome Res*, Page.
19. Rashid, K. A., Hevi, S., Chen, Y., Le Caherec, F., and Chuck, S. L. (2002) A proteomic approach identifies proteins in hepatocytes that bind nascent apolipoprotein B. *J Biol Chem* 277, Page.
20. Back, J. W., de Jong, L., Muijsters, A. O., and de Koster, C. G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* 331, Page.
21. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein
structures and protein-protein interactions. Mass Spectrom Rev 25, Page.
22. Orlando, V., and Paro, R. (1993) Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. Cell 75, Page.
23. Orlando, V., Strutt, H., and Paro, R. (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods 11, Page.
24. Schmitt-Ulms, G., Hansen, K., Liu, J., Cowdrey, C., Yang, J., DeArmond, S. J., Cohen, F. E., Prusiner, S. B., and Baldwin, M. A. (2004) Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. Nat Biotechnol 22, Page.
25. Vasilescu, J., Guo, X., and Kast, J. (2004) Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. Proteomics 4, Page.
26. Kim, S., Wong, P., and Coulombe, P. A. (2006) A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. Nature 441, Page.
27. Tardiff, D. F., Abruzzi, K. C., and Rosbash, M. (2007) Protein characterization of Saccharomyces cerevisiae RNA polymerase II after in vivo cross-linking. Proc Natl Acad Sci U S A 104, Page.
28. Guerrero, C., Tagwerker, C., Kaiser, P., and Huang, L. (2006) An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26S proteasome-interacting network. Mol Cell Proteomics 5, Page.
29. Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L., and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol Cell Proteomics 5, Page.
30. Trakselis, M. A., Alley, S. C., and Ishmael, F. T. (2005) Identification and Mapping of Protein-Protein Interactions by a Combination of Cross-Linking, Cleavage, and Proteomics. Bioconjugate Chem. 16, Page.
31. Vasilescu, J., and Figeys, D. (2006) Mapping protein-protein interactions by mass spectrometry. Curr Opin Biotechnol 17, Page.
32. Novak, P., and Giannakopoulos, A. E. (2007) Chemical cross-linking and mass spectrometry as structure determination tools. Eur J Mass Spectrom (Chichester, Eng) 13, Page.
33. Bennett, K. L., Kussmann, M., Bjork, P., Godzxon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. Protein Sci 9, Page.
34. Seebacher, J., Mallick, P., Zhang, N., Eddes, J. S., Aebersold, R., and Gelb, M. H. (2006) Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. J Proteome Res 5, Page.
35. Trakselis, M. A., Kamada, K., Burley, S. K., Fenyo, D., Chait, B. T., and Muir, T. W. (2003) A modular cross-linking approach for exploring protein interactions. J Am Chem Soc 125, Page.
36. Sinz, A., Kalkhof, S., and Ihling, C. (2005) Mapping protein interfaces by a trifunctional cross-linker combined with MALDI-TOF and ESI-FTICR mass spectrometry. J Am Soc Mass Spectrom 16, Page.
37. Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008) Identification of cross-linked peptides from large sequence databases. Nat Methods, Page.
38. Tang, X., Yi, W., Munske, G. R., Adhikari, D. P., Zakharova, N. L., and Bruce, J. E. (2007) Profiling the membrane proteome of Shewanella oneidensis MR-1 with new affinity labeling probes. J Proteome Res 6, Page.
39. Anderson, G. A., Tolic, N., Tang, X., Zheng, C., and Bruce, J. E. (2005) Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. Anal Chem 77, Page.
40. Anderson, G. A., Tolic, N., Tang, X., Zheng, C., and Bruce, J. E. (2007) Informatics strategies for large-scale novel cross-linking analysis. J Proteome Res 6, Page.
41. Green, N. S., Reisler, E., and Houk, K. N. (2001) Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. Protein Sci 10, Page.
42. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3, Page.
43. Lomant, A. J., and Fairbanks, G. (1976) Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). J Mol
44. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, Page.

