

Large Scale Chemical Cross-linking Mass Spectrometry Perspectives

Boris L. Zybailov^{1*}, Galina V. Glazko², Mihir Jaiswal³ and Kevin D. Raney¹

¹Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

²Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

³UALR/UAMS Joint Bioinformatics Program, University of Arkansas Little Rock, Little Rock, AR, USA

Abstract

The spectacular heterogeneity of a complex protein mixture from biological samples becomes even more difficult to tackle when one's attention is shifted towards different protein complex topologies, transient interactions, or localization of PPIs. Meticulous protein-by-protein affinity pull-downs and yeast-two-hybrid screens are the two approaches currently used to decipher proteome-wide interaction networks. Another method is to employ chemical cross-linking, which gives not only identities of interactors, but could also provide information on the sites of interactions and interaction interfaces. Despite significant advances in mass spectrometry instrumentation over the last decade, mapping Protein-Protein Interactions (PPIs) using chemical cross-linking remains time consuming and requires substantial expertise, even in the simplest of systems. While robust methodologies and software exist for the analysis of binary PPIs and also for the single protein structure refinement using cross-linking-derived constraints, undertaking a proteome-wide cross-linking study is highly complex. Difficulties include i) identifying cross-linkers of the right length and selectivity that could capture interactions of interest; ii) enrichment of the cross-linked species; iii) identification and validation of the cross-linked peptides and cross-linked sites.

In this review we examine existing literature aimed at the large-scale protein cross-linking and discuss possible paths for improvement. We also discuss short-length cross-linkers of broad specificity such as formaldehyde and diazirine-based photo-cross-linkers. These cross-linkers could potentially capture many types of interactions, without strict requirement for a particular amino-acid to be present at a given protein-protein interface. How these short-length, broad specificity cross-linkers be applied to proteome-wide studies? We will suggest specific advances in methodology, instrumentation and software that are needed to make such a leap.

Keywords: Chemical cross linking; Mass spectrometry; Proteomics; Large-scale PPI

Abbreviations: PPI: Protein-Protein Interaction; AP-MS: Affinity Purification Mass Spectrometry; Y2H: Yeast Two Hybrid; SRM: Single-Reaction Monitoring; PTM: Post-Translational Modification; SH2: Src Homology 2; PDB: Protein Data Bank; FRET: Fluorescence Resonance Energy Transfer; EDC: 1-Ethyl-3-(3-Dimethylaminopropyl)-Carbodiimide; CXMS: Cross-Linking Mass Spectrometry; iTRAQ: Isobaric Tag for Relative and Absolute Quantitation; MS²: Tandem Mass-Spectrometry, or recording of a fragmentation spectrum of a molecule; MS³: MS-to-the-third, or recording of a fragmentation spectrum of a peak from MS²; NHS: N-Hydroxysuccinimidyl; PEG: Polyethylene Glycol; HCD: High Energy Collision Dissociation; SIA: Succinimidyl iodacetate; SDA: Succinimidyl-Diazirine; SDAD: NHS-SS-Diazirine ; DSS: Disuccinimidylsuberate; BS3: Bissulfosuccinimidylsuberate; MALDI: Matrix-Assisted Laser Desorption Ionization; ESI: Electrospray Ionization; qTOF: quadrupole Time-of-Flight; FA: Formaldehyde; FFPE: Paraffin-Embedded Formalin-Fixed; CLIP: Click-Enabled Linker for Interacting Proteins; BDRG: Biotin-Aspartate-Rink-Glycine; PIR: Protein Interaction Reporter; ETD: Electron-Transfer Dissociation; ECD: Electron-Capture Dissociation

Protein-Protein Interactions: Research Strategies

It has been long recognized, that in living systems, genes and proteins rarely act individually. Indeed, a particular response to environmental stimuli, such as disease, growth, or development is always an integration of multitude of interactions, an intricate web of connections between genes, proteins, and small molecules (Figure 1). Reconstruction, cataloguing, and functional categorization of these interactions is the major goal of “-omics” technologies. Functional

understanding of Protein-Protein Interactions (PPIs) and protein-nucleic acid interactions is essential part of describing how a particular genotype yields a corresponding phenotype. Furthermore, having detailed interactional information allows predicting systems behavior in response to a perturbation. For example, interactional information can be used to suggest new drug targets for therapeutic intervention [1-3].

Given its importance, mapping PPI networks has been on proteomics agenda for quite some time: for many organisms the most persistent interactions, either physical, genetic or computationally predicted has been mapped and catalogued. Before the advent of high-throughput technologies, most computational approaches to predict PPIs studied subunit interfaces, employing the information from the protein structure database [4]. The impetus in computational prediction of PPI has been gained by the availability of full genomic information, leading to the appearance of genomic context approaches that consider genome sequences to predict interactions [5]. These approaches use the fact that the genes of functionally interacting proteins are genomically associated with each other. Originally, non-random genomic co-occurrence, such as gene fusion [6,7], the conservation of gene order

***Corresponding author:** Boris L. Zybailov, Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA, Tel: 501-686-7254; Fax: 501-686-8169; E-mail: BLZybailov@UAMS.edu

Received January 07, 2013; **Accepted** February 04, 2013; **Published** February 08, 2013

Citation: Zybailov BL, Glazko GV, Jaiswal M, Raney KD (2013) Large Scale Chemical Cross-linking Mass Spectrometry Perspectives. J Proteomics Bioinform S2: 001. doi:10.4172/jpb.S2-001

Copyright: © 2013 Zybailov BL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

