A new cross-linking strategy: protein interaction reporter (PIR) technology for protein–protein interaction studies

Xiaoting Tang[a] and **James E. Bruce**[*,b]

[a]Novo Nordisk Inflammation Research Center, Seattle, Washington, USA

[b]Department of Genome Sciences, University of Washington, Box 358050, Seattle, Washington 98195, USA

Abstract

Chemical cross-linking coupled with mass spectrometry, an emerging approach for protein topology and interaction studies, has gained increasing interest in the past few years. A number of recent proof-of-principle studies on model proteins or protein complex systems with improved cross-linking strategies have shown great promise. However, the heterogeneity and low abundance of the cross-linked products as well as data complexity continue to pose enormous challenges for large-scale application of cross-linking approaches. A novel mass spectrometry-cleavable cross-linking strategy embodied in Protein Interaction Reporter (PIR) technology, first reported in 2005, was recently successfully applied for *in vivo* identification of protein–protein interactions as well as actual regions of the interacting proteins that share close proximity while present within cells. PIR technology holds great promise for achieving the ultimate goal of mapping protein interaction network at systems level using chemical cross-linking. In this review, we will briefly describe the recent progress in the field of chemical cross-linking development with an emphasis on the PIR concepts, its applications and future directions.

Introduction

Protein–protein interactions are key determinants of critical cellular function and serve to orchestrate functionality in space and time.[1,2] Identification and analysis of protein–protein interactions are, therefore, critical to the assignment and comprehension of protein function within biological systems. Many techniques currently allow acquisition of information on protein interactions, such as the yeast two-hybrid,[3-5] *in vivo* FRET,[6] immunoprecipitation (IP),[2] and IP through affinity tag such as FLAG tag[7] and TAP tag.[8] However, most of these approaches generally require system perturbation such as genetic modification of the native cells or disruption of the original cellular context, which leads to both false positive and false negative results.[9,10]

Chemical cross-linking strategies have been pursued for many years with the goal to fulfill two primary needs in proteomics research: (i) identification of protein interaction network and (ii) low-resolution mapping of protein and protein complex structures. Cross-linking has been widely applied in protein–protein interaction studies due to the inherent potential these approaches hold for stabilizing *bona fide* interactions and freezing transient or labile interactions with covalent bonds.[11,12] Furthermore, *in vivo* applications of chemical cross-linking coupled with immunoprecipitation and affinity tag techniques have significant

[*]jimbruce@u.washington.edu; Fax: +1 206-616-0008; Tel: +1 206-543-0220 .

advantages and have been extensively reported[2,13-17,18-24] and reviewed.[11,12,25-29] Chemical cross-linking also has significant advantages for structural characterization of individual purified proteins and protein complexes[11,27-30] where the identified cross-linked residues/ peptides provide distance constraints that help define 3-D structural models. However, identification of cross-linked peptides is not trivial even for purified protein complexes which may be available in large quantity. In this regard, the challenges associated with cross-linking approaches arise primarily from three aspects: (i) enormous complexity inherent in the cross-linked samples which contain predominantly unmodified peptides, inter-cross-linked peptides, intra-cross-linked peptides, dead-end labeled peptides (nomenclature of cross-linked peptides was reported by Schilling *et al.*[31]), multiple-labeled peptides, and non-specific labeled peptides; (ii) low abundance of the cross-linked species; and (iii) data complexity of MSMS spectra which may contain fragment ions from both peptides as well as fragments from modified versions of each peptide where the peptide ions are observed with partial or full sequence of the other peptide still attached through the cross-link. Significant effort has been dedicated to designing novel cross-linkers that allow improved capability for cross-linked peptide identification *via* a signature pattern in the data or reduction of sample complexity by enrichment. Alternative efforts have also been applied to informatics software development with the use of the conventional cross-linkers. We have pursued concepts embodied in a novel class of cross-linkers called Protein Interaction Reporters (PIRs). In this review, we begin by discussing recent progress on cross-linking development, specifically targeting the area of protein interaction identification. Next we introduce the unique attributes of the PIR approach and describe how we incorporate novel experimental and data analysis strategies to the PIR technology for enabling large-scale profiling of protein interactions in living cells and recent applications in a model bacterial cell system. Finally, we finish with a forward-looking prospective describing where the PIR and similar developing strategies will likely go and problems these approaches will be able to address.

Cross-linking principles, applications, and potentials

A cross-linker by definition is simply a chemical reagent with two or more reactive groups connected by a spacer or linker region. The selection of the reactive groups depends on the target molecules to be cross-linked. For the purposes of labeling proteins, the reactive ends can be amine-, sulfhydryl-, or photo-reactive. The length of the spacer chain is often used as a "ruler" for estimating the distance of the two linked residues, providing approximate topological information on proteins or protein complexes. Many structural variants of the basic, conventional cross-linker (two reactive groups and a simple spacer chain) are commercially available. Pierce Biotechnology (Rockford, IL) has been a leading source in this area and provides a wide variety of cross-linkers and technical publications with a highly useful summary and guidance on cross-linker selection.[32,33]

Although in principle many reactive groups such as maleimides, diazoacetate-esters, or azides can be implemented in cross-linkers, *N*-hydroxysuccinimide (NHS) esters have been the most frequently-used reactive group[27] due to multiple factors. First, NHS esters form stable amide linkage to the primary amines in proteins at physiological condition (pH 7.0—7.5), which is important for *in vivo* labeling. Second, 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 frequency of lysine residues in most proteins. Third, the NHS ester reaction has very rapid reaction rates with primary amines.[34] Our studies[35] showed that a 5 min reaction time was sufficient for completion of labeling of intact cells at pH near 7. 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. Finally, lysine or N-terminus modification can be used as additional search

constraint with software tools. In that regard however, it is worth noting that recently several research groups have observed that NHS esters do not react exclusively with primary amines and other groups including serines, tyrosines, and threonines also show significant reactivity.[36-39] The Sinz research group reported that about 12.5% of overall cross-linker containing products resulted from serine labeling, 4.3% from tyrosine, and 3% from threonine when BS^3 and BS^2G were used as cross-linkers.[38] Therefore, cross-link assignment using software tools should also take into account possible modifications of Ser, Tyr, and Thr residues.