45. Ledford, E. B., Jr., Rempel, D. L., and Gross, M. L. (1984) Space charge effects in Fourier transform mass spectrometry. *Anal Chem* 56, Page.

46. Nepomuceno, A. I., Muddiman, D. C., Bergen, H. R., 3rd, Craighead, J. R., Burke, M. J., Caskey, P. E., and Allan, J. A. (2003) Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 75, Page.

47. Peterman, S. M., Dufresne, C. P., and Horning, S. (2005) 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, Page.

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

49. Shi, L., Chen, B., Wang, Z., Elias, D. A., Mayer, M. U., Gorby, Y. A., Ni, S., Lower, B. H., Kennedy, D. W., Wunschel, D. S., Mottaz, H. M., Marshall, M. J., Hill, E. A., Beliaev, A. S., Zachara, J. M., Fredrickson, J. K., and Squier, T. C. (2006) 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* 188, Page.

50. Ross, D. E., Ruebush, S. S., Brantley, S. L., Hartshorne, R. S., Clarke, T. A., Richardson, D. J., and Tien, M. (2007) Characterization of protein-protein interactions involved in iron reduction by Shewanella oneidensis MR-1. *Appl Environ Microbiol* 73, Page.
Tables:

Table 1. A summary of inter-cross-linked peptides and proteins identified using Rink-based PIR.

Figure Captions:

Figure 1. PIR structure.  a) Conceptual modular design of novel cross-linkers, protein interaction reporters (PIRs).  b) The specific fragmentation pattern of PIR-labeled peptide distinguishes dead-end, intra-, and inter-cross-linked peptides. The neutral mass of the precursor ion equals the sum of the neutral masses of its product ions.  c) Structure of a biotinylated Rink-based PIR. 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 yellow.  The singly- and doubly-charged reported ions are 1122.5 and 561.7, respectively.  d) Structure of a biotinylated DP-based PIR. 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 yellow.  The singly charged reporter ion is 752.4.

Figure 2. Diagram of two-stage mass spectrometric strategy.  After in vivo labeling, protein extraction, and affinity capture of labeled proteins, enriched PIR-labeled proteins are divided into two parts. Stage 1 involves digestion and shotgun LC/MS/MS for protein identification to constitute a restricted protein database.  Stage 2 is performed by digesting PIR-labeled proteins first, followed by another 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 3. Chromatograms of a multiplexed LC/FTICR-MS of PIR-labeled peptides with alternating low energy (-4 v) and high energy (-22 v) applied in the collision cell.  

a) Base peak chromatogram (BPC).  
b) Extracted ion chromatogram (EIC) of the reporter ion.  The inset in a) is the overlaid EICs of a PIR-labeled precursor ion (red) and its released peptide ion (blue) and reporter ion (yellow).

Figure 4. An example of inter-cross-link determination from multiplexed spectra based on the mathematical relationship (precursor = peptide 1 + peptide 2 + reporter).  

a) Overlaid EICs of precursor (m/z 9785+, red), released peptide 1 (m/z 12432+, green), released peptide 2 (m/z 1281+, blue), and reporter (m/z 1122+, yellow) ions, respectively.  
b) Overlaid FTICR mass spectra extracted from scan 765 (low energy, black) and scan 766 (high energy, red).  All major products resultant from cleavage one and two labile bonds in PIR are assigned and labeled schematically.