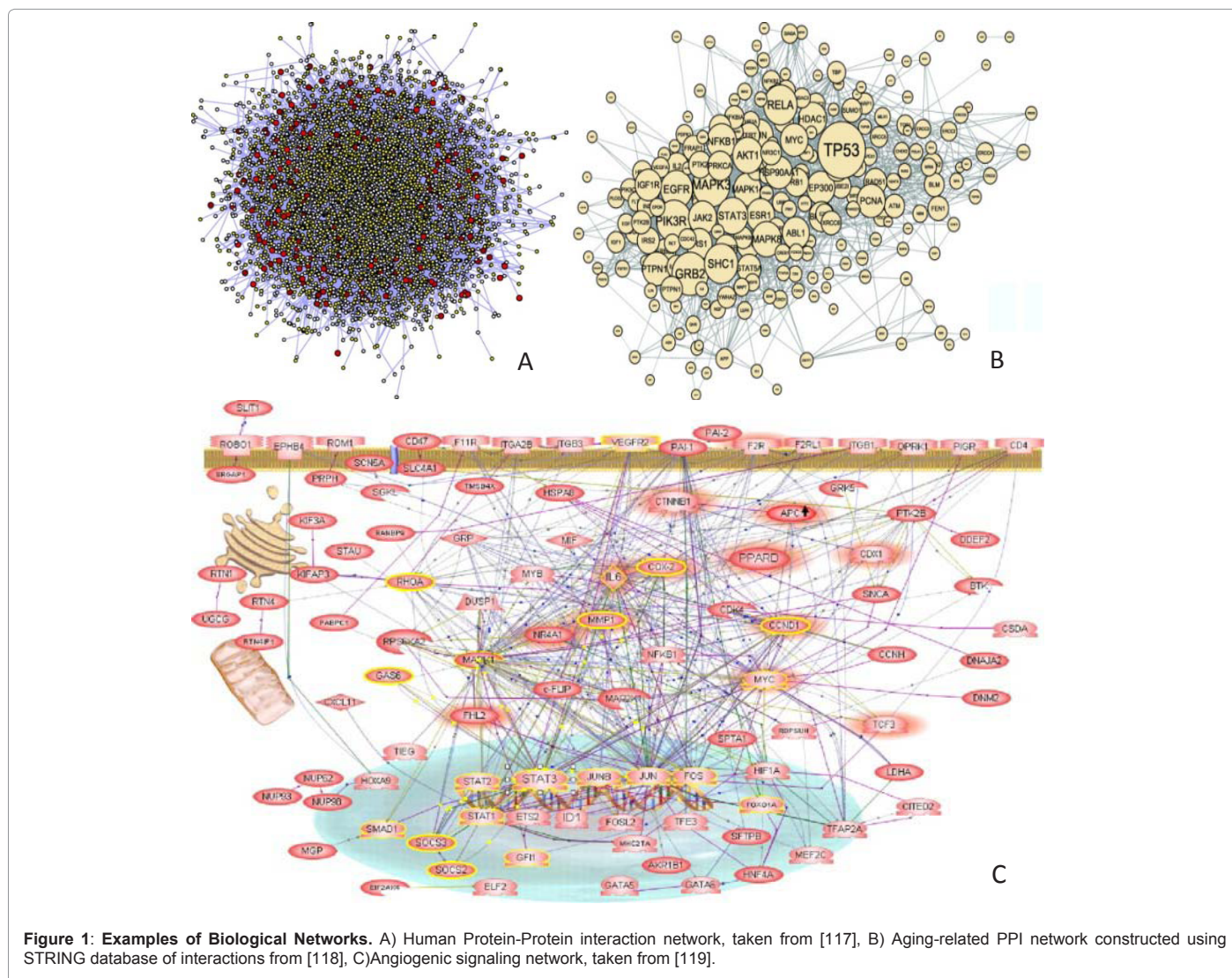


Figure 1: Examples of Biological Networks. A) Human Protein-Protein interaction network, taken from [117]. B) Aging-related PPI network constructed using STRING database of interactions from [118]. C) Angiogenic signaling network, taken from [119].

[8,9] and co-occurrence of genes among sequenced genomes [10] was used to get insights into PPIs.

Majority of the experiments to detect physical PPIs has been relying on Affinity-Purification-Mass-Spectrometry (AP-MS) [11,12] or Yeast-Two-Hybrid (Y2H) [13,14], or Mammalian-Two Hybrid [15] methods. These experiments provided a rich source of protein interactions, deposited in databases such as BioGRID, HPRD, IntAct, DIP and GeneMania [16-20]. Proteome-wide PPIs were collected in a high-throughput screen form a PPI network. In this network, a protein is treated as a node and an indirect link between two proteins (edge) represents physical interaction. This representation re-states the problem of computational prediction of protein interactions, in particular protein complexes, as a well-familiar problem of finding network modules. A module (cluster) in a network is a highly interconnected group of nodes, connected to the nodes outside the group with a few links [21]. The modularity of biological networks was observed for many types of networks [22,23]. It is generally believed that modules tend to represent groups of functionally associated genes/proteins that work together to perform a biological function. In a PPI network functional association refers to a protein complex.

Many algorithms have been developed to extract network modules [21,24,25], to identify the functional relationships between nodes and to further predict functional links. It should be noted that the identification of the network modules is possible if the number of edges is to some extent comparable to the number of nodes, otherwise the distribution of edges among nodes is too homogeneous [26]. Several algorithms were developed specifically to identify protein modules-molecular complexes-in a large PPI network (e.g. MCODE [27], RNSC [28], LCMA [29,30]).

Identifying protein complexes from PPI networks requires the high quality of the underlying network. However, the accuracy of high-throughput PPI data is generally low [31], undermined by specific biases, pertinent to different experimental approaches. For example, Y2H interaction networks tend to be enriched for transient interactions [32], while AP-MS tends to map indirect interactions [33]. This is why the algorithms for identifying densely connected PPI network components, although abundant, are not widely used for the inference of protein complexes.