Since the span of the cross-linker has been used for providing topological measurement for the low-resolution studies of protein three-dimensional structures, small-sized or shorter cross-linkers are often considered to be useful for obtaining structural constraint information. However, our previous studies using a PIR cross-linker with the calculated maximum length of 43 Å showed identification of a cross-link between the two lysine residues in the distance of 14 Å in ribonuclease S.[40] Furthermore, this same cross-link was also reported later by others using the smaller-sized cross-linkers.[41,42] Our recent studies[43] using PIR cross-linkers for labeling several protein standards resulted in identification of many intra-protein cross-links which were reported by others using DSS, a small cross-linker with 11 Å spacer chain.[43,44] These results suggest that in solution, the cross-linkers are 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.*[45] 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 wider range of distances is most critical to enable identification of larger numbers of interactions. The latter point highlights a significant difference in the overall goals for cross-linker designs that target protein interactions and those of more conventional cross-linking efforts that aim to provide actual distance constraints on purified proteins and complexes. However, 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,[46,47] analytical techniques such as co-crystallization, and computational methods.[48,49] For example, site-directed mutagenesis has been employed by Wells, to map critical structural features for selected interactions *in vivo* and identify "hot spots" for these selected interactions.[50] Thus, chemical cross-linking can be considered as potentially fulfilling two separate and distinct needs in proteomics research: protein interaction network identification and protein and complex structural measurements. Both are important and both represent unique opportunities to provide critical information on biological systems.

For cross-linking application and protein interaction studies, this concept can be further considered as two distinct areas (Fig. 1) since the challenges associated and potential utility from either area vary greatly. The first involves the identification of interactions with a given specific protein target. The second involves the large-scale determination of interacting proteins. Chemical cross-linking coupled with the IP approach has been consistently employed for targeted interactions for several years. The observation of higher molecular weight bands in the SDS-gel image after cross-linking and IP often suggests the detection of interaction. For identification of these interactions, the analysis and identification of two or more unique tryptic peptides is generally sufficient for protein identification[51,52] as in other approaches such as co-IP or TAP-tag. In fact, a significant advantage of cross-linking approach over other techniques is that interactions can be stabilized with covalent bonds in cells or under conditions more likely to preserve native, weak, and transient interactions prior to cell lysis; thus identified interactions may be more physiologically relevant. This approach is highly useful for membrane protein interactions

since effective solubilization of these species normally requires conditions that are likely non-conducive to preservation of protein interactions. In principle, cross-linking methods can be applied on a large scale without the need of IP, since the detection of the cross-link itself is sufficient to indicate interactions. Importantly, the identification of cross-linked peptides not only suggests protein interactions but also provides the information on *how* proteins interact *in vivo, i.e.*, provides critical data on the sites of close proximity of X and Y. In this regard, largescale protein interaction studies that are based on identification of cross-linked peptide pairs resultant from reactive protein sites with close proximity during cross-linking reaction (either within a single protein or two proteins that form complex) are highly specific. In contrast, the IP approach is susceptible to detection of non-specific binding of abundant proteins to the target protein, antibody and solid support. As such, cross-linking offers a complimentary approach to conventional co-IP studies and potentially may yield increased specificity of the detected interactions. Furthermore, highly abundant proteins that are labeled during cross-linking reaction and compose a majority of the detected cross-linked peptides during a single analysis can be removed by affinity selection performed with cross-linked samples, allowing detection of less abundant cross-linked peptides and greater dynamic range. Finally, structural models can utilize these data to enable improved comprehension of identified interactions and identify factors that cause binding or increase binding specificity. For interactions that give rise to undesirable functional consequences (*e.g.*, release of inflammatory cytokines upon ligand/receptor binding) and thus, might be targeted for disruption or blocking, identification of interacting regions can conceptually lead to more rapid development of molecules that can inhibit binding with desired specificity. In this regard, the actual interfacial region is impossible to retrieve with all other large-scale protein interaction technology, and yet, is perhaps the most critical piece of information related to protein interactions in biological systems.

The challenges associated with the identification of the cross-linked peptides have prevented large-scale application. To improve the capability of cross-link identification, there have been two primary strategies, *i.e.*, reducing the sample complexity through enrichment of cross-linked peptides and reducing the data complexity by introducing signature pattern in cross-link containing spectra. For the purpose of reducing sample complexity and data complexity, much work has centered on modification of the spacer chain of the cross-linker, which includes incorporating an affinity tag,[53-58] differential isotope label,[41,59-62] chemically cleavable bonds,[61,63-65] and mass spectrometry-cleavable bonds.[40,42,66-69] Implementation of isotope patterns in the data has also been achieved through incorporation of isotope labels in peptides *via* proteolytic digestion using ^{18}O water[56,70,71] or in proteins with ^{15}N labeling.[72] Recently, the Brodbelt group employed an IR chromophore group in the spacer chain and demonstrated that cross-linked peptides were more prone to IRMPD fragmentation thus allowing peptide identification from unmodified product ions.[73] Additionally, several groups have initiated efforts on informatics software development to identify cross-linked peptides with modified database search strategies.[41,74,75]

Despite significant advancement in cross-linking technology, further proteome-wide *in vivo* application of cross-linking strategies for studying protein interactions with topology information has remained challenging due to the complexity of both the database and the possibility of multiple cross-link product types. For example, a digest mixture with n peptides can lead to n^2 possible cross-linked products and the desired inter-cross-linked peptides are less than 0.1% of the total theoretical peptide combinations.[41] With proteome-wide consideration, this becomes intractable for most interesting biological organisms in the absence of additional information that can help decipher reaction product type and protein identity.[12] Furthermore, a living organism is a dynamically changing system, in which many protein–protein interactions are transient both in time and space. As a consequence, mapping protein–protein interactions and their interaction topology in the native cellular milieu will

provide most relevant and crucial information for understanding biological function. Therefore, a method that is well-suited to the challenging task of large-scale determination of protein interaction networks will likely benefit from some type of reporter molecules that can infiltrate living systems and provide analytically-useful feedback to allow visualization of these networks.

PIR concept

We have undertaken chemistry development efforts to further advance chemical and photochemical cross-linking strategies that can be combined with mass spectrometry to provide largescale information on protein interactions and topological features in living cells. One of the biggest hurdles for large-scale application with traditional cross-linkers is the difficulty in spectral interpretation which prevents peptide/protein identification. In general, during MSMS of the cross-linked peptides, fragments are generated from the cleavage of both peptide chains and cross-linker itself, which results in a highly complex spectrum precluding protein identification with the use of the conventional database search algorithm. To simplify the situation, one can envision that an ideal cross-linker should allow measurement of the cross-linked peptide complex first and then dissociation of the cross-linked complex to two intact peptide chains which can then be measured and sequenced individually with commonly-used database search algorithms such as Sequest and Mascot. Efforts with chemically cleavable cross-linkers such as DSP and DTSSP have been attempted and these applications using differential MALDI mass mapping between the spectra of before and after thiol-cleavage demonstrated utility for a simple pure protein complex measurements.[63] However, in cases where LC is necessary for fractionating the digest mixture from two or more cross-linked protein complexes, the relationship between the cross-linked complex precursor and its released intact peptides becomes intractable. Furthermore, the off-line sample handling with chemical cleavage and purification steps could result in additional sample loss.

In 2005, we reported our efforts to develop a new class of cross-linker, Protein Interaction Reporter (PIR),[40] which can be cleaved directly in the mass spectrometer to release the two individual peptide chains. A key enabling factor in the PIR cross-linker is the introduction of the labile bonds in the cross-linker (Fig. 2a). The labile bonds in the cross-link can be specifically cleaved *in situ* to release two intact peptide chains. Each peptide chain can then be sequenced separately and identified either by MSMS fragmentation or accurate mass measurement through conventional database search algorithm. One advantage of this approach is that cleavage allows release and subsequent mass analysis of the intact peptide, allowing the peptide precursor mass to be used in the identification step as is commonly employed in standard protein identification experiments. The PIR cross-linker design is modular to allow tunable chemistry to be applied to PIR structure in terms of labile bonds, reaction groups, and affinity groups. First, either one or two labile bonds can be included in the PIR cross-linkers. The labile bonds can be dissociated *in situ* in the mass spectrometer by CID, ECD, IRMPD, or other dissociation mechanism; or cleaved with photo activation in-line after LC separation and prior to the entrance of mass spectrometer. When two labile bonds are included, a signature ion which we call reporter is specifically released along with the two intact peptide chains, providing important basis for informatics development for large-scale application which will be discussed below. Additionally, two cleavable groups yield a common, known residual modification on the peptides after release, simplifying the database search strategy. We employed the Rink structure[76] for the MS labile bonds in our initial attempt (Fig. 2c).[40] Alternative groups/structures can also be investigated for low-energy MS cleavage functionality. One example is the use of peptide bond between aspartic acid and proline (D–P) which has shown more susceptible to cleavage by CID as compared with other peptide bonds. Incorporation of one D–P bond by Soderblom and Goshe[67] and

two D–P bonds in the cross-linkers by us (Fig. 2d)[77] have both shown the desired low-energy cleavage of the cross-linkers during CID activation. Building on the mass spectrometry cleavable concept, Petrotchenko *et al.* recently introduced a mass spectrometry-cleavable cross-linker, BiPS, which can be dissociated by the MALDI laser during analysis.[42] Lu *et al.* employed a sulfonium ion in the spacer chain and the C–S bond showed favored cleavage upon CID MSMS. All these applications demonstrated the advantages of cleaving cross-linked peptides in the mass spectrometer directly to release intact peptides for further MSMS fragmentation and peptide identification. To further explore the versatility of PIR chemistry, we recently incorporated two photo-cleavable bonds in PIR cross-linkers.[43] These bonds are cleaved by a UV laser in-line after LC fractionation and before introduction to the mass spectrometer. Since the cleavage occurs after LC fractionation, the cross-linked peptide complex "co-elutes" with the released intact peptides and thus the relationship of the precursor and product can be tracked as is normally done with CID-cleavable PIR samples. The additional advantage using the photo-cleavable PIR is that the intact peptides are released in solution instead of gas phase and thus, the peptides can be observed with multiple charge states which results in more comprehensive MSMS fragmentation.

As discussed above, the general strategy to improve cross-link identification is to reduce data and sample complexity. Dissociation of the cross-linked peptide complex *in situ* allows simplified data interpretation. For reduction of the sample complexity, an affinity tag can be included for enrichment of low-abundant cross-linked peptides. Our initial attempt employed a biotin group in the PIR structures (Fig. 2c and d). However in principle, other groups such as an alkyne group which can form covalent bonds with an azide group conjugated solid support (the reaction is referred as "click chemistry"[78,79]), or any peptide or small molecular tag that can be used for immunoprecipitation could be coupled to allow enrichment.

PIR informatics and *in vivo* application

PIR technology incorporates two labile bonds that can be cleaved with high specificity within the mass spectrometer to allow specific activation and release of the central core portion of the cross-linker or reporter ion along with the intact peptide ions. These released species can then be further identified by MSMS or accurate mass measurements.[40,66] As a consequence, the identification of the cross-links using the informatics tools is a two-step process, *i.e.*, (i) identifying which ions contain the cross-link and the type of cross-links (inter- and intra-cross-linked as well as dead-end labeled) and (ii) identifying the peptide sequence and its protein of origin. For the first step, the reporter ion can be used to focus informatics analysis on regions of the data that contain the expected cleavage products. Furthermore, the predictable mass relationships (Fig. 2b) between the released peptide ions and the measured cross-linked products can be used for assignment of PIR cross-link types from spectra that may contain many different PIR reaction products. Thus, the engineered mass relationships resultant from the PIR concept enable multiplexed measurements of cross-linked complexes and released peptides, since one can associate released peptides with specific cross-linked products even in complex spectra. A custom search algorithm, X-links, was developed to utilize PIR mass relationships and enable identification of PIR-labeled products and assign cross-link types.[80] Because the masses of the cross-linked products and released peptides can be measured with high mass measurement accuracy, multiplexed activation of a collection of PIR-labeled peptides can be carried out simultaneously. For example, a false discovery rate (FDR) less than 6% for cross-link pair assignment was estimated with Monte Carlo simulations with random peptide masses, spectral complexity up to 100 components and modest mass measurement accuracy of 5 ppm.[80] Further improvements in mass measurement accuracy allow either a lower FDR at equal spectral

complexity or higher spectral complexity with an equal FDR. Most importantly however, PIR technology allows the determination of the mass of each peptide from a cross-linked complex and subsequent identification of each peptide using conventional mass spectrometric methods. An example of an identified inter-cross-linked peptide complex from multiplexed MSMS FTICR mass spectra with the activation off (black trace) and on (red trace) is shown in Fig. 3. Once the assignment of PIR labeled products/cross-link types is established, the subsequent task is to identify the peptide sequence and its protein of origin. The capability of analysis of the individual peptides by dissociation of the labile bonds allows unambiguous identification of cross-linked proteins in a manner that is unique to PIR technology. Peptide/protein identification can be achieved *via* database search algorithm either using the accurate masses of the released peptides or fragmentation patterns of the peptides. Fig. 4 shows an example of CID MSMS spectrum of a released peptide from a PIR-labeled product.

We have recently published the first-ever demonstration of unbiased identification of protein interactions from cross-linking experiments with intact cells using PIR technology.[57] These results were acquired by cross-linking of live cells, lysis to extract proteins, digestion and affinity capture of cross-linked peptides followed by LC/MS analysis. Importantly, these results demonstrate that protein interactions in cells can be identified with PIR technology and MS analysis on the cross-linked peptide pairs. During LC/MS analysis, intact cross-linked peptide complexes were alternately measured and then activated to release the cross-linked peptides within the mass spectrometer. Accurate peptide masses were then used to define cross-link relationships and identify protein–protein interactions in cells. Finally, several of the protein interactions identified by the PIR approach were also identified by other means,[77,81] illustrating the feasibility of PIR application for macromolecular complex analyses in cells.

Future aspects

Looking forward it seems likely that molecular biology techniques, such as 2-hybrid approaches, TAP tags and other tagging methods, will continue to dominate the field of the large-scale protein-protein interaction discovery. However, chemical cross-linking strategies that enable identification of protein interactions present many opportunities for measurements that strongly complement these approaches. Since cross-linking can in principle be applied on nearly any biological system, these methods will offer opportunity to discover interactions in systems not readily applicable to molecular biology approaches, such as biological fluids or native cells, communities of cells, host pathogen interactions, and possibly, tissues. In addition, chemical methods offer the opportunity to provide quantitative measurement of protein interactions in these and other systems where such information is currently beyond the scope of any current technology. For example, stable isotopes can be incorporated in much the same way as with other leading tools used for quantitative proteomics, such as ICAT,[82] ITRAQ[83] and other common approaches, to yield relative quantitation of protein interactions between two or more systems. Since many human diseases are resultant from abnormal protein–protein interactions involving proteins that are normally present, proteins from pathogens or both,[84] large-scale measurements of protein abundance, even if comprehensive, may not reveal this critical level of information. A quantitative, chemical strategy for protein interaction measurement will yield information on these molecular changes that are currently beyond reach. Chemical cross-linking can conceivably be carried out with normal and diseased cells and/or tissue samples and this approach will ultimately be useful for visualization of changes in functional protein interaction networks associated with the disease which could enable better comprehension of the disease mechanism. Several routes exist to promote a quantitative PIR approach, such as incorporation of isotopes in the reporter ion region, the remaining tag region, or a

combination of isotopic labels in multiple regions of the PIR compound. An example of incorporation of isotope in the PIR cross-linker is illustrated in Fig. 5. Implementing a ^{13}C labeled lysine in the reporter region will result in a 6 Dalton difference between heavy and light PIR compounds. Incorporation of the deuterium or ^{13}C labeled succinic acid group will result in an 8 Dalton shift in observed PIR-labeled mass between the two samples. As another area of future development, *functionally*-directed PIRs can be developed for labeling specific classes of proteins and their interactions. Mechanism-based cross-linkers have been reported previously by Shokat and co-workers with an adenosine-analog cross-linker development for identification of kinase-substrate pairs.[85] Similarly, PIRs can be synthesized in a way that incorporates specific sequence motifs, tags, or structural features that can direct PIR labeling to certain classes of proteins. In this case, the presence of the cleavable groups in functionally-directed PIRs can help identify protein interactions in selected classes of protein targets *in vivo*. Such an approach may help map functional interaction pathways in cells that have remained elusive due to specific, but weak protein interactions that are difficult to purify by co-IP methods. In summary, as a chemical biology approach, PIR/cross-linking with its unique attributes holds great potential for mapping protein–protein interactions on a large-scale. With continued development, it is certain that larger numbers of complexes can be mapped *in vivo*. However, perhaps its greatest contribution will be to provide *in vivo* data on actual sites of interactions of proteins. Even if only applicable to a small subset of complexes, this level of detail is not possible with current technology, but could strongly impact the general understanding of protein complexes in live cells and yield new avenues for therapeutic intervention.

Biography
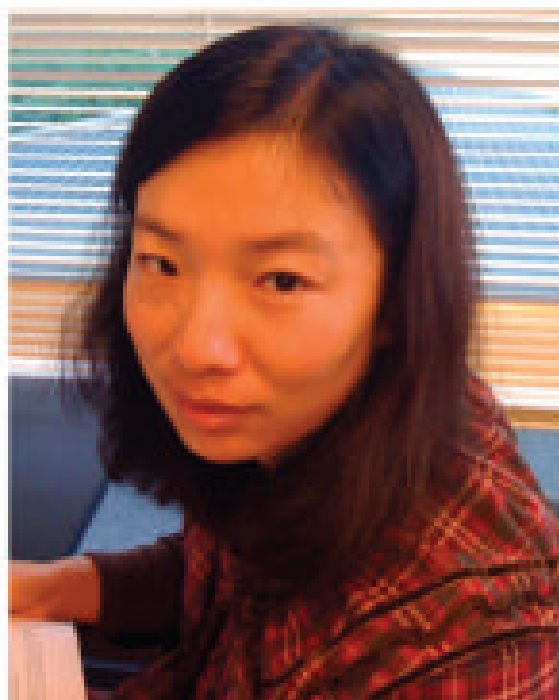
Xiaoting Tang

Dr Xiaoting Tang received an MS from Purdue University and a PhD in Washington State University. Her PhD research involved development of a novel cross-linking strategy (PIR)

that can enable largescale protein-protein interaction studies and development of hybrid instrumentation that combines atmospheric pressure ion mobility spectrometry (IMS) with FTICR/MS. Dr Tang has held positions as Associate Director of Proteomics and Biological Mass Spectrometry at Washington State University, Research Scientist in the Department of Genome Sciences in the University of Washington and currently is a Proteomics Research Scientist at Novo Nordisk Inflammation Research Center in Seattle.

James E. Bruce

James E. Bruce, Ph.D. is Professor of Genome Sciences at the University of Washington in Seattle, WA. Professor Bruce received a Ph.D. in physical chemistry from the University of Florida, held a postdoctoral researcher, staff scientist, and research fellow positions at the Pacific Northwest National Laboratory and Merck Research Labs before returning to academia. His current research interests include chemical biology and technology developments for advance proteomics research in areas including protein interactions, membrane transport processes, development of drug resistance and bioremediation.

Abbreviations

BiPS	bimane bisthiopropionic acid *N*-succinimidyl ester
BS^3	bis(sulfosuccinimidyl)suberate
BS^{2}G	bis[sulfosuccinimidyl] glutarate
CID	collision induced dissociation
DSP	dithiobis (succinimidyl propionate)
DSS	disuccinimidyl suberate
DTSSP	3,3′-dithio-bis(succinimidylproprionate)
ECD	electron capture dissociation
FTICR	Fourier transform ion cyclotron resonance
IP	immunoprecipitation
IRMPD	infrared multi-photon dissociation
MALDI	matrix-assisted laser desorption/ionization
NHS	*N*-hydroxysuccinimide
PIR	protein interaction reporter

References

1. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. Nat. Biotechnol. 2003; 21:315–318. [PubMed: 12577067]

2. Hall DB, Struhl K. J. Biol. Chem. 2002; 277:46043–46050. [PubMed: 12297514]

3. Fields S, Song O. Nature. 1989; 340:245–246. [PubMed: 2547163]

4. Li B, Fields S. FASEB J. 1993; 7:957–963. [PubMed: 8344494]

5. Yang M, Wu Z, Fields S. Nucleic Acids Res. 1995; 23:1152–1156. [PubMed: 7739893]

6. Damelin M, Silver PA. Mol. Cell. 2000; 5:133–140. [PubMed: 10678175]

7. 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, Goudreault M, Muskat 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, Figeys D, Tyers M. Nature. 2002; 415:180–183. [PubMed: 11805837]

8. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. Nat. Biotechnol. 1999; 17:1030–1032. [PubMed: 10504710]

9. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Nature. 2002; 417:399–403. [PubMed: 12000970]

10. Aloy P, Russell RB. Nat. Biotechnol. 2004; 22:1317–1321. [PubMed: 15470473]

11. Back JW, de Jong L, Muijsers AO, de Koster CG. J. Mol. Biol. 2003; 331:303–313. [PubMed: 12888339]

12. Gingras AC, Gstaiger M, Raught B, Aebersold R. Nat. Rev. Mol. Cell Biol. 2007; 8:645–654. [PubMed: 17593931]

13. Orlando V, Paro R. Cell (Cambridge, Mass.). 1993; 75:1187–1198.

14. Orlando V, Strutt H, Paro R. Methods. 1997; 11:205–214. [PubMed: 8993033]

15. Orlando V. Trends Biochem. Sci. 2000; 25:99–104. [PubMed: 10694875]

16. Kim DH, Sarbassov DD, Ali SM, King JE, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM. Cell (Cambridge, Mass.). 2002; 110:163–175.

17. Meunier L, Usherwood YK, Chung KT, Hendershot LM. Mol. Biol. Cell. 2002; 13:4456–4469. [PubMed: 12475965]

18. Kametaka S, Shibata M, Moroe K, Kanamori S, Ohsawa Y, Waguri S, Sims PJ, Emoto K, Umeda M, Uchiyama Y. J. Biol. Chem. 2003; 278:15239–15245. [PubMed: 12586838]

19. Schmitt-Ulms G, Hansen K, Liu J, Cowdrey C, Yang J, DeArmond SJ, Cohen FE, Prusiner SB, Baldwin MA. Nat. Biotechnol. 2004; 22:724–731. [PubMed: 15146195]

20. Vasilescu J, Guo X, Kast J. Proteomics. 2004; 4:3845–3854. [PubMed: 15540166]

21. Guerrero C, Tagwerker C, Kaiser P, Huang L. Mol. Cell. Proteomics. 2006; 5:366–378. [PubMed: 16284124]

22. Kim S, Wong P, Coulombe PA. Nature. 2006; 441:362–365. [PubMed: 16710422]

23. Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P. Mol. Cell. Proteomics. 2006; 5:737–748. [PubMed: 16432255]

24. Tardiff DF, Abruzzi KC, Rosbash M. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:19948–19953. [PubMed: 18077427]

25. Trakselis MA, Alley SC, Ishmael FT. Bioconjugate Chem. 2005; 16:741–750.

26. Vasilescu J, Figeys D. Curr. Opin. Biotechnol. 2006; 17:394–399. [PubMed: 16822661]

27. Sinz A. Mass Spectrom. Rev. 2006; 25:663–682. [PubMed: 16477643]

28. Novak P, Giannakopulos AE. Eur. J. Mass Spectrom. 2007; 13:105–113.

29. Sinz A. Anal. Bioanal. Chem. 2005; 381:44–47. [PubMed: 15616788]

30. Jin Lee Y. Mol. Biosyst. 2008; 4:816–823. [PubMed: 18633483]

31. Schilling B, Row RH, Gibson BW, Guo X, Young MM. J. Am. Soc. Mass Spectrom. 2003; 14:834–850. [PubMed: 12892908]

32. In Thermo Scientific Pierce Crosslinking Technical Handbook. http://www.piercenet.com/

33. Hermanson, GT. Bioconjugate Techniques. 2nd edn. Academic Press; London, UK: 2008.

34. Lomant AJ, Fairbanks G. J. Mol. Biol. 1976; 104:243–261. [PubMed: 957432]

35. Tang X, Yi W, Munske GR, Adhikari DP, Zakharova NL, Bruce JE. J. Proteome Res. 2007; 6:724–734. [PubMed: 17269728]

36. Swaim CL, Smith JB, Smith DL. J. Am. Soc. Mass Spectrom. 2004; 15:736–749. [PubMed: 15121203]

37. Leavell MD, Novak P, Behrens CR, Schoeniger JS, Kruppa GH. J. Am. Soc. Mass Spectrom. 2004; 15:1604–1611. [PubMed: 15519227]

38. Kalkhof S, Sinz A. Anal. Bioanal. Chem. 2008; 392:305–312. [PubMed: 18724398]

39. Madler S, Bich C, Touboul D, Zenobi R. J. Mass Spectrom. 2009; 44:694–706. [PubMed: 19132714]

40. Tang X, Munske GR, Siems WF, Bruce JE. Anal. Chem. 2005; 77:311–318. [PubMed: 15623310]

41. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Nat. Methods. 2008; 5:315–318. [PubMed: 18327264]

42. Petrotchenko EV, Xiao K, Cable J, Chen Y, Dokholyan NV, Borchers CH. Mol. Cell. Proteomics. 2009; 8:273–286. [PubMed: 18838738]

43. Yang, L.; Tang, X.; Munske, GR.; Bruce, JE. Novel photo-cleavable Protein Interaction Reporter (pcPIR) cross-linking and data-dependent MS/MS strategy for studying protein-protein interactions; 5th ASMS Conference on Mass Spectrometry; Philadelphia, PA. 2009;

44. Weisbrod, C.; Li, Y.; Tang, X.; Bruce, JE. Mapping protein-protein interactions in human serum with the protein interaction reporter (PIR)/chemical cross-linking strategy; 5th ASMS Conference on Mass Spectrometry; Philadelphia, PA. 2009;

45. Green NS, Reisler E, Houk KN. Protein Sci. 2001; 10:1293–1304. [PubMed: 11420431]

46. Cunningham BC, Jhurani P, Ng P, Wells JA. Science. 1989; 243:1330–1336. [PubMed: 2466339]

47. Cunningham BC, Wells JA. Science. 1989; 244:1081–1085. [PubMed: 2471267]

48. Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ. BMC Biol. 2007; 5:43. [PubMed: 17925020]

49. Halperin I, Wolfson H, Nussinov R. Structure (London). 2004; 12:1027–1038.

50. Wells JA. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:1–6. [PubMed: 8552582]

51. Higdon R, Kolker E. Bioinformatics. 2007; 23:277–280. [PubMed: 17121779]

52. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. Mol. Cell. Proteomics. 2004; 3:531–533. [PubMed: 15075378]

53. Trester-Zedlitz M, Kamada K, Burley SK, Fenyo D, Chait BT, Muir TW. J. Am. Chem. Soc. 2003; 125:2416–2425. [PubMed: 12603129]

54. Sinz A, Kalkhof S, Ihling C. J. Am. Soc. Mass Spectrom. 2005; 16:1921–1931. [PubMed: 16246579]

55. Ahrends R, Kosinski J, Kirsch D, Manelyte L, Giron-Monzon L, Hummerich L, Schulz O, Spengler B, Friedhoff P. Nucleic Acids Res. 2006; 34:3169–3180. [PubMed: 16772401]

56. Chu F, Mahrus S, Craik CS, Burlingame AL. J. Am. Chem. Soc. 2006; 128:10362–10363. [PubMed: 16895390]

57. Zhang H, Tang X, Munske GR, Tolic N, Anderson GA, Bruce JE. Mol. Cell. Proteomics. 2009; 8:409–420. [PubMed: 18936057]

58. Chowdhury SM, Du X, Tolic N, Wu S, Moore RJ, Mayer MU, Smith RD, Adkins JN. Anal. Chem. 2009; 81:5524–5532. [PubMed: 19496583]

59. Muller DR, Schindler P, Towbin H, Wirth U, Voshol H, Hoving S, Steinmetz MO. Anal. Chem. 2001; 73:1927–1934. [PubMed: 11354472]

60. Pearson KM, Pannell LK, Fales HM. Rapid Commun. Mass Spectrom. 2002; 16:149–159. [PubMed: 11803535]

61. Petrotchenko EV, Olkhovik VK, Borchers CH. Mol. Cell. Proteomics. 2005; 4:1167–1179. [PubMed: 15901824]

62. Seebacher J, Mallick P, Zhang N, Eddes JS, Aebersold R, Gelb MH. J. Proteome Res. 2006; 5:2270–2282. [PubMed: 16944939]

63. Bennett KL, Kussmann M, Bjork P, Godzwon M, Mikkelsen M, Sorensen P, Roepstorff P. Protein Sci. 2000; 9:1503–1518. [PubMed: 10975572]

64. Back JW, Sanz MA, De Jong L, De Koning LJ, Nijtmans LG, De Koster CG, Grivell LA, Van Der Spek H, Muijsers AO. Protein Sci. 2002; 11:2471–2478. [PubMed: 12237468]

65. Peterson JJ, Young MM, Takemoto LJ. Mol. Vision. 2004; 10:857–866.

66. Chowdhury SM, Munske GR, Tang X, Bruce JE. Anal. Chem. 2006; 78:8183–8193. [PubMed: 17165806]

67. Soderblom EJ, Goshe MB. Anal. Chem. 2006; 78:8059–8068. [PubMed: 17134140]

68. Soderblom EJ, Bobay BG, Cavanagh J, Goshe MB. Rapid Commun. Mass Spectrom. 2007; 21:3395–3408. [PubMed: 17902198]

69. Lu Y, Tanasova M, Borhan B, Reid GE. Anal. Chem. 2008; 80:9279–9287. [PubMed: 19551991]

70. Reynolds KJ, Yao X, Fenselau C. J. Proteome Res. 2002; 1:27–33. [PubMed: 12643523]

71. Back JW, Notenboom V, de Koning LJ, Muijsers AO, Sixma TK, de Koster CG, de Jong L. Anal. Chem. 2002; 74:4417–4422. [PubMed: 12236350]

72. Taverner T, Hall NE, O'Hair RA, Simpson RJ. J. Biol. Chem. 2002; 277:46487–46492. [PubMed: 12235153]

73. Gardner MW, Vasicek LA, Shabbir S, Anslyn EV, Brodbelt JS. Anal. Chem. 2008; 80:4807–4819. [PubMed: 18517224]

74. Singh P, Shaffer SA, Scherl A, Holman C, Pfuetzner RA, Larson Freeman TJ, Miller SI, Hernandez P, Appel RD, Goodlett DR. Anal. Chem. 2008; 80:8799–8806. [PubMed: 18947195]

75. Nadeau OW, Wyckoff GJ, Paschall JE, Artigues A, Sage J, Villar MT, Carlson GM. Mol. Cell. Proteomics. 2008; 7:739–749. [PubMed: 18281724]

76. Rink H. Tetrahedron Lett. 1987; 28:3787–3790.

77. Zhang H, Tang X, Munske GR, Zakharova NL, Yang L, Zheng C, Wolff MA, Tolic N, Anderson GA, Shi L, Marshall MJ, Fredrickson JK, Bruce JE. J. Proteome Res. 2008; 7:1712–1720. [PubMed: 18303833]

78. Kolb HC, Finn MG, Sharpless KB. Angew. Chem., Int. Ed. 2001; 40:2004–2021.

79. Kolb HC, Sharpless KB. Drug Discovery Today. 2003; 8:1128–1137. [PubMed: 14678739]

80. Anderson GA, Tolic N, Tang X, Zheng C, Bruce JE. J. Proteome Res. 2007; 6:3412–3421. [PubMed: 17676784]

81. Shi L, Chen B, Wang Z, Elias DA, Mayer MU, Gorby YA, Ni S, Lower BH, Kennedy DW, Wunschel DS, Mottaz HM, Marshall MJ, Hill EA, Beliaev AS, Zachara JM, Fredrickson JK, Squier TC. J. Bacteriol. 2006; 188:4705–4714. [PubMed: 16788180]

82. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Nat. Biotechnol. 1999; 17:994–999. [PubMed: 10504701]

83. Ross PL, Huang YN, Marchese JN, 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, Pappin DJ. Mol. Cell. Proteomics. 2004; 3:1154–1169. [PubMed: 15385600]

84. Ryan DP, Matthews JM. Curr. Opin. Struct. Biol. 2005; 15:441–446. [PubMed: 15993577]

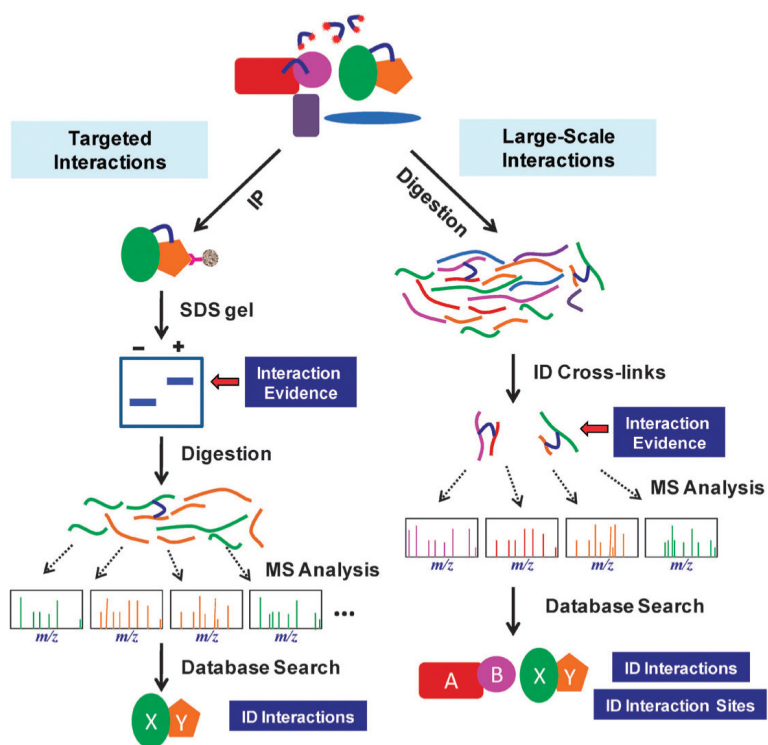85. Maly DJ, Allen JA, Shokat KM. J. Am. Chem. Soc. 2004; 126:9160–9161. [PubMed: 15281787]

Fig. 1.
Cross-linking applications: targeted interactions and large-scale interactions.
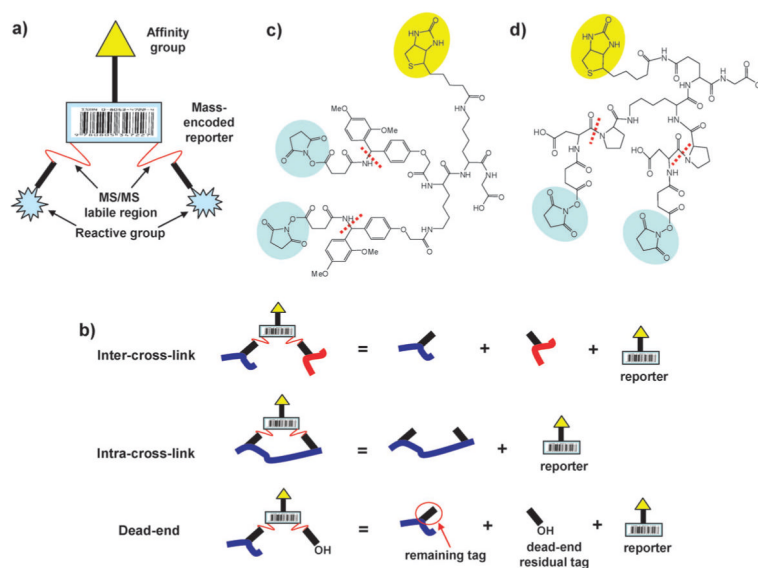
Fig. 2.
PIR structure. (a) Conceptual modular design of novel cross-linkers, protein interaction reporters (PIRs). The functional groups are color coded with structure examples given in (c) and (d). (b) The specific fragmentation pattern of PIR-labeled peptide distinguishes dead-end, intra-, and inter-cross-linked 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 and (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.
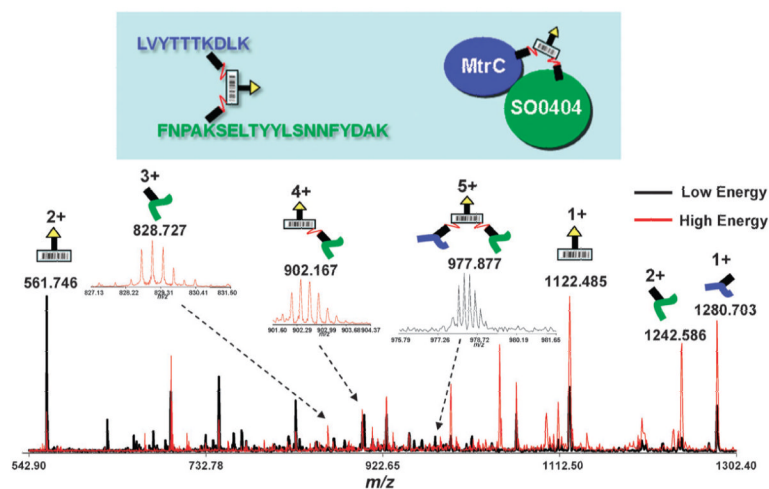
Fig. 3.
An example of inter-cross-link determination from multiplexed FTICR-MS spectra based on the mathematical relationship (precursor = peptide 1 + peptide 2 + reporter). A biotinylated Rink-based PIR (Fig. 2c) is used. The given example is a cross-linked peptide pair from protein MtrC and SO0404 in *Shewanella oneidensis*.
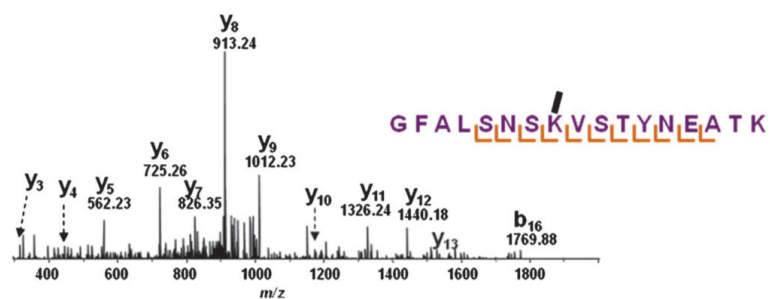
Fig. 4.
An example of MSMS spectrum of the peptide ion released from CID activation of cross-linked precursor ion using an ion trap mass spectrometer. The remaining stump modification on residue K is C4H5N1O2 with a mass of 99.089.
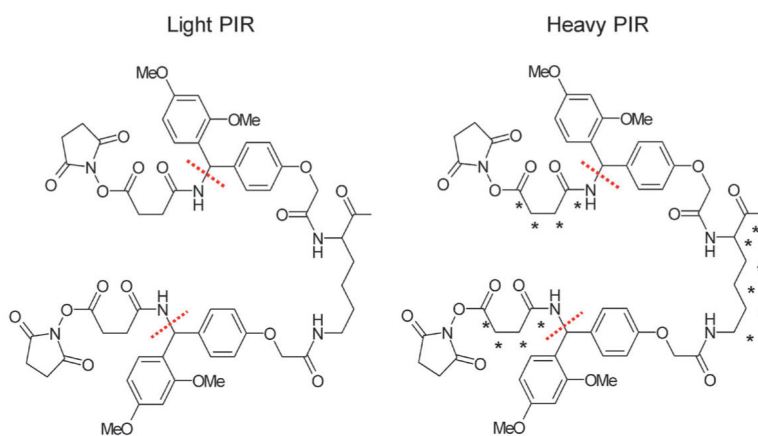
Fig. 5.
Incorporation of stable isotopes in PIR structure will result in heavy and light versions that can be used for relative quantitation among multiple samples. * indicates ^{13}C and/or deuterium-substituted compound available as a building block from commercial sources. The resultant PIR pairs will differ by 6 to 8 Daltons, depending on which site turns out to be optimal for isotope incorporation. The mass differentiation of labeled peptides from each sample will allow relative quantitation based on observed peak intensities.