Figure 5. An example of validation of inter-cross-link determination and peptide sequence identification for the precursor ion with neutral mass of 4884.3531.  

a) Isolation and MS² of precursor ion (m/z 9785+).  All peaks produced from cleavage of one and two PIR labile bonds are labeled schematically.  
b) Isolation and MS³ of the released peptide ion with m/z 1243²+ and 1280¹, respectively.
| Precursor Mass | Peptide Mass | Identified peptide | Error (ppm) | Protein of origin | Type | MW/ kDa |
|---------------|-------------|--------------------|-------------|-------------------|------|--------|
| 5047.3774     | 1911.9354   | TDNTPGTKVPTSFAWK   | 0.67        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
|               | 1815.8954   | GFALSNKSVSTYNEATK  | 2.69        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| 4916.3157     | 2271.0398   | LMSEDQIKTTFADVAGDEAK | 1.35 | SO1197|Cell division protein (ftsH) | IM | 71 |
|               | 1325.7302   | NTPVEKGIGQVJK      | 3.38        | SO2756|antioxidant | C | 22 |
| 4906.2116     | 2641.1930   | ASFSGNSDIMLVPDSEFFKYMKEKNTPSER | 5.68 | SO0606|inf protein (hflC) | IM | 33 |
| 945.4501      | 2641.1930   | ASFSGNSDIMLVPDSEFFKYMKEKNTPSER | 1.57 | SO0606|inf protein (hflC) | IM | 33 |
| 4884.3531     | 2384.1255   | FNPAKSELTYLYLNNFYDAK | 2.92 | SO0404|hypothetical protein | OM | 130 |
| 1180.6630     | 2384.1255   | FNPAKSELTYLYLNNFYDAK | 6.56 | SO0404|hypothetical protein | OM | 130 |
| 4800.2140     | 2384.1324   | FNPAKSELTYLYLNNFYDAK | 3.61 | SO0404|hypothetical protein | OM | 130 |
| 1096.5665     | 2384.1324   | FNPAKSELTYLYLNNFYDAK | 1.68 | SO0404|hypothetical protein | OM | 130 |
| 4717.2362     | 2580.3263   | KFGTFTVAETGWPALDPLFIENK | 2.32 | SO4680|conserved hypothetical protein | C | 41 |
| 817.4073      | 2580.3263   | KFGTFTVAETGWPALDPLFIENK | 1.17 | SO4680|conserved hypothetical protein | C | 41 |
| 4429.1018     | 1815.8948   | GFALSNKSVSTYNEATK  | 3.02        | SO1779|decaheme cytochrome c (mtrC) | OM | 71 |
| 1293.6500     | 1815.8948   | GFALSNKSVSTYNEATK  | 5.01        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| 4353.0545     | 2222.1268   | SQINPKFQDLTADVFSGR | 2.85        | SO0404|hypothetical protein | OM | 130 |
| 811.3956      | 2222.1268   | SQINPKFQDLTADVFSGR | 7.10        | SO0404|hypothetical protein | OM | 130 |
| 4256.0728     | 2119.1187   | GKPHERAFSFLHNHVNANK | 0.02        | SO1778|decaheme cytochrome c (mtrC) | OM | 71 |
| 817.4082      | 2119.1187   | GKPHERAFSFLHNHVNANK | 1.90        | SO1778|decaheme cytochrome c (mtrC) | OM | 71 |
| 4228.0116     | 524.3454    | LHKK                | 2.52        | SO1778|decaheme cytochrome c (mtrC) | OM | 71 |
| 2384.1452     | 524.3454    | LHKK                | 4.66        | SO0404|hypothetical protein | OM | 130 |
| 3920.8081     | 1468.7459   | GYOQWQYINAKK        | 0.27        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| 1132.535      | 1468.7459   | GYOQWQYINAKK        | 0.99        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| 3793.8837     | 1293.6629   | GLKTDNTYPGTK        | 5.21        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
|                | 1293.6629   | GLKTDNTYPGTK        | 0.88        | SO1779|decaheme cytochrome c (mtrC) | OM | 71 |
| 3744.775      | 1448.6644   | WNGTDTNSSAAEKR      | 1.81        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| 976.5662      | 1448.6644   | WNGTDTNSSAAEKR      | 7.54        | SO1779|decaheme cytochrome c (omcA) | OM | 79 |
| Accession | PSM   | Description                                                                 | Type     | Prec. | Charge | m/z   |
|-----------|-------|------------------------------------------------------------------------------|----------|-------|--------|-------|
| 3702.7157 | 1293.6484 | GLKT D N TYP GTK                                                              | OM       | 79    | 1      | 3702.7157 |