Other experimental methods such as fluorescence resonance energy transfer (FRET [34,35]), and to lesser degree electron spin

resonance using spin-labeled pairs [36] has been used to further validate and refine the interaction geometry and protein-protein contacts. Additionally, “hot-spots” of protein-protein interactions - the residues, which stabilize the interaction - has been studied using site-directed mutagenesis approaches (e.g. alanine scanning [37]). The major drawback of these methods is, perhaps, their large-scale applicability-it takes a lot of effort to map the full interactome using these approaches.

At the same time, it is anticipated that the research demand for fast large-scale network mapping will increase as biologists start to ask questions not just about how different protein levels differ between one condition vs. the other, but how do the interaction networks differ [38,39]. Excellent example of the differential network study is the recent paper by Bisson et al. [40], where the authors examine signaling PPI network, associated with GRB2 adaptor. To map the GRB2-centered PPIs, the authors used a mass-spectrometry approach based on affinity purification of GRB2 and Single-Reaction Monitoring (SRM). Comparing summed SRM intensities of GRB2 interacting partners before and after stimulation of signaling, allowed the authors to quantitate difference in strengths of the interactions. Notably, Bisson et al. [40] focused on the key hub protein, GRB2. However, if we were to pose the question of the network dynamics on the scale of the full interactome, even if focusing only on hubs, it would take considerable effort to get to the answer using the affinity-purification methods.

In this regard, taking a snap-shot of interactome using chemical cross-linking may offer a significant advantage. Potentially, the cross-linking could not only give identities of interacting partners, but it can also provide topology of their interactions in the same setting. Despite being conceptually attractive, chemical cross-linking is notoriously difficult to implement on the large scale. What are the challenges and technological limitations that need to be overcome for the large-scale cross-linking to be an effective PPI research tool? In this review we focus on the following three: 1) In digests of complex protein mixtures, the non-cross-linked peptides always dominate; 2) If the previous problem is solved, say, by enriching for the cross-linked peptides, still, cross-links within the same molecule (intra-protein) will be several folds more abundant than cross-links between different molecules (inter-protein); 3) How does one select a cross-linking reagent to capture most of the interactions, given the high structural and chemical heterogeneity of protein-protein interfaces?

In other words, are the cross-linking chemistry and the current mass spectrometry instrumentation adequate to solve the double ‘needle-in-a-haystack problem’-finding cross-linked peptides amongst non-cross-linked species and discriminating between inter- and intra-cross-links? There are promising studies coming out recently from several proteomics laboratories that point to the positive answer to this question.

Chemistry of Protein-Protein Interactions

Biologically relevant PPIs are commonly classified into persistent (stable protein complexes) and transient. A given protein may exhibit broad spectrum of affinities to its substrates, which are modulated and regulated by the cell. This is especially evident in the case of transient interactions, which are prevalent during signal transduction processes, where a given protein’s affinity towards its interacting partner is modulated by a certain Post-Translational Modification (PTM). For example, proteins with SH2 domains recognize phosphorylated

tyrosines, proteins with bromo-domains recognize acetylated lysines, and proteins with chromo-domains recognize methylated lysines. In principle, by examining physico-chemical properties of known stable and weak protein-protein interfaces it is possible to construct a classifier, which will with reasonable accuracy predict if a given interface is persistent or transient [41,42]. Therefore, in general, the chemistry of interactions is different in transient vs. persistent PPI interfaces.

From the purpose of designing effective cross-linking methodology, which would aim at capturing broad range of transient and persistent interactions, it is also useful to know likelihood that a particular amino-acid or an amino-acid pair is present at a PPI interface. One approach to assess this likelihood is to examine structures of protein complexes deposited into the Protein Data Bank (PDB) and to count amino acid occurrences at the interfaces [43,44]. Table 1 summarizes these findings and shows relative frequencies of aminoacids at protein interfaces. Interestingly, the interface propensity of an amino acid correlates not with its hydrophobicity, but with its propensity to form under-wrapped hydrogen bonds, or “dehydrons”, (Table 1) ([45] for the detailed discussion of the dehydrons). Overall, hydrogen bonds and salt bridges play the most significant role at the interfaces [43], while hydrophobic interaction is the second in the order of importance [46]. On average, a PPI interface has about ten hydrogen bonds, and about two salt bridges [46]. For a random interface there is still a good chance to find amino acids Leu, Ilu, and Val due to their high overall abundance. In the order of decreasing frequency, the following amino acids are over represented at PPI interfaces compared to other surfaces: Asn, Thr, Gly, Ser, Asp, Ala, and Cys [45]. In terms of amino-acid to amino-acid contacts, amino acids pairing with the highest preferences are Cys-Cys, Trp-Pro, Asp-His, Arg-Trp, Asp-Ser, and Asp-Thr [47]. Additional examination of PPI interfaces reveals that structural water and metal ions often participate in the interaction at the interfaces, and often shield charged residues from each other [48].

Amino Acid	Interface Propensity [45]	Total Abundance [45]	Rim/Core frequency [49]	Dehydron Propensity [45]	Hydrophathy
Asn	1.28	3.36	1.19	1.63	-3.50
Thr	1.10	4.87	1.19	1.41	-0.70
Gly	0.99	7.30	1.16	1.42	-0.40
Ser	0.60	4.66	1.04	0.80	-0.80
Asp	0.34	5.42	1.48	0.76	-3.50
Ala	0.29	7.77	0.95	0.60	1.80
Cys	0.25	0.78	0.45	0.24	2.50
Val	0.20	8.17	1.09	-0.31	4.20
Met	0.10	3.00	0.67	0.10	1.90
Tyr	0.10	2.41	0.54	0.10	-1.30
His	-0.25	4.35	1.24	-0.25	-3.20
Pro	-0.25	1.92	0.52	-0.25	-1.60
Trp	-0.33	1.02	0.32	-0.40	-0.90
Arg	-0.35	8.91	0.82	-0.40	-4.50
Leu	-0.35	6.27	1.19	-1.10	3.80
Phe	-0.40	3.61	0.33	-0.40	2.80
Lys	-0.42	7.76	2.16	-0.38	-3.90
Glu	-0.50	8.59	1.87	-0.11	-3.50
Gln	-0.62	3.15	1.03	-0.60	-3.50
Ile	-0.70	6.66	0.76	-0.92	4.50

Table 1: Amino acid frequencies within protein-protein interfaces.

Further examination of PDB structures of interacting proteins reveals differences in amino-acid composition deep inside the interface (core) and on the periphery (rim). Table 1 summarizes these results in the order of decreasing frequency, residues Lys, Glu, Asp, Pro, Asn, Pro are likely to occur on the rim, while residues Trp, Phe, Cys, His, Met are likely to occur within the core [49]. Importantly, in cores of the interfaces amino acid residues are protected from the outside environment and most likely won't interact with bulky long-length chemical cross-linkers, especially if the protein-protein interaction is stable. On the other hand, the shorter the length of the cross-linker is, the more likely it will get closer to the interface. In certain cases it is possible to engineer cross-linking groups into the protein itself, which would capture interaction exactly at the interface (e.g. photo-reactive amino acids [50]). In addition, in the case of a transient interaction, there is a possibility that modification with a small, short-length cross-linker prior to the interaction might still capture the interface after the interaction is formed.

It is reasonable to expect long-length cross-linkers to capture interactions away from the interface, while shorter cross-linkers will be capturing the interaction closer to the interface. Also, if we will choose very specific cross-linker, for example zero-length cross-linker 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), which catalyzes condensation of primary amines with carboxylic acids [51,52]; we will obtain only those interactions, which by chance will have the lysines and aspartates/glutamates at the right positions. Consequently, EDC will be blind to all the other interactions. Also, in cases where a carboxylic acid/amine pair is deeply-buried within PPI interface, EDC might not even get to it, thereby ignoring most stable interactions. Yet, at the same time, other interactions that EDC does capture will be close to the PPI interface, because it is a zero-length cross-linker.

In contrast, if we choose a very long, flexible cross-linker coupling a certain amino acid pair there is a good chance that majority of interactions will have at least one such pair at right position within the cross-linker's reach to be captured. However, it is unlikely that in this scenario, we will get to the PPI interfaces themselves. Another thing to consider is that long-length cross-linker may detect spurious, non-specific interactions and alter the native protein structure more than the short-length cross-linkers would [52].

We can, therefore, conclude that long-length cross-linkers, selective for frequently occurring amino acids, are good for establishing identities of interacting partners; while shorter, broadly selective cross-linkers are better suited for capturing interacting interfaces themselves.

Chemistry of Protein-to-Protein Cross-Linking

Figure 2 illustrates common steps undertaken during a typical Cross-Linking Mass Spectrometry (CXMS) study. The exact number and order of the steps depends on the task at hand. Figure 3 provides a general scheme for protein cross-linking. An archetypical cross-linker has at least two functional groups, which are either the same (homo-bifunctional), or different (hetero-bifunctional). Reaction 1 involves activation of the first protein with the first functional group on the cross-linker. Reaction 2 involves cross-linking event itself using the second functional group of the cross-linker. If the reaction 2 takes place within the same protein, the result is called intra-protein cross-link. If it occurs with another protein molecule, it is called inter-protein cross-link. It should be noticed that costs in entropy are different between reactions 1 and 2. For example, if the interaction is formed

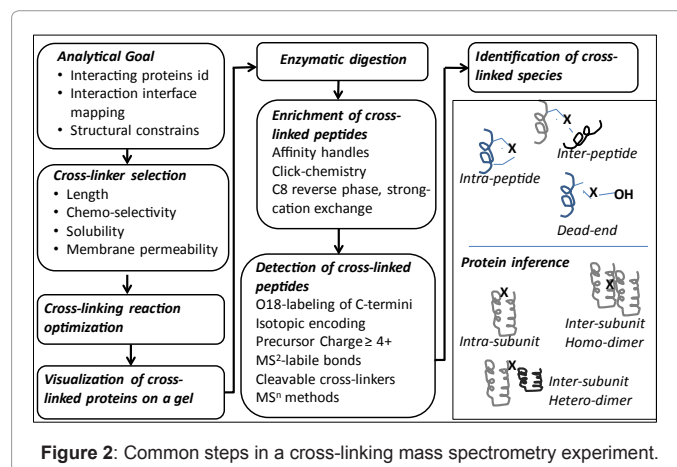


Figure 2: Common steps in a cross-linking mass spectrometry experiment.

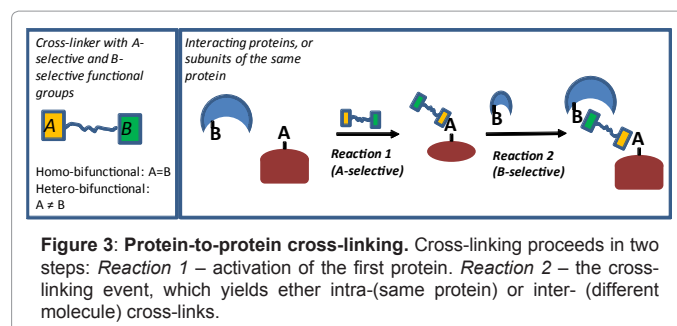


Figure 3: Protein-to-protein cross-linking. Cross-linking proceeds in two steps: Reaction 1 – activation of the first protein. Reaction 2 – the cross-linking event, which yields either intra-(same protein) or inter-(different molecule) cross-links.

prior to the cross-linking, and the amino-acids to be linked are at the right geometric positions-the only entropic cost for the reaction 2 is in curbing the rotation of the cross-linker, which was already bound during the reaction 1. In such cases we will expect the chemo-selectivities to be slightly different: even for the homo-bifunctional cross-linkers the reaction 2 will be less selective than the reaction 1. Corollary to the observation of differences in energetics between reaction 1 and 2, is that we may potentially get different results, depending on which protein gets activated first in the reaction 1.

In the following section we will look at what physicochemical properties of cross-linked peptides can be exploited for their detection, and also what functional groups are available for effective cross-linking. We limit our survey to the approaches, which could be potentially useful on the large-scale. For a more detailed discussion, see the recent review by Paramelle et al. [53].

Physicochemical Properties of Cross-Linked Peptides

A complex protein mixture from a biological specimen is highly heterogeneous, where individual proteins differ considerably in hydrophobicity, molecular weight, and isoelectric point. Enzymatic digestion, while resulting in increased number of distinct molecules (peptides derived from the starting proteins), shrinks the physicochemical space, thereby simplifying peptide detection, peptide identification, and protein inference. This is the basis for “bottom-up” protein identification experiment, where each of the protein from the starting protein mixture can be identified by several different peptides [54].

In a digest by trypsin protease, which cuts after Lys and Arg residues, peptides on average are around 12-15 amino acid in length,

with majority having charge 2⁺ at pH 2.5 (which is usual pH during ionization/mass spectrometry in a bottom-up proteomics experiment). One of the two charges is due to protonation of peptide N-terminus, and the other one is due to protonation of the C-terminal Lys or Arg. Small proportion of lower charges is due to non-specific cleavages, protein C-termini, and occasional loss of a proton during ionization process. Similarly, small proportion of higher charges is due to incomplete digestion, His residues, and occasional capture of a proton during ionization process.

When using non-specific protease (e.g. proteinase K), one could manipulate reaction times and temperature to produce peptides of desired average length [55]. In this case, however, because of the absence of the selection for Lys- and Arg-containing peptides, higher proportion of lower charges is expected for the shorter peptides.

Next, when the enzymatic digestion of cross-linked proteins occurs, we can expect the cross-linked peptides to be different from the non-cross-linked ones in the following properties: i) charge to be doubled on average; ii) length to be doubled on average; iii) cross-linked species to have at least two C-terminal carboxy- groups and two N-terminal amino-groups. Noticeably, the properties i-iii) are not dependent of the nature of the cross-links. It is, therefore, advisable to use these properties for the large-scale studies when employing short-length, zero-length, or natural cross-links.

Enrichment by strong-cation exchange resin [56,57] (increase in charge), or using size exclusion chromatography [57] (increase in size) have been used as effective cross-link enrichment strategies. On a side note, the authors [57] also showed that the use of Asp-N protease parallel to trypsin boosted the number of detectable cross-links. Detection of highly charged cross-linked peptides is facilitated further by the high resolution mass spectrometry, available on modern hybrid instruments [58]. With the high resolution, charge of a peptide can be unambiguously assigned based on the separation of its isotopic peaks on the m/z axis, thereby enabling charge-driven acquisition of the fragmentation spectra. This strategy has been proved very effective to detect the cross-linked peptides [59].

The third property-increase in the number of carboxy and N-termini within cross-linked peptides can be exploited via selective chemical labeling. Digestion by trypsin in O¹⁸ water, which introduces two O¹⁸ atoms into a peptide C-terminus, has been used in quantitative proteomics for a long time. In a typical O¹⁸ quantitation experiment, one sample is digested in O¹⁸ and another in O¹⁶ followed by 1:1 mixing. Because the chromatographic properties are not affected, O¹⁸ and O¹⁶ peptides co-elute and enter mass detector at the same time, with O¹⁸:O¹⁶ ratio indicative of relative abundance of the protein in the mixture [60,61]. When applied to the cross-linked samples, one half of the sample is digested in O¹⁸ and another in O¹⁶ water. Next, the samples can be analyzed either after mixing, or separately [62,63]. As a result, the cross-linked peptide pairs will be separated by ~8 Da (two labeled C-termini), while non-cross-linked peptides, will be separated by ~4 Da (one labeled C-termini). Using this approach one also needs to be aware of the additional incorporation of O¹⁸ atoms during non-enzymatic deamidation of Asn and Gln [64-66] as well as back-exchange [67] and incomplete labeling [68,69]. Similar to the counting of C-termini using incorporation of O¹⁸, albeit less frequently used, N-termini also can be counted by post-digestion chemical modifications [70]. In fact, the authors of [70] demonstrated that if a 1:1 heavy:light mixture is used to modify N-termini, inter-peptide crosslinks exhibit a distinct isotopic signature (a 1:2:1 ratio). Just like with the deamidation of asparagines

producing artifacts in the O¹⁸ method, here, special attention needs to be paid to miss-cleaved lysines residues and reduction in charge.

It would be interesting to combine the N-terminal and C-terminal chemical labeling of cross-linked peptides; as we anticipate that such methods could be particularly useful for large-scale differential PPI studies. Recently, Zhang et al. [71] reported successful use of similar strategy for analysis of N-glycosylation sites. The authors combined an N-terminus-labeling reagent, isobaric tag for relative and absolute quantitation (iTRAQ) [71-73] with O¹⁸ labeling to quantify the glycosylated and non-glycosylated peptides. The glycosylation sites were marked via PNGase F catalyzed labeling in O¹⁸/O¹⁶ water. Using the multiplexing capability of the iTRAQ reagents, the authors performed simultaneous analysis of four samples.

The above methods can be employed on the large scale to improve detection of any type of protein-to-protein cross-links, including short-length and zero-length. For longer cross-linkers, on the case-by-case basis, their chemical structure can be manipulated further to enhance enrichment, detection, and analysis of the cross-linked species. These strategies include isotopic coding, affinity handles, click-chemistry handles, ionization enhancers, reporter ions, and MS²-labile bonds [53].

In the following sections we will examine functional groups on commercial and custom-synthesized cross-linkers, which could be potentially useful in large-scale studies.

Amine-Reactive Functional Groups

N-hydroxysuccinimidyl-ester (NHS-ester) is the most popular functionality in this category. Perhaps, the most common cross-linking reagents - DSS (Figure 4) and its soluble analog BS3 are used in various types of cross-linking experiments to yield irreversible cross-links. In addition to primary amines, lysine side chains and protein N-termini-NHS group has some reactivity towards serine, threonine, and tyrosine [74], especially at the second step of the cross-linking for the entropic reasons already discussed. Overall, the DSS cross-linking reaction is reasonably fast and selective, giving accurate snap-shot of PPIs present in solution. Cross-link introduced by DSS is 11.3 Å long, but shorter and longer versions of homo-bifunctional NSH-esters are also available; including cleavable reagents and molecules with Polyethylene Glycol (PEG) linkers to increase the reagent's solubility.

Recently, other amine-reactive functional groups -N-hydroxyphthalimide, hydroxybenzotriazole, and 1-hydroxy-7-azabenzotriazole have been shown to have faster reaction times and improved efficiency compared to DSS [75]. Imido-esters and thioimido-esters another promising functional group for the amine-to-amine cross-linking studies. Advantage of these reagents over DSS is that they preserve positive charge, thereby minimizing alteration of native protein structures [76]. In addition, the higher charges often imply easier enrichment and detection, and in many cases more complete fragmentation in tandem MS [77].

Because lysine is quite abundant amino acid residue (Table 1), the use of amine-to-amine cross-linkers is quite effective in variety of applications. For example, variable-length amine-to-amine cross-linkers are often used to constrain the Lys-to-Lys distances. Even though these constrains provide rather weak structural information, using cross-links in conjunction with molecular modeling is usually sufficient to determine the structure of a protein complex or single protein at a moderate resolution [78-80]. There are also several studies,

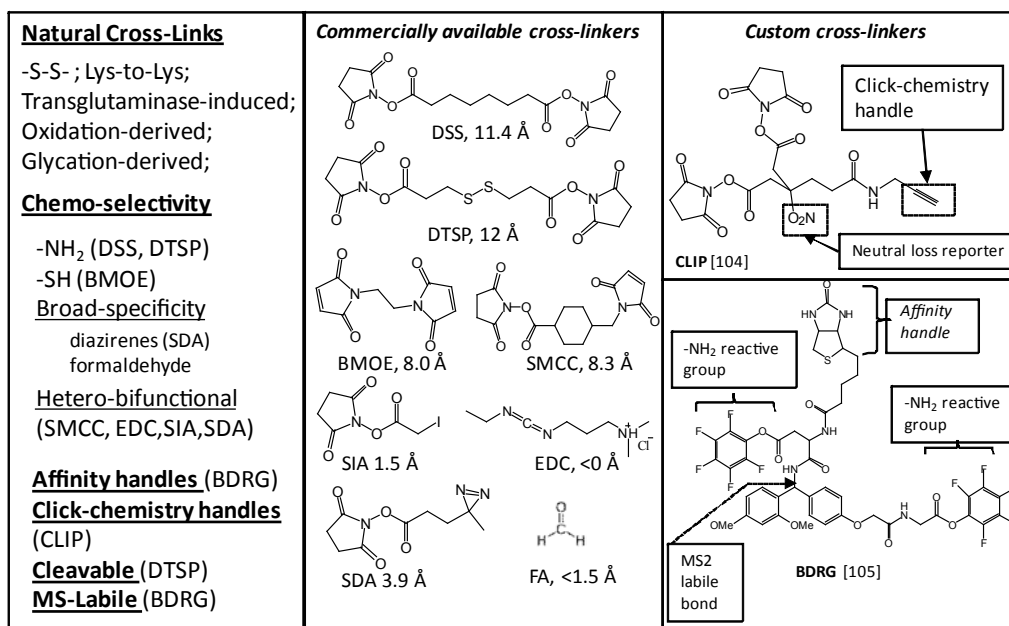


Figure 4: Cross-linking reagents commonly used in PPI studies. Cross-linker selection: depending on the analytical task at hand, a cross-linker is chosen based on its length and physico-chemical properties (left panel). A variety of cross-linking reagents are commercially available (middle panel). Custom-synthesized cross-linkers include MS-labile groups, affinity handles, and conjugation-chemistry handles for facilitating detection and purification are shown in the right panels.

which use DSS type cross-linkers on large scale. For example, Rinner et al. [81] used *E. coli* whole-lysate cross-linking to demonstrate utility of the xQuest algorithm. To enable the large-scale analysis the authors employed isotopically coded DSS cross-linker (DSS-d0/d12 pair) and focused on highly charged precursors. Majority (above 3000), of the detected cross-links were intra-protein, while only 71 were inter-protein cross-links, which included subunits of Serine hydroxymethyltransferase (GLYA), GroEL, Tryptophanase (TNA) and the small ribosomal subunit. The authors further validated most of these inter-protein *E. coli* cross-links by examining corresponding X-ray structures.

Similarly, in the more recent analysis by Yang et al. [82], introducing the software pLINK, using BSS-d0/d4 isotopically coded cross-linking and high-resolution High-energy Collision Dissociation (HCD), identified 394 interlinks from *E. coli* lysates.

Sulfhydryl-Reactive Functional Groups

Alkyl iodides and maleimides [83], and less often, thiopyridines [84] are used to selectively modify cysteine residues in proteins. Homo-bifunctional bis-maleimido cross-linkers are often employed on a small scale for the Cys-Cys coupling [85]. Because Cys residue is rarer compared to Lys (Table 1), sulfhydryl-reactive functionality may be less useful in large-scale PPI studies. In a hetero-bifunctional format, e.g. NHS-Maleimide, Lys to Cys cross-linking has its uses as a structural technique to derive Lys-Cys distance constrains [86]. Succinimidyliodoacetate (SIA) is a short-length (~1.5 Å) hetero-bifunctional cross-linker, which could be potentially useful for mapping of rims of PPI interfaces.

Hydroxyl-Reactive Functional Groups

Isocyanate moiety, -NCO, is used for the purpose of hydroxyl-specific reaction, but is rarely used in the PPI studies. This is probably because the -NCO reacts well with stronger nucleophiles-primary

amines and sulfhydryls. It does require un-protonated amine for the reaction, therefore, at physiological pH this leaves threonines, serines, and histidines as likely sites of the -NCO action. Another reason for the limited use of the -NCO chemistry is that the isocyanate coupling is a rather slow reaction; and for effective cross-linking faster reaction times are preferred. However, the activated -OH groups, such as in the active sites of serine proteases, are very reactive towards -NCO [87]. In fact, this is one application where -NCO cross-linking chemistry could be useful-selective cross-linking of serine proteases with their substrates.

Carboxyl-Reactive Groups

Carbodiimides, such as already discussed EDC, activate carboxyl groups for condensation with primary amines via amide bond formation [88,89]. We are not aware of much other carboxyl-selective functionality used in chemical cross-linking of carboxyl groups. One other possibility is dehydration induced carboxy-to-amino condensation in the solid state [88]. Potentially the latter approach could be more useful in capturing amino-to-carboxy ionic bridges deep inside stable PPI cores, which EDC would ignore.

Photo-Reactive Groups

Photo-reactive groups which are used in many types of bioconjugation including protein cross-linking, are diazirines [90], azides [91], and benzophenones [92] (Figure 4). Upon photolysis, these groups generate highly reactive free radical species, which show almost no chemo-selectivity. Typically, the photo-reactive functionalities are used in hetero-bifunctional cross-linkers that are highly chemo-selective in the reaction 1, while the non-selective photo-reactive moiety is employed in the reaction 2. NHS-diazirine cross-linkers (Figure 4, SDA) are of this variety. We are not aware of any large-scale study using such cross-linking reagents. On the smaller scale, however, SDA was proved highly effective. For example, Gomes and Gozzo studied cross-linking of model peptides and equine myoglobin using 13.5 Å

long, cleavable NHS-diazirine cross-linker, SDAD [93]. As expected, the NHS-diazirine generated higher number of cross-links, compared to DSS. Also, using MALDI-MS/MS and ESI-qTOF-MS/MS platforms, the authors showed that all cross-linked spectra had characteristic ions indicative of carbene insertion.

We therefore expect that when used on the large-scale the sheer number of captured interactions with NHS-diazirine cross-linkers should be higher than those obtained with Lys-to-Lys cross-linking. Moreover, the existence of predictable fragmentation of the points of carbene insertion as demonstrated [93] can aid in the cross-link identification.

Photo-active Leu and Met analogues containing diazirine groups are especially interesting, as Leu and Met are very abundant (Table 1). Suchanek et al. [50] demonstrated that these amino-acids can be incorporated into proteins in live cells by native cellular translation machinery. Using this method, Suchanek et al. [50] discovered a novel interaction of the progesterone-binding membrane protein PGRMC1 with Insig-1, a regulator of cholesterol homeostasis.

Other genetically encoded un-natural amino acids, which have photo-activable functional groups has also been used to explore protein-protein interactions [94-97].

Formaldehyde

Arguably, formaldehyde (FA) is the most important cross-linker. It is frequently used on the large scale to fix, conjugate, and cross-link proteins and protein-to-nucleic acids. FA has been shown to aid in affinity purification; stabilizing specific interactions within protein complexes [98]. There are numerous tissue banks, which are filled with Paraffin-Embedded Formalin-Fixed (FFPE) samples that potentially hold therapeutically valuable PPI information. FA reaction times are fast; FA easily penetrates membranes, and therefore is next-to-ideal cross-linker for *in vivo* studies. It is also easily reversible at elevated temperatures.

Generally, FA is considered a broad-specificity cross-linker and has potential to cross-link any nucleophile to any nucleophile in a protein. In practice, reaction 1 proceeds the fastest with primary amines, and the reaction 2 with primary amines, histidines, asparagines, glutamines, tyrosines, and arginines [99]. The reaction products are quite heterogeneous in context, and structure-dependent [100] and, in addition to simple methylene bridges, the cyclic and polymeric end-products are abundant. It is therefore important to optimize the FA-cross-linking conditions to maximize the benefit. If more selective cross-linking is desired, reaction times should be short, and formaldehyde concentration low. In contrast, if broader specificity is desired, and there is not much concern about the ill-defined products of formaldehyde modification, reaction times could be longer. Concentration range at 0.1-2% is a typical range for the PPI studies using FA. Low FA concentrations (0.1-0.5%) are also more selective and yield mostly Lys- and Trp-directed cross-links [99].

The FA's broad specificity and chemical diversity of the end products complicates MS-based proteomics analysis. Despite such difficulties, successful strategies for FA cross-link identification are advancing, and we anticipate seeing more such studies in the future [101,102].

Multi-Functional Cross-Linker Design

For the purposes of enrichment and facilitation of cross-link

detection during data acquisition, variety of other functionalities in addition to chemo-selective groups has been explored for the design of cross-linkers. Most frequent are click-chemistry handles, biotin handles, reporter ions and fragmentation specific cleavages [53]. Click-chemistry, i.e. copper-catalyzed alkyne-azide conjugation [103], is an attractive cross-link enrichment strategy for the large scale studies because the reaction is very fast and selective; plus, alkyne handle itself is compact and inert.

Chowdhury et al. [104] demonstrated effectiveness of this approach in the form of a novel Lys-to-Lys cross-linker ((Figure 4), CLIP). The CLIP cross-linker has alkyne group for the click-chemistry capture, and the reporter-NO₂ group, which allows for neutral loss scanning and also increases water solubility of the reagent. The authors further evaluated the click-chemistry effectiveness in enriching cross-linked proteins from complex backgrounds, by mixing the CLIP-cross-linked samples with *E. coli* lysates. As a result, for as low as 1:100 mixing into the non-cross-linked protein background, enrichment efficiency of the cross-linked peptides remained exceptionally high. Importantly, CLIP has the same chemo-selectivity as DSS, and is almost the same length. We, therefore, anticipate that CLIP can easily replace DSS in large-scale cross-linking and structure analysis, because in comparison to DSS; enrichment yields more interactions and provides higher quality mass spectrometric evidence.

In a similar study, using biotin-handle along with MS²-labile bond engineered into the cross-linker ((Figure 4), BDRG), Luo et al. [105] demonstrated effectiveness of this strategy on the scale of large protein complexes. Luo et al. [105] used the MS²-labile bond, and the subsequent MS³ scans of resultant fragments to enable rapid identification of the cross-linked peptides using regular database search.

In the latter study the MS²-labile bond is called Rink bond; due to the stability of its fragmentation products this bond is more labile than peptide bonds. One could also use two Rink functionalities in the same molecule, so that cross-linker is cleaved out during fragmentation process and serves as a reporter. Such strategy was implemented by Zhang et al. [106] using novel cross-linker called Protein Interaction Reporter (PIR). PIR has also biotin handle in addition to the two Rink groups for the enrichment of cross-linked proteins. Zhang et al. [106] also studied another MS² mobile bond - Asp-Pro - as an alternative to Rink. The authors applied the PIR method to *Shewanella oneidensis* bacteria and showed that PIR is a particularly well suited for capturing interactions of membrane proteins.

Disulfide Bonds

While being rare, Cys is enriched at PPI interfaces (Table 1). Plus, whenever it does occur in PPI interface it is most likely to form disulfide bond with another Cys from the interacting protein [47]. Therefore, it is beneficial to map disulfide bridges within large scale studies of PPIs, in conjunction with other approaches, in order to increase number of captured interactions.

Because disulfide bonds are easily reduced, overlaying LC-MS maps in reduced vs. oxidized samples allows for the cross-link detection. Peptides originally carrying a disulfide bond are recognized due to the shift in both retention time and m/z values, whereas peptides containing no cysteine stay the same. Such approach was undertaken by Evaristo et al. [107] in the identification of disulfide bridges within skin venom of several amphibian species.

Another approach for the disulfide-specific detection is to exploit

the disulfide affinity for electrons in the gas phase. Fragmentation techniques, such as Electron-Transfer Dissociation (ETD) and Electron-Capture Dissociation (ECD), can be used to fragment cross-linked peptides selectively at the disulfide bond [108]. Next, performing MS³ on the resulting fragments allows for accurate identification of the cross-linked sequences [109]. Ultraviolet irradiation in the gas phase also has been reported to selectively cleave the disulfide bonds [110].

Cross-Link Identification Algorithms

For a small-scale analysis, it is relatively easy to design a cross-link search algorithm, which uses a database consisting of pair-wise combination of peptides from interacting proteins. Next, this database is used to constrain masses of possible cross-linked precursors. Examples of such algorithms are abundant: CLPM [111], xComb [112], GPMW [113], X!Link [114], StavroX [98], MassMatrix [115].

Within the Massmatrix it is also possible to use multi-staged strategy, which can be applicable on large-scale [115]. Notably, cross-linking analysis is not the primary function of MassMatrix, which is really a complex proteomics platform, including stand-alone database search engine with parallel computation capabilities, use of multiple fragmentation modes, and different quantitation strategies. The multi-stage implies running regular protein identification search first, and limiting the subsequent cross-links search to reliably identified proteins.

For complex protein mixtures the database of peptide-pairs becomes impractically large and different approaches are required. Generally, we need software platforms, which could use information from isotopic coding, O¹⁸-labeling, reporter ions, and MS³ experiments to reliably detect cross-linked species. Additionally, we need algorithms incorporating *de-novo* sequencing, open-search modifications strategies, high-resolution mass spectrometry and multiple fragmentation modes to reliably identify the cross-linked species. Software platforms such as pLINK [82], xQuest [81], XLink-Identifier [116] are in this category. Cross-link identification pipe-lines using existing algorithms for open modification search [59] and *de novo* sequencing are of particular interest. Indeed, as already discussed, while the use of non-specific cross-linker's groups, such diazirines may complicate the bioinformatics analysis, it is exceptionally attractive to use on the large-scale; and *de novo* approaches could be the key technology here.

Conclusion

In summary, we suggest that the short-length, broad-specificity cross-linkers are the most suited for the large-scale studies, including cross-linking in live cells. In addition, it is always useful to perform the disulfide mapping, as Cys-Cys coupling is frequent at the PPI interfaces. Therefore, methods, which exploit most general physicochemical properties of cross-linked peptides-higher charge, bigger size, and higher number of C-termini, are the most suited for the large-scale analysis. At the same time, complementary analysis with specific, long-length cross-linkers will be desirable for the validation and sensitivity-evaluation purposes. The current stage of technology-high resolution mass spectrometry, hybrid instruments, methods of multidimensional separation, multi-stage fragmentation techniques-is adequate, in our opinion for the