Published in final edited form as:

J Proteome Res. 2012 February 3; 11(2): 1027–1041. doi:10.1021/pr200775j.

In Vivo Application of Photocleavable Protein Interaction Reporter Technology

Li Yang¹, Chunxiang Zheng², Chad R. Weisbrod³, Xiaoting Tang³, Gerhard R. Munske¹, Michael R. Hoopmann³, Jimmy K. Eng³, and James E. Bruce^{3,*}

¹Department of Chemistry, Washington State University, Pullman, Washington, 99164

²Department of Chemistry, University of Washington, Seattle, Washington, 98109

³Department of Genome Sciences, University of Washington, Seattle, Washington, 98109

Summary

In vivo protein structures and protein-protein interactions are critical to the function of proteins in biological systems. As a complementary approach to traditional protein interaction identification methods, cross-linking strategies are beginning to provide additional data on protein and protein complex topological features. Previously, photocleavable protein interaction reporter (pcPIR) technology was demonstrated by cross-linking pure proteins and protein complexes and the use of ultraviolet light to cleave or release cross-linked peptides to enable identification. In the present report, the pcPIR strategy is applied to *E. coli* cells and *in vivo* protein interactions and topologies are measured. More than 1600 labeled peptides from *E. coli* were identified, indicating many protein sites react with pcPIR *in vivo*. From those labeled sites, 53 *in vivo* inter-cross-linked peptide pairs were identified and manually validated. Approximately half of the interactions have been reported using other techniques, although detailed structures exist for very few. Three proteins or protein complexes with detailed crystallography structures are compared to the cross-linking results obtained from *in vivo* application of pcPIR technology.

I. Introduction

Mapping protein interaction networks is a critical goal of proteomics research to better understand the biological function of proteins in cells. Traditional methods to study protein-protein interactions from cells include yeast two-hybrid (YTH)¹, immunoprecipitation (IP)² and tagged immunoprecipitation³, eg. Flag-tag or TAP-tag. Chemical technologies such as chemical cross-linking have been applied for many years, but have gained increasing interest in recent years^{4–7}. Cross-linker application can result in creation of new covalent bonds between proteins that allow visualization of interacting regions. Exceptionally large numbers of potential protein interactions can be identified with YTH or IP-based methods^{8–10}. However, in very few cases are any topological features of the interacting regions known or revealed from these studies. Cross-linking strategies, on the other hand, can provide unique information including the identities of native interacting partners and topological features of the interacting regions *in vivo*. In the past, cross-linking strategies were used on limited applications to study topologies of one or a few proteins. Recently, a few applications on complex biological systems, including cells and cell lysates have been reported^{11–13}. Specifically, PIR cross-linking *E. coli* experiments revealed novel *in vivo* topological information on interacting proteins. Excitingly, these efforts resulted in identification of

Corresponding author: James E. Bruce, Current address: Department of Genome Sciences, University of Washington, Seattle, Washington, 98109, jimbruce@u.washington.edu, Phone: 206-616-0010, Fax: 206-616-0008.

interacting regions on proteins that are resistant to crystallization¹³. Such applications indicate cross-linking technologies are complimentary to existing crystallography and other protein interaction measurements, and can yield unique *in vivo* topological information on proteins or protein complexes.

Despite this progress, few reports of cellular application of cross-linking methods currently exist, which underscores the need to develop new cross-linker molecules and data analysis capabilities. A major challenge includes the difficulty to reliably identify large numbers of cross-linked peptides. Part of the reason for this difficulty is the incredible sample complexity and wide dynamic range resultant from cross-linking reactions carried out with cells. Intercross-linked peptide pairs, compared to non-reacted peptides and dead-end labeled peptides normally comprise only a small portion of the sample. In addition, the lack of robust and reliable mass spectrometric identification strategies further hinders this progress. In cross-linking strategies with non-cleavable cross-linkers, such as DSS or BS3, the identification of cross-linked peptides is extremely challenging, because the database of cross-linked peptides increases greatly with increasing number of candidate proteins. For example, the database of wild type *E. coli* (K12) contains 4178 proteins, assuming each protein can produce 100 peptides, with non-cleavable cross-linkers the database of cross-linked pairs is a combination of all of these possible peptides and an approximate database of all possible cross-linked species contains 8.7×10^{10} candidates. As such, cross-linking experiments have normally been restricted to low complexity samples, such as those containing only one or a few proteins⁵. To overcome these challenges and enable large database searches, two strategies have been developed. One method is to develop new database search algorithms. In the recent publication from Aebersold and coworkers, a new database search software xQuest was reported¹¹. With this strategy, Aebersold and co-workers were able to identify inter-cross-linked peptide pairs from *E. coli* cell lysate, which is an important breakthrough in the area of cross-linking studies¹¹. Another approach is to use a cleavable cross-linker to overcome the identification problem¹⁴. Protein interaction reporter (PIR) technology developed by Tang et al. was first to show the utility of cleavable cross-linker technologies that allowed identification of released peptides^{12, 13, 15, 16}. PIR cross-linker is fragile under low energy mass spectrometric collision conditions; therefore with collision induced dissociation (CID) or in-source activation (ISCID), the cross-linker is cleaved at a certain bond to release the cross-linked peptides; MS/MS and accurate mass can then be used to identify the released peptide. In this way, each released peptide is modified with a small and defined modification group from the cross-linker and the database search strategy is analogous to a traditional proteomics search with a variable modification. Search algorithms like Mascot and SEQUEST are compatible with this protein interaction reporter (PIR) strategy. In the first cellular application of PIR strategy¹², the cross-linker was cleaved under low-energy CID conditions, and the released peptides were subjected to MS³ fragmentation. The accurate mass of released peptide was used to identify the peptide's sequence and MS³ was used to validate the identity. However MS³ is time consuming and it typically produces lower quality fragmentation spectra than common MS/MS experiments. An advancement on the method includes cleavage of the PIR cross-linker under in-source collisionally induced dissociation (ISCID) conditions. Specifically, in a recent *E. coli* cellular cross-linking study¹³, the cross-linker was cleaved under optimized ISCID condition and the released peptides were fragmented in the followed MS/MS scan event in the LTQ instrument. With this method, 65 peptide pairs cross-linked in *E. coli* cells were reported and novel native topological information was derived¹³. Furthermore, advancements in the PIR technology include modifications of cross-linker structure, PIR activation methods and MS/MS to allow improved identification of released peptides. A photocleavable cross-linker pcPIR was developed and demonstrated with pure proteins and protein complexes¹⁶, where application of UV laser light on the ESI spray tip leads to release of the cross-linked peptides after LC separation but before mass spectrometric

analyses. In this way, the released peptides are produced in solution phase, which potentially can produce ions with higher charge states and thus result in better MS/MS fragmentation results.

In this report, the initial application of pcPIR technology to cells is described. In total, 1602 labeled peptides were identified, from which 53 inter-cross-linked peptide pairs were identified and validated. Many of these pairs were also identified with our previous PIR compound in a report elsewhere¹³. These pairs as well as several pairs unique to the present report show excellent agreement with the protein crystal structure data where available. The unique information resultant from pcPIR application to cells relevant to *in vivo* protein and protein complex topological features further suggests that cross-linking strategies will enable greater insight on protein interactions and function *in vivo*.

II. Experimental Procedures

1. Materials

Fmoc-amino acids and hydroxyethyl photolinker were purchased from Novabiochem (San Diego, CA). The pcPIR cross-linker was synthesized as described previously¹⁶, except that the fmoc-Arginine residue was replaced with fmoc-Glycine and the hydroxymethyl photolinker was replaced with hydroxyethyl photolinker. Deprotection and purification methods all followed the previous report¹⁶. Dimethyl sulfoxide (DMSO) and Tergitol solution (70% NP40 in water) were purchased from Sigma-Aldrich (St. Louis, MO) and used without purification. Monomeric avidin ultralink resin and mass spectrometry-grade Trypsin endoproteinase were purchased from Pierce (Rockford, IL); and the phenylmethylsulfonyl fluoride (PMSF) was from GBiosciences (Maryland Heights, MO).

2. Cross-linking sample preparation

a. *In vivo* cross-linking of the intact *E. coli* cells—*E. coli* cells (K12) were harvested at the mid-log phase and washed 5 times with 1 mL cold phosphate buffer saline (PBS) before cross-linking. The cells were pelleted after washing and 100 μ L cell pellets were suspended in 1 mL PBS solution. An aliquot of 10 μ L of a pcPIR solution (100mM in DMSO) was added to the suspension (1 mL cells in 0.1% NP40/PBS), which resulted in final concentrations of 1 mM pcPIR and 1% DMSO. The mixture was incubated at 37 °C for 30 min. A final concentration of 1 mM pcPIR was used in the experiment mainly because of the saturation of pcPIR in PBS buffer. The current identification results are obtained after two steps of affinity enrichment, therefore lower the concentration of cross-linker would likely result in worse identification results. The cells were then washed 5 times with 1 mL cold PBS each time and then resuspended in a 1 mL solution of 0.1% NP40 in PBS. The cells were then lysed with ultra sound sonication and subjected to centrifugation at 4 °C at 15,000 G for 30 min. The soluble portion was used directly in the next step. The insoluble portion was dissolved again with 200 μ L of 8M urea in 100mM Tris-HCl buffer and then diluted with 0.1% NP40 in PBS to a final volume of 1 mL. These two solutions were combined for the next step.

b. Enrichment of the cross-linked products—The enrichment of the cross-linked products was achieved by a double avidin capture strategy. An aliquot of 100 μ L monomeric avidin slurry (50 μ L of avidin beads) was added to the cell lysis. The mixture was incubated on a temperature-controlled shaker (Eppendorf Thermomixer R) at 1000 rpm at room temperature for 30 min. Then the beads were washed 3 times with 1 mL of a 0.1% NP40/ NH_4HCO_3 solution. The beads were resuspended in 100 μ L of an NH_4HCO_3 solution, reduced and alkylated with final concentrations of 5 and 10 mM TCEP and IAA, respectively. Five micrograms of trypsin and 1 μ L of a 0.1 M CaCl_2 solution were added to

the suspension of beads. The digestion was incubated 3 hours at 37 °C. Then, the trypsin inhibitor PMSF was added to yield a final concentration of 1 mM. The mixture was incubated at room temperature for 30 min. Another 100 µL aliquot of avidin slurry was added and the mixture was incubated on the shaker at 1000 rpm for another 30 min. The beads were washed 3 times with 0.1% NP40/NH₄HCO₃, and then another 3 times with NH₄HCO₃. Finally the enriched products were eluted from the beads with 200 µL of elution buffer (75% acetonitrile and 0.5% formic acid in DI water) 4 times. The eluent was placed under vacuum and dried overnight (Thermo Speed Vacuum SPD131DDA). Hasher elution methods can also be used to elute the cross-linked products from avidin beads. Here acid and acetonitrile were used to avoid further clean up steps, which will also cause sample loss. In the future, the elution methods can be optimized.

c. Off-line photocleavage of the cross-linked products—The dried eluent was redissolved in 50 µL of a 0.1% formic acid solution. One microliter of the sample was diluted into 10 µL with water, and placed under a UV lamp for four hours, the wavelength of which centers at 365 nm. The remainder of the sample was used for the LC-mass spectrometry experiments described below.

5. Mass spectrometry analyses

a. On-line photocleavage and LC-MS analyses for identification of cross-linking relationships—The enriched cross-linked products were further fractionated with HPLC (nanoAcquity Waters, Milford, MA) and detected with a Velos-FT MS, which was built in house and described elsewhere¹⁷. High mass accuracy resultant from FT-MS detection is necessary in the identification of cross-linking relationships¹⁸. A 35-cm long C18 column was made in house by packing frit fused silica capillary (360 µm × 75 µm) with MAGIC C18AQ 100A 5U beads (Michrom Bioresources, Inc. Auburn, CA). A 4-cm long electrospray ionization (ESI) tip was made by pulling a 75 µm i.d. fused silica capillary with a laser puller (Sutter Instrument Co.). The ESI tip was then connected to the analytical column. A 2-cm long trap column was prepared similarly by packing a frit fused silica capillary (360 µm × 100 µm) with MAGIC C18AQ 200A 5U beads. For on-line photocleavage, a nitrogen laser with output at 337.1 nm was set up as described elsewhere¹⁶. The following LC gradient was used: 0–120 min 5%–40% buffer B, 121–135 min flushing with 80% buffer B and 136–160 min equilibrating with 5% buffer B. Buffers A and B consisted of 0.1% formic acid in DI water and 95% acetonitrile, 0.1% formic acid in DI water, respectively. The MS instrument resolving power was set at 50K to obtain high mass accuracy data. A novel instrument control language (ITCL) program was written in the lab, which was designed to trigger the laser by employing the atmospheric pressure corona discharge (APCI – presently unused) trigger for method based experiments. This allows the user to tailor methods for interleaving low energy and high energy activation scans. The UV laser was coupled with LC-MS to enable four consecutive scans with laser turned on and then four scans with laser off. This cycle was repeated throughout the entire LC-MS run to enable cross-linked peptide relationships to be identified as described below.

b. LC-MS/MS identification of released peptides—The sample prepared with off-line photocleavage was analyzed with LC-MS/MS to allow identification of released peptides. The LC-MS/MS setup included one MS scan with 25K resolving power in the FT analyzer followed by 10 data-dependent MS/MS measurements in the Velos analyzer. The dynamic exclusion repeat and exclusion duration were each 15 sec.

6. Data analyses

The MS data were searched with software BLinks¹⁹ and/or X-links¹⁸ to identify the cross-linking relationships. Mass tolerance for accepted cross-linking relationships in BLinks was

set to 15 ppm. *X-links* parameters included: filter *m/z* range 300–2000, filter isotopic fit 0.0–0.2 and mass tolerance 15 ppm. The LC-MS/MS data were searched against the *E. coli* K12 database with Mascot (version: 2.3.01). The search parameters included: up to 3 possible missed cleavages, precursor error tolerance 25 ppm and fragmentation error tolerance 0.6 Da. Modifications included: fixed modification of Carbamidomethyl on cysteine residues, and variable modification of oxidation on methionine residues. The remaining tag from pcPIR cross-linker (mass = 100.0160) was treated as a variable modification on lysine residues or protein N-termini. The identities of the released peptides were matched with accurate masses (within 10 ppm) to the cross-linking relationships with software *X-links* or BLinks.

7. Mass spectrometric verifications and measuring the cross-linking distances

The cross-linking relationships and the identities of the released peptides were manually verified (parent verification and released peptide verification). For parent verification, a mass-and-time targeted inclusion list of the cross-linked parents was generated from the search results with *X-links* and BLinks. These parent ions were subjected to an LC-MS/MS experiment with the same LC setup and gradient as the LC-MS run to maintain the retention time information. The parent ions were cleaved in the MS/MS scan events with a CID energy of 30, so that the released peptides and the reporter ion can be produced and mass relationships can be verified. It should be noted that pcPIR molecules are designed to allow specific fragmentation under UV irradiation to enable identification of cross-linked relationships and peptides. However, CID of pcPIR cross-linked peptides is useful for validation of mass relationships and peptides identified from photo dissociation experiments. A second mass-and-time targeted MS/MS experiment targeting the released peptides was performed to allow real time peptide identification. Applied laser pulse frequency was 20 Hz to enable maximum cleavage of pcPIR to produce the released peptides. These collected verification results were manually compared to the Mascot results to ensure unambiguous identification of released peptides. In addition, the chromatographic features of the cross-linked products were also manually inspected. The extracted ion chromatograms (EICs) of the cross-linked parent and the released peptides ions must be observed with matching LC elution profiles, as described previously¹⁹.

Finally the cross-linking results were used to determine protein and protein interaction topologies that existed in cells during the application of pcPIR molecules. For cases with existing crystal structure data, the distances between the cross-linked lysine residues were measured from their positions within the crystal structures. The identified protein interactions were compared with EciD (*E. coli* interaction database, <http://ecid.bioinfo.cnio.es/>) to highlight protein interactions previously identified with other technologies. A novel homodimer complex identified in this work (kdud) was modeled with tools available on the SymmDock web server²⁰ with a symmetry order of 2 using pcPIR-derived cross-linking constraints.

III Results

Major challenges in cross-linking research include the difficulties in identification and verification of large numbers of cross-linking relationships. The difficulty in identifying lots of relationships is mainly due to sample complexity, as mentioned above. In the cross-linking mixture, the labeled or cross-linked peptides are a small portion compared with the non-reacted peptides, because of the cross-linker/protein ratio and the hydrolysis of the cross-linker. Two avidin affinity purification steps were performed on the protein and peptide levels separately, to accumulate the labeled products. With the advancements in the pcPIR approach, large numbers of released peptides were identified. In this study, 1602 non-redundant *in vivo* labeled peptides were identified, the sub-cellular localizations of which are

predicted with PSORT²¹ and shown in Supplementary Figure 1. This large number of identified peptides indicates that many sites on the *E. coli* proteins were labeled or cross-linked *in vivo*, although most of the released peptides may come from the dead-end, intra-cross-linked products or as yet, unidentified pcPIR cross-linked peptide pairs. Dead-end products indicate the surface accessible sites, and are useful for identifying which sites are accessible in cells. Importantly, these sites identify regions in proteins that are solvent- and cross-linker-accessible as they exist within cells. A total of 53 cross-linked pairs were identified and manually verified in this first cellular application of the pcPIR technology. The inter-cross-linked peptide pairs are listed in Table 1. Although the number is still small compared with the traditional methods, such as IP-based or YTH approaches, the cross-linking data show novel information of new interactions or topological features of the interfacing areas from the *in vivo* interactions. *E. coli* was chosen as an experimental model because it is commonly used in different kinds of genomics and proteomics research. Previous knowledge of *E. coli* protein structures and interactions is relatively rich and many NMR, crystallography or modeling structures are available for *E. coli* proteins and protein complexes. For the inter-cross-links where two peptides belong to the same protein and three-dimensional protein structures exist, the distances between the two cross-linked lysine side chains were measured from published structures and listed in Table 1. In this dataset, 17 intra-protein, 9 homodimers and 27 inter-protein inter-cross-links were found, which can provide valuable information on *in vivo* protein interactions or protein complex topologies. Among the inter-protein cross-links, 13 interactions are previously identified with other approaches, which together with the intra-protein cross-links are marked with bold font in the table. Here several of the known protein complexes are chosen and discussed, because their detailed three-dimensional structures are available. The information obtained from the cross-linking experiment can then be compared with these 3D structures.

IV. Discussion

1. pcPIR strategy

The general concept of the pcPIR technology has been described in an earlier paper¹⁶. Here it will be described briefly. The structure of the present pcPIR cross-linker is shown in Figure 1a, which contains an ultraviolet (UV) light cleavable bond on each arm (Figure 1b, c). Compared to the previous generation of pcPIR, the cleavage group in the new pcPIR is changed from a hydroxymethyl photolinker to a hydroxyethyl photolinker. As discussed in Bradley's review²², adding a methyl group on the photolinker improved the UV cleavage efficiency by 5–7 fold, which makes the cleavage of the new pcPIR more efficient under UV laser irradiation. Such specific cleavage at the UV sensitive bond results in the specific product/precursor mass relationships of the cross-linked pairs. However, non-reacted peptides are unaffected by UV irradiation (Figure 2) and thus are readily distinguished from pcPIR reaction products. An advantage of the pcPIR and PIR strategies is that with specific mass relationships created from specific cleavage, both the cross-linking relationships among peptides and the types of different cross-linking products can automatically be identified. Our software *X-links* was developed to allow analysis of the LC-MS data to enable the identification of the cross-linked relationships with PIR molecules¹⁸ and is useful for pcPIR cross-linked molecules as well. BLinks¹⁹ builds on the results from *X-links* and includes correlation of chromatographic profiles to enable p-value estimates of PIR-derived relationships. Therefore, a four-step mass spectrometric identification-and-verification strategy was developed to allow identification of the cross-linked peptide relationships, the released peptides and the manual verifications of both. These four steps are listed and briefly described in Figure 3. The first step included LC-MS separation of pcPIR-labeled peptides coupled with UV cleavage. Throughout the whole LC-MS run, the UV laser was alternately turned on or off for four consecutive scans. Such UV laser control enables

the observation of the intact cross-linked parent ions in the “laser off” scans and the cleavage products, e.g. the released peptides and sometimes the reporter, in the “laser on” scans. *X-links* and *BLinks* analyses were then used to analyze the datasets from the laser off and laser on scans to generate a list of possible cross-linked peptide relationships. Two mass-and-time targeted inclusion lists of cross-linked parent ions and released peptide ions were generated separately. These two lists were used in the second and the third LC-MS/MS runs respectively. The purpose of the second LC-MS/MS experiment was to confirm the cross-linked peptide relationships identified with *X-links* or *Blinks* analyses (step 2). A pcPIR cross-linked parent ion can also be fragmented under CID condition to produce the two released peptides. Therefore in the second LC-MS/MS run, isolation and fragmentation of the parent ions were performed with the mass-and-time targeted inclusion list; the UV laser was not used for this run. In this case, CID energy similar to that normally used for peptide fragmentation was employed to cleave the bonds in these cross-linkers; pcPIR bonds fragment under collisional activation but do not appear to be as labile as those in PIR molecules (data not shown). The third LC-MS/MS measurement (step 3) was used to identify or verify the released peptides. Therefore the maximum power of UV laser was used to enable the highest efficiency in generating the released peptides. Those released peptides were fragmented with a mass-and-time targeted CID experiment. The cross-linked peptide pairs were cleaved on-line before they entered the mass spectrometer, allowing MS/MS to be carried out on the released peptides that originated from identified pcPIR cross-linked relationships. These released peptides can be considered as tryptic digested peptides with a cross-linker-related variable modification. Thus database search algorithms such as Mascot and SEQUEST can be used to reliably identify the released peptide sequences. In addition, pcPIR can be completely cleaved under extended UV light irradiation (step 4). Therefore a 4-hr photocleavage (which is likely more than adequate to cleave all pcPIR molecules) with a UV lamp was used to generate high yield of the released peptides, which were then subjected to another data-dependent LC-MS/MS measurement. Off-line photocleavage provides the best chance to allow identification of the released peptides. If a released peptide was involved in several different cross-linking relationships, after off-line cleavage the released peptide accumulated under one chromatographic peak. The abundance of the specific released peptide was increased, which facilitated the identification. In addition, the samples produced from the off-line cleavage experiments did not require any further sample treatment and they were analyzed directly in the mass spectrometers. Therefore no additional sample loss was incurred in the off-line photocleavage step. The on-line fragmentation spectra were then compared with those obtained from the off-line cleavage experiment as the manual verification of the released peptides. A general workflow of the mass spectrometric analyses of the pcPIR technology is shown in Figure 3a, and the four-step LC-mass spectrometric identification-and-verification of the cross-links is shown in Figure 3b. With improved instrument control software, future PIR analyses to identify the cross-linking relationships and the released peptides will be combined into one run.

An example inter-cross-linked peptide pair identified from *E. coli* is shown in Figure 4 to demonstrate the overall workflow. In the first LC-MS experiment, specific mass relationships resultant from alternating UV cleavage were found with *X-links* (Figure 4a). The neutral masses of the m/z peaks of 472.767²⁺, 762.409²⁺ and 1084.467¹⁺ (the reporter) sum to match the mass of the m/z peak of 710.964⁵⁺ with an error of 0.5 ppm (0.002 Da), which indicates that ions of 472.767²⁺ and 762.409²⁺ are possible released peptides and 710.964⁵⁺ is a putative inter-cross-linked peptide pair. Such indication was supported by the patterns observed in EIC traces (Figure 4b). The EIC of the inter-cross-link parent ion is complementary to those of the released peptides and the reporter. For example, the parent peak has higher intensities in the laser off scans and lower intensities in the laser on scans, while the released peptides and the reporter are opposite. The on-line photocleavage efficiency and on/off duty cycle have been discussed in a previous report¹⁶.

In addition, the EIC profile of a cross-linked peptide pair from the *E. coli* cross-linking experiment is compared with the EIC of a standard peptide Angiotensin from a separate LC-MS/MS run (Figure 4c). The EIC profile of the non-reacted standard peptide is smooth and does not show specific temporal patterns. The two released peptides in Fig. 4b were identified with Mascot search of K12 *E. coli* database and both sequences originate in the protein tryptophanase/L-cysteine desulfhydrase (tnaA). pcPIR reacted internal lysine residues in these peptides as shown in the designated cross-linked sites (Figure 4d-e). The two cross-linked sites were produced in the intact protein as it existed in cells and appear in the sequence at Lys459 and Lys467, respectively. To verify the pcPIR cross-linked relationship, the cross-linked parent ion was isolated and fragmented under CID conditions. The two released peptides were found to be the major fragments (Figure 4f). The MS/MS spectra shown in Fig. 4d-e were from the “on-line cleavage and identification” run (the third LC-MS/MS). Taken together, these experiments allow identification and confirmation of the cross-linking relationship and the exact cross-linked residue. It is also worth pointing out that these same two sites in tnaA were identified as cross-linked in cells using non-photocleavable PIR technology¹³, further supporting the present identification. With this information, the two cross-linked residues were mapped on the crystal structure of tnaA (Figure 4g, PDB entry 2OQX). The distance between the two cross-linked residues was measured to be 16.7 Å from the lysine side chains, or 13.7 Å from the lysine α carbons. In cases where protein crystal structures are unknown, cross-linking results provide novel information regarding sites that are close to one another *in vivo*. In cases where inter-cross-linked peptide pairs were identified from known protein complexes, topological features of the interacting regions were studied. Most importantly, novel protein interactions with interface information were revealed with the inter-cross-links between proteins without known interactions. Such information is important in understanding the functions and mechanisms of proteins and protein complexes.

The distances between the cross-linked lysine side chains were measured and listed in Table 1 where detailed crystal structures are available. All of the distances fell below 33 Å. The cross-linking constraint of DSS has been reported from other groups by measuring the distances between the cross-linked lysine α carbons⁵. Although the theoretical length of DSS is very short (11 Å), the cross-linking constraint was observed to be 30 Å^{5, 11}, which is likely due to the preferred orientations of the cross-linked lysine residue side chains. In some cases, it is also possible that the distances between amino acid residues in solution phase differ from those measured from crystal structures. To compare with these previous results, the pcPIR cross-linking distances between lysine α carbons were also measured and listed in Table 1. All of the distances are smaller than 30 Å, which is comparable to the DSS results. As discussed elsewhere¹⁶, for protein topology studies, a shorter cross-linker can be beneficial. But for protein interaction studies, longer but flexible cross-linkers may be advantageous, because a larger range of distances can be covered by such cross-linkers.

2. *In vivo* protein interactions

Many protein complexes in *E. coli* are homomultimers and complexes are formed by packing several proteins with the same sequence together. According to a summary report of the EcoCyc database, the 1008 reactions of *E. coli*'s small molecule metabolic steps are performed by 918 enzymes, among which 354 are known homomultimers²³. With such a large portion of homomultimers in all the protein complexes, one would expect that many inter-cross-linked peptides reported in literature should arise from homomultimers. Unfortunately, unambiguous identification of such homomultimers with cross-linking can be challenging because the two cross-linked peptides may originate within a single monomer or within a dimer. However, for cases where two peptides are observed with exactly the same sequence and this sequence occurs only once within the protein, unambiguous multimer

identification can be achieved. For all homodimer cross-links reported here, the two cross-linked peptides were observed with the same sequence. Additionally, two cross-linked peptides have been occasionally observed with differing, but overlapping sequences that contain the same labeled lysine residue. These cases are also considered as an unambiguous homodimer cross-link¹³. In this report, several interesting homodimer cross-links with detailed 3D structures are discussed in detail.

One example is the 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (kdud) homodimer cross-link, which was identified from pcPIR cross-linking experiments but not in our previous PIR experiments with *E. coli*. The spectra of the cross-linked parent and the released peptide ions are shown in Figure 5a-b. The released peptide was identified both in the on-line and offline cleavage experiments. The MS/MS spectrum from on-line cleavage and the Mascot search result are shown in Figure 5c. The cross-linked lysine site is underlined on the sequence. In addition, the cross-linked peptide pair was isolated in mass spectrometer and cleaved with CID fragmentation, which is shown in Figure 5d, to yield a manual verification of the cross-linked relationship. Kdud is an NAD(H) dependent enzyme which belongs to the short-chain dehydrogenases/reductases (SDR) family involved in the D-gluconate metabolism. D-gluconate is both important for bacterial survival and pathogenicity as an important carbon and energy source, and it is critical for the colonization of *E. coli* on mouse intestine models^{24, 25}. In addition, only certain species of bacteria contain D-gluconate enzymes, which makes those proteins good targets for therapeutic developments²⁶. Therefore, studies of the structures and functions of D-gluconate enzymes can be critical to help improve drug development efforts against bacteria. However, little knowledge about the *E. coli* kdud protein complex topology exists either *in vivo* or *in vitro*. Most information currently available is resultant from similar enzymes belonging to the SDR family in other bacteria; the three-dimensional structure of *E. coli* kdud has not yet been resolved. Only recently the protein structure of Gluconate 5-dehydrogenase (Ga5DH) from *Streptococcus suis* (*S. suis*) was published²⁶. A kdud structural model derived from this *S. suis* Ga5DH (41% sequence identity homology) is therefore used here (PDB entry 3CXR). The sequences of Ga5DH from *S. suis* and kdud from *E. coli* are aligned and compared in Figure 5e. The cross-linked Lys162 site from kdud corresponds to Lys167 in Ga5DH; both residues are indicated with a red box on Fig. 5e. Figure 5f shows the published three-dimensional structure model, where the cross-linked Lys167 is marked with red color. Lys167 is known as one of several binding sites of ligands from searching of the UniProt Knowledge Base (www.uniprot.org). Other known binding sites are boxed with orange color on the sequence. Importantly, a known catalytic tetrad of Arg104-Ser150-Tyr163-Lys167 is critical for binding and orienting the substrates (Figure 5f, yellow and red), where the cross-linking site resides (Figure 5f, red). Lys167 is also conserved across different bacteria²⁶, and it is involved in the characteristic motif of the SDR family (YXXXXK, underlined on sequence). Taking all the information together, the cross-linked lysine in this case is a highly characteristic and conserved site and it is critical for the binding and catalytic activities of this enzyme. Here the cross-linking data indicates that the region where Lys167 resides is important for protein interactions. Furthermore, the cross-linking result was used to predict the complex structure of kdud. Kdud could form a dimer of dimers complex structure by similarity to the Ga5DH protein, the tetramer structural model of the latter was generated with software PISA and reported elsewhere²⁶. Here we used the docking software SymmDock²⁰ to calculate the possible dimer structures and filtered these results with the cross-linking data. The resultant complex model was then compared with the dimer portion of the Ga5DH model predicted with PISA. With the monomer model 3CXR and a symmetry order of 2, 100 dimer predictions were generated. The highest scored 10 candidates were examined with the 33 Å cross-linking constraint of pcPIR, and only the top scoring model met this requirement. The distance between the lysine side chains was 22.3 Å. Figure 5g shows this model, and the two cross-linked lysine residues are shown in red color. The

resultant dimer model which is calculated with Symmdock, is highly similar to the published dimer model which was predicted with PISA²⁶. The previously predicted tetramer model is supported by the *in vivo* cross-linking data, which helped to filter modeling results with the cross-linking constraint requirement.

The cross-linked lysine in kdud is an important site for the binding and enzyme activity. Similar cases were also observed in several other protein complex cross-links. For example, another homodimer cross-link between the two subunits of glyceraldehyde 3-phosphate dehydrogenase (GAPDH or gapA) was identified from our *in vivo* pcPIR experiments. GAPDH is known to form a homotetrameric complex²⁷. Like kdud, GAPDH is also an NAD(H) dependent enzyme and serves a variety of functional roles, including catalysis of the sixth step of glycolysis. The cross-linked sites observed *in vivo* on GAPDH are near the NAD⁺ binding region. Again, the spectra of the cross-linked parent ion, the released peptide, the MS/MS fragmentation and parent verification are shown in Figure 6a-d. The two released peptides have the same mass and sequence, which indicates that two subunits from the GAPDH homotetramer were cross-linked together. The 3D structure of the protein complex is shown in Figure 6e, and the cross-linked lysine residues (Lys 108) are shown in red color. In addition, several NAD⁺ binding sites known from searching the UniProt Knowledge Base (Lys12, 13, 34, 78 and 314) are highlighted with the yellow color on Figure 6f, which appear near the cross-linked lysine residues. In a previous *in vivo* application of a mass spectrometry cleavable PIR cross-linker, the same Lys108 was observed to cross-link to two other lysine residues (Lys124 and 192)¹³. Such observation indicates that Lys108 exhibits higher reactivity in both pcPIR and PIR experiments. Whether these two inter-cross-links were formed between two protein subunits or within the same subunit could not be determined unambiguously. The cross-linking distances are 25.5 Å and 32.1 Å when the cross-links are within one subunit. Lys124 and 192 also reside in the NAD binding region. Similar to the kdud example discussed above, the observed inter-cross-linked sites are observed in sequence regions that are important for binding and activity.

As discussed in Kuriyan and Eisenberg's insightful review²⁸, true interacting proteins are likely resultant through evolution from random interactions that give rise to some specific biological advantage, which must originate from specific partner orientation and specific topological features. The observed cross-linked peptides in mass spectrometry experiments normally require on the order of femtomoles or more to yield detectable signals. Thus, any cross-linked peptide pair observed in the mass spectrometer must have formed hundreds of millions to billions of times or more during cross-linker reaction with cells (within 10–30 min), in all cases these interactions occurred with the same orientation so as to allow the same two lysine side chains to be cross-linked. Such a high frequency of these proteins being in the same proximity, with the same orientation is a hallmark of a specific protein interaction. Many cross-linked peptides are observed where both sequences originate within the same protein, some with exactly the same sequence (which define an unambiguous homomeric interaction). This too strongly suggests cross-linking experiments can yield useful topological information on non-random, specific interactions in cells.

Conclusions

In this paper, the first application of the pcPIR technology to the *E. coli* cellular system is reported. In the pcPIR technology, the specific photocleavage of the UV sensitive bonds in pcPIR resulted in the type-specific mass relationships between the cross-linked peptide pairs and the released peptides. The released peptides were identified with the LC-MS/MS data and traditional database search. Both the cross-linking relationships and the released peptide identities were validated with mass-and-time targeted LC-MS/MS experiments. In this *in vivo* cross-linking study, a total of 1602 non-redundant released peptides were identified.

Among them, 53 inter-cross-linked peptide pairs were identified and validated. Seventeen of the intercross-links were formed between different peptides within one protein, 9 were homodimer cross-links and the other 27 pairs were between different proteins. Thirteen of the protein interactions were identified with the use of other methods, e.g. the yeast-two-hybrid approach. Approximately half of the identified pairs include intra-molecular and homodimer cross-links, which are expected to be predominant compared to inter-molecular cross-links. Further experiments will result in larger datasets and a better understanding of the composition of *in vivo* cross-linked species. Many of the cross-linked proteins have detailed 3D structures. Therefore several protein complex examples with 3D structures were discussed in detail. In the kdud example, a dimer complex was computed and it is highly similar to a published dimer model built with another docking software tool. The regions where the cross-linked sites reside also showed high binding activity. Combining with the results in our previous study, the identified cross-linking regions *in vivo* appear to have features of high solvent accessibility, binding activity and flexibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Michael W. Senko from Thermo Fisher Scientific Inc. for help with modification of mass spectrometry control software to enable laser control. This research was supported by NIH grants 7S10RR025107, 5R01GM086688 and 5R01RR023334. It was also supported by the University of Washington's Proteomics Resource (UWPR95794).

References

1. Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature*. 1989; 340(6230):245–6. [PubMed: 2547163]
2. Anderson NG. Co-immunoprecipitation. Identification of interacting proteins. *Methods in Molecular Biology*. 1998; 88:35–45. [PubMed: 9664296]
3. Gould KL, Ren L, Feoktistova AS, Jennings JL, Link AJ. Tandem affinity purification and identification of protein complex components. *Methods*. 2004; 33(3):239–44. [PubMed: 15157891]
4. Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol*. 2003; 331(2):303–13. [PubMed: 12888339]
5. Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics*. 9(8):1634–49. [PubMed: 20360032]
6. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*. 2006; 25(4):663–82. [PubMed: 16477643]
7. Jin Lee Y. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst*. 2008; 4(8):816–23. [PubMed: 18633483]
8. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreau M, Muskut B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figgeys D, Tyers M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*. 2002; 415(6868):180–3. [PubMed: 11805837]
9. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology*. 1999; 17(10):1030–2.

10. Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods*. 2001; 24(3):297–306. [PubMed: 11403578]
11. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. *Nature methods*. 2008; 5(4):315–8. [PubMed: 18327264]
12. Zhang H, Tang X, Munske GR, Tolic N, Anderson GA, Bruce JE. Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Molecular and Cellular Proteomics*. 2009; 8(3):409–20. [PubMed: 18936057]
13. Zheng C, Yang LHM, Eng JK, Tang X, Weisbrod CR, Bruce JE. Cross-linking measurements of in vivo protein complex topologies. *Molecular and Cellular Proteomics*. 2011
14. Tang X, Bruce JE. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. *Mol Biosyst*. 6(6):939–47. [PubMed: 20485738]
15. Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Analytical chemistry*. 2005; 77(1):311–8. [PubMed: 15623310]
16. Yang L, Tang X, Weisbrod CR, Munske GR, Eng JK, von Haller PD, Kaiser NK, Bruce JE. A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. *Analytical chemistry*. 2010; 82(9):3556–66. [PubMed: 20373789]
17. Weisbrod CR, Hoopmann MR, Senko MW, Bruce JE. A New Dual Linear Ion Trap Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: the Velos-FT. *J Prot Res*. 2011 Submitted 12/10.
18. Anderson GA, Tolic N, Tang X, Zheng C, Bruce JE. Informatics strategies for large-scale novel cross-linking analysis. *J Proteome Res*. 2007; 6(9):3412–21. [PubMed: 17676784]
19. Hoopmann MR, Weisbrod CR, Bruce JE. Improved Strategies for Rapid Identification of Chemically Cross-linked Peptides. *Journal of Proteome Research*. 2010 Accepted.
20. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*. 2005; 33(Web Server issue):W363–7. [PubMed: 15980490]
21. Gardy JL, Spencer C, Wang K, Ester M, Tusnady GE, Simon I, Hua S, deFays K, Lambert C, Nakai K, Brinkman FS. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res*. 2003; 31(13):3613–7. [PubMed: 12824378]
22. Guillier F, Orain D, Bradley M. Linkers and cleavage strategies in solid-phase organic synthesis and combinatorial chemistry. *Chem Rev*. 2000; 100(6):2091–158. [PubMed: 11749285]
23. Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I, Collado-Vides J, Gama-Castro S, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Bonavides-Martinez C, Ingraham J. Multidimensional annotation of the Escherichia coli K-12 genome. *Nucleic Acids Res*. 2007; 35(22):7577–90. [PubMed: 17940092]
24. Adachi O, Shinagawa E, Matsushita K, Ameyama M. Crystallization and properties of 5-keto-D-gluconate reductase from *Gluconobacter suboxydans*. *Agric Biol Chem*. 1979; 43:75–83.
25. Sweeney NJ, Laux DC, Cohen PS. Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. *Infect Immun*. 1996; 64:3504–3511. [PubMed: 8751891]
26. Zhang Q, Peng H, Gao F, Liu Y, Cheng H, Thompson J, Gao GF. Structural insight into the catalytic mechanism of gluconate 5-dehydrogenase from *Streptococcus suis*: Crystal structures of the substrate-free and quaternary complex enzymes. *Protein Sci*. 2009; 18(2):294–303. [PubMed: 19177572]
27. Yun M, Park CG, Kim JY, Park HW. Structural analysis of glyceraldehyde 3-phosphate dehydrogenase from *Escherichia coli*: direct evidence of substrate binding and cofactor-induced conformational changes. *Biochemistry*. 2000; 39(35):10702–10. [PubMed: 10978154]
28. Kuriyan J, Eisenberg D. The origin of protein interactions and allostery in colocalization. *Nature*. 2007; 450(7172):983–90. [PubMed: 18075577]

Figure 1. Structure of new generation of pcPIR cross-linker, cross-linking reaction and photocleavage reaction

1. Inter-cross-link

2. Intra-cross-link

3. Dead-end

4. Normal peptides

Figure 2. Specific mass relationships between cross-linked parents and released peptides of different cross-linking relationships

Figure 3. General experimental process of the modified pcPIR technology

3a. Experimental scheme of pcPIR technology.

3b. Four-step LC-mass spectrometric identification-and-verification of the cross-links.

Figure 4. An inter-cross-link example from tryptophanase/L-cysteine desulfhydrase (tnaA)

- a. Comparison between laser on and laser off scans: the cross-linked parent peak decreased from laser off to laser on; the released peptides and reporter were generated in the laser on scan; and the sum of neutral masses of released peptides and reporter equals the mass of cross-linked parent with an error of 0.5 ppm.
- b. Extracted Ion Chromatograms (EICs) of the tnaA cross-linking related products: cross-linked parent, released peptides and the reporter.
- c. Comparison of EICs between an identified tnaA cross-linked peptide and a non-cross-linked standard peptide Angiotensin from a separated LC-MS/MS experiment.
- d-e. MS/MS fragmentations and sequences of the two released tnnA peptides
- f. Manual verification of the tnnA cross-linking relationship
- g. Mapping the intra-protein tnnA cross-link on its crystallography structure (PDB entry 2OQX). The distance between the two lysine side chains is 16.7 Å.

Monoisotopic mass of neutral peptide Mr(calc): 1738.8560
Fixed modifications:
 Carbamidomethyl (C) (apply to specified residues or termini only)
Variable modifications: K8:
 Xlink:pcPIR_K (K)
Ions Score: 76
Expect: 1e-007

tr A4VVQ2 A4VVQ2_STRSY	SUISSTRAINZYHGNSSUPESVMNQFSLDQFSLKGKIALVTGASYGIGF	50
sp P37769 KDUD_ECOLI	-----MILSAFSLGKVAVVTGCDTGLGQ	24
	: * . ***:***:***. *:*	
tr A4VVQ2 A4VVQ2_STRSY	AIASAYAKAGATIVFNDINQELVDRGMAAYKAAGINAHGYVCDVTDEDGI	100
sp P37769 KDUD_ECOLI	GMALGLAQAGCDIVG--INIVEPTETIEQVTALGRRFLSLTADLRKIDGI	72
	.:* . *:*** ** ** : . * * : . ***	
tr A4VVQ2 A4VVQ2_STRSY	QAMVAQIESEVGIIIDILVNNAGIIRRVPMIEMTAAQFRQVIDIDLNAPFI	150
sp P37769 KDUD_ECOLI	PALLDRAVAEFGHIDILVNNAGLIIRREDALEFSEKDWDDVMNLNIKSVFF	122
	:: : :: * *****:*** :*: : :*: :*: : *	
tr A4VVQ2 A4VVQ2_STRSY	VSKAVIPSMIKKGHG-KIINISMMSELGRETVSAYAAAKGGLKMLTKNI	199
sp P37769 KDUD_ECOLI	MSQAAAKHFIAQGNGGKIINIASMLSFQGGIRVPSYTASKSGVMGVTRLM	172
	:*: . : * :*: * *****:*** * :*: :*: .*: :*: :	
tr A4VVQ2 A4VVQ2_STRSY	ASEYGEANIQCNGIGPGYIATPQTAPLRELQKDGSRHPFDQFIIAKTPAA	249
sp P37769 KDUD_ECOLI	ANEWAKHNINVNAIAPGYMATNNTQQLRADEQR-----SAEILDRIAPAG	216
	*.: : * : * :*:***: * * * : : : * :*: :*	
tr A4VVQ2 A4VVQ2_STRSY	RWGEAEDLMGPAVFLASDASNFVNGHILYVDGGILAYIGKQPEA	293
sp P37769 KDUD_ECOLI	RWGLPSDLMGPVFLASSASDYVNGYTIADVGGWLLAR-----	253
	*** .***** *****:***: : * * * *	

e

f

Dimer model

Figure 5. Homodimer inter-cross-link of kdud

- a. Inter-cross-linked parent spectrum of kdud peptide pair
- b. Released peptide spectrum from kdud
- c. On-line MS/MS and Mascot search result of the kdud released peptide
- d. Manual verification of the kdud cross-linking relationship
- e. Alignment of kdud from *E. coli* and Ga5DH from *S. suis*. Red box of lysine residues are cross-linked lysine sites. Orange boxed residues are binding sites from UniProt Knowledge Base. Red under lined sequence (including cross-linking site) is the characteristic motif of SDR family. * indicates conserved residue sites, : shows conservative substitution, and . shows semi-conserved sites.
- f. Kdud monomeric model. The cross-linked lysine is marked red, and the other three residues in the catalytic tetrad Arg104-Ser150-Tyr163-Lys167 are marked yellow.
- g. Homodimer kdud protein model docked with PDB entry 3CXR and software SymmDock. The two red residues are the two cross-linked lysine. The cross-linking distance resultant from this model structure was found to be 22.3 Å between lysine side chains.

Figure 6. Homodimer inter-cross-link of GAPDH

- a. Inter-cross-linked parent spectrum of GAPDH peptide pair
- b. Released peptide spectrum from GAPDH

- c. On-line MS/MS and Mascot search result of the GAPDH released peptide
- d. Manual verification of the GAPDH cross-linking relationship
- e. GAPDH tetramer complex crystallography structure and the homodimer cross-link
- f. The cross-linked lysine residues are marked red and the other NAD⁺ binding sites are marked yellow

Table 1

Summary of the *in vivo* inter-cross-links from *E. coli*.

parent mass	released peptide mass	peptide sequence	Expect score	gene name	Distance between K side chains (Å)	Distance between α carbons (Å)	protein
3350.643	1327.615	YADMLAMSAKK	0.084	tnaA	27	21.9	gi 90111643
	939.539	VIEPVKR	0.0086	tnaA			gi 90111643
3350.648	1522.824	GLTFTYEPKVLRL	0.00061	tnaA	18.5	12.5	gi 90111643
	744.369	DPKTKGK	0.042	tnaA			gi 90111643
3398.767	1570.914	GAEQIYIPVLKK	0.00035	tnaA	31.5	24.3	gi 90111643
	744.375	DPKTKGK	0.042	tnaA			gi 90111643
3549.808	943.513	HFTAKLK	0.089	tnaA	16.7	13.7	gi 90111643
	1522.803	GLTFTYEPKVLRL	0.00061	tnaA			gi 90111643
3597.897	1570.914	GAEQIYIPVLKK	0.00035	tnaA	19.7	23.5	gi 90111643
	943.521	HFTAKLK	0.089	tnaA			gi 90111643
3649.699	1073.535	EQEKGLDR	0.0051	tnaA	15	14.2	gi 90111643
	1492.723	KYDIPVMDSAR	0.00019	tnaA			gi 90111643
3676.762	1649.765	KDAMVPMGGLLCMK	0.027	tnaA	29.3	29.6	gi 90111643
	943.513	HFTAKLK	0.089	tnaA			gi 90111643
3777.930	1522.823	GLTFTYEPKVLRL	0.00061	tnaA	16.7	13.7	gi 90111643
	1171.630	HFTAKLKEV	0.055	tnaA			gi 90111643
3914.029	1887.010	AVEIGSFLGRDPKTKG	0.004	tnaA	32.9	16.6	gi 90111643
	943.513	HFTAKLK	0.089	tnaA			gi 90111643
4177.210	1570.911	GAEQIYIPVLKK	0.00035	tnaA	11.3	15.2	gi 90111643
	1522.817	GLTFTYEPKVLRL	0.00061	tnaA			gi 90111643
4188.117	1522.803	GLTFTYEPKVLRL	0.00061	tnaA	15.3	12.1	gi 90111643
	1581.819	TGKQLPCPAELLR	0.0011	tnaA			gi 90111643
4493.328	1887.031	AVEIGSFLGRDPKTKG	0.004	tnaA	18.5	12.5	gi 90111643
	1522.819	GLTFTYEPKVLRL	0.00061	tnaA			gi 90111643
4541.422	1887.021	AVEIGSFLGRDPKTKG	0.004	tnaA	31.5	24.3	gi 90111643
	1570.914	GAEQIYIPVLKK	0.00035	tnaA			gi 90111643

parent mass	released peptide mass	peptide sequence	Expect score	gene name	Distance between K side chains (Å)	Distance between α carbons (Å)	protein
3491.763	1265.662	SKATNLLYTR	0.047	dps	10.5	13.5	gi 16128780
	1142.630	KATVELLN	0.017	dps			gi 16128780
4520.434	2294.321	AVQLGGVALGTTQVINSKTPPK	2.50E-08	dps	12.6	17.4	gi 16128780
	1142.634	KATVELLN	0.017	dps			gi 16128780
3578.861	1540.813	EAKDLVESAPAALK	0.0034	rpIL	14.2	10.8	gi 16131816
	954.590	VAVIKAVR	4.00E-05	rpIL			gi 16131816
4734.284	2229.033	VDSSMINKVGNFMDDSAITAK	2.20E-07	osmY			gi 16128715
	1421.760	VKAALVDHNDIK	0.0073	osmY			gi 16128715
2932.480	924.503	KHITAGAK	0.00083	gapA			gi 16129733
	924.503	KHITAGAK	0.00083	gapA			gi 16129733
2970.510	943.513	HFTAKLK	0.089	tnaA			gi 90111643
	943.513	HFTAKLK	0.089	tnaA			gi 90111643
3172.519	1053.524	VGFYGYGKAR	0.03	rpsE			gi 16131182
	1035.512	KFISIEAE	0.059	rpmA			gi 16131075
3271.560	1161.567	LAKEDPSFR	0.062	fusA			gi 16131219
	1026.509	TSGEKHLR	0.0043	rpmF			gi 16129052
3336.572	1187.578	KISNGEGVER	0.00023	rpIS			gi 16130527
	1065.528	KLFSGMQR	0.016	rodZ			gi 16130441
3349.561	1153.540	VYKNYDPR	0.0067	gluA			gi 16128695
	1112.525	QAKGYYGAR	0.0038	rpIT			gi 16129672
3385.597	1252.609	KNIEFFEAR	0.054	rpII			gi 16132025
	1049.518	ATKLTMINR	0.041	purT			gi 16129802
3424.594	1179.559	MGKTYQQPK	0.011	ypfJ			gi 16130400
	1161.567	LAKEDPSFR	0.062	fusA			gi 16131219
3454.605	1197.603	QHVIYKEAK	0.0038	rpmG			gi 16131507
	1173.563	EAAGSALKGDR	0.0086	hemB			gi 90111123
3473.692	1284.650	TMKAQQPPR	0.0029	pheS			gi 16129670
	1105.566	FSVEAPKTK	0.012	rpID			gi 16131198
3478.675	1252.609	KNIEFFEAR	0.054	rpII			gi 16132025

parent mass	released peptide mass	peptide sequence	Expect score	gene name	Distance between K side chains (Å)	Distance between α carbons (Å)	protein
	1142.630	KATV L LN R	0.017	dps			gi 16128780
3580.780	1343.693	LN T LSPA E GS K K	0.046	rpIO			gi 16131180
	1153.650	TG K AA R	0.012	rpIS			gi 16130527
3611.683	1053.524	VG F GYG K AR	0.03	rpsE			gi 16131182
3632.743	1492.723	KYDIPV V MD S AR	0.00019	tmaA			gi 90111643
	1274.626	YAPNA K DLA G R	0.00011	sdhA			gi 16128698
	1274.626	YAPNA K DLA G R	0.00011	sdhA			gi 16128698
3634.785	1343.693	LN T LSPA E GS K K	0.046	rpIO			gi 16131180
	1207.595	ASDPAN H L K R	0.0074	gstB			gi 16128798
3636.658	1276.597	EIAEK M VE G R	0.00032	tsf			gi 16128163
	1276.597	EIAEK M VE G R	0.00032	tsf			gi 16128163
3638.739	1277.662	TLAASG I KD F R	0.0033	fabI			gi 16129249
	1277.662	TLAASG I KD F R	0.0033	fabI			gi 16129249
3724.738	1320.641	STDISV K TD Q K	0.0000015	osmY			gi 16132194
	1320.641	STDISV K TD Q K	0.0000015	osmY			gi 16132194
3740.835	1511.751	RTAEICE H L K R	0.0081	glpK			gi 16131764
	1145.593	AAV K SGSEL G K	0.012	adk			gi 16128458
3802.7510	1472.736	LE K GEDLEA T IR	0.00026	tdcE			gi 49176316
	1246.594	NSG K FNPL D R	0.02	tolB			gi 16128715
3866.783	1073.499	EQE K G L DL R	0.0051	tmaA			gi 90111643
3871.871	1709.858	ANITV N KNSVP N DP K	0.00078	glyA			gi 16130476
	1844.937	ITI K ASSGL N EDEI Q K	0.0000054	dnaK			gi 16128008
	943.5127	HFTA K L K	0.089	tmaA			gi 90111643
3958.933	1437.735	FGA K SISTIAES K	0.0025	gadB			gi 16129452
	1437.735	FGA K SISTIAES K	0.0025	gadB			gi 16129452
4007.987	1472.772	GVETA D K V L K GE K	0.0053	rbsB			gi 16131619
	1451.726	AEAPAA A PA A K A EG K	0.00043	aceF			gi 16128108
4016.842	1796.963	L K GN T GENT L LAL L EG R	0.00079	rpsD			gi 16131175
	1136.557	GLSA K SFD G R	0.075	rpIE			gi 16131187

parent mass	released peptide mass	peptide sequence	Expect score	gene name	Distance between K side chains (Å)	Distance between α carbons (Å)	protein
4041.048	1478.773	QASLLKTNVYVS R	0.0039	mdtE			gi 16131385
	1478.773	QASLLKTNVYVS R	0.0039	mdtE			gi 16131385
4090.919	1252.609	KNIEFF E AR	0.054	rpII			gi 16132025
	1754.851	VPSYTA S KSGVMGV T R	6.50E-08	kduD			gi 16130746
4163.049	1192.591	IAFVN K MDR	0.00025	fusA			gi 16131219
	1887.010	AVEIGSFLGRDPK T GK	0.004	tnaA			gi 90111643
4419.151 [#]	1913.949	EIPMRPGQLFMDP K R	0.024	gadB			gi 16129452
	1913.949	EIPMRPGQLFMDP K R	0.024	gadA			gi 16131389
	1421.751	VKAAALVDH D NIK	0.0073	osmY			gi 16132194
4433.221	2462.252	DMALLGKALIHDPVEYAI H K	0.005	Dacc			gi 16128807
	887.482	LSEK R R	0.043	rnhB			gi 16128176
4462.116	1414.705	NLTGKEADAALGR	0.00016	glyA			gi 16130476
	1964.004	HILLKPSPI M TDEQAR	0.00091	surA			gi 16128047
4561.212	1738.856	VPSYTA S KSGVMGV T R	0.0000022	kduD			gi 16130746
	1738.856	VPSYTA S KSGVMGV T R	0.0000022	kduD			gi 16130746
4622.279	1290.634	AF A ENWL G KR	0.0031	sucC			gi 16128703
	2248.170	VDIITGTLGKALGGASGGY T AAR	1.30E-07	kbl			gi 16131488
4632.266	1290.636	AF A ENWL G KR	0.0031	sucC			gi 16128703
	2258.183	KYDFSTPYTISGIQALV K K	0.066	fliY			gi 16129867
5268.632	2517.250	LDLNPITGTPFQLQQYQ K DSR	0.0021	dppA			gi 16131416
	1667.884	QAGELQE K LIAVNR	0.0000069	rpsE			gi 16131182
5325.538	1154.593	AKVNNVD P AK	0.064	pta			gi 16130232
	3087.480	ERGEGFQQA V AAHKFNVLASQP A DEF D R	0.00007	rbsB			gi 16131619

Underlined K: pcPIR labeled Lysine residues.

Underlined M: oxidized Methionine residues.

[#] One of the released peptides can come from either of the two reported proteins. The two proteins have the same sequence of the released peptide region.

BOLD font: Interactions identified with other technologies by searching *E. coli* Interaction Database (Ecid)