| 1089.5310 |       | EVTDEAGNKK                                                                    | OM       | 61    | 1      | 3702.7157 |
| 3440.6819 | 729.3697 | KEYK                                                                         | OM       | 40    | 1      | 3702.7157 |
| 1391.7772 |       | YQQLVK TNAISK                                                                 | OM       | 40    | 1      | 3702.7157 |
| 3355.6241 | 1015.5170 | MENPEIKR                                                                    | C        | 56    | 1      | 3702.7157 |
| 1020.5345 |       | GLSAKAFDGR                                                                    | C        | 20    | 1      | 3702.7157 |
| 3300.5526 | 848.4773 | MKSKLDK                                                                      | OM       | 66    | 1      | 3702.7157 |
| 1132.5320 |       | EKVGETEADR                                                                   | OM       | 79    | 1      | 3702.7157 |
| 3291.5894 | 982.5242 | SKEAANHVK                                                                    | C        | 59    | 1      | 3702.7157 |
| 989.5303  |       | EISMNIKR                                                                     | C        | 13    | 1      | 3702.7157 |
| 3272.5790 | 937.5061 | IAGSMGFKK                                                                    | C        | 77    | 1      | 3702.7157 |
| 1015.5167 |       | MENPEIKR                                                                     | C        | 56    | 1      | 3702.7157 |
| 3246.6089 | 987.5094 | AEFA KTHGK                                                                    | C        | 102   | 1      | 3702.7157 |
| 939.5683  |       | GKPVVWVR                                                                     | IM       | 21    | 1      | 3702.7157 |
| 3218.6457 | 850.4869 | FNTLT KK                                                                     | C        | 52    | 1      | 3702.7157 |
| 1048.6404 |       | IPVVPSPKGR                                                                   | C        | 52    | 1      | 3702.7157 |
| 3179.4984 | 920.4166 | WADKMDR                                                                      | OM       | 84    | 1      | 3702.7157 |
| 939.5683  |       | GKPVVWVR                                                                     | IM       | 21    | 1      | 3702.7157 |
| 3021.4109 | 811.3898 | FKSGDMK                                                                      | OM       | 130   | 1      | 3702.7157 |
| 890.4986  |       | LGVFGKDR                                                                     | OM       | 150   | 1      | 3702.7157 |
| 2194.0687 | 437.2634 | KYK                                                                           | C        | 103   | 1      | 3702.7157 |
| 437.2634  |       | YKK                                                                           | OM       | 9     | 1      | 3702.7157 |

*Type of identified proteins; OM, IM, and C denotes outer-membrane, inner membrane, and cytoplasmic proteins, respectively.
Figure 1

a) Affinity group
   Mass-encoded reporter
   MS/MS labile region
   Reactive group

b) Inter-cross-link
   =
   + reporter

Intra-cross-link
   =
   + reporter

Dead-end
   =
   + remaining tag
   + dead-end residual tag
   + reporter
Figure 2.

PIR reaction → Clean up → Cell lysis → Enrich PIR-labeled peptides → Stage 1 → Digestion → Restricted protein database → Enrich PIR-labeled peptides → Stage 2 → Multiplexed LC/FTICR-MS

2D LC/MS/MS Protein identification → Protein-protein interactions Regions of interactions

Restricted protein database
Figure 3.

a) BPC

- PIR labeled precursor
- Released intact peptide
- Reporter

b) EIC of reporter ion
Figure 4.

EIC
- **Precursor 978**
- **Peptide 1 1243**
- **Peptide 2 1281**
- **Reporter 1122**

Low Energy
High Energy

Scan #

m/z

a)

b)
Figure 5.

a) MSMS of 9785+ precursor ion

b) MSMS of 12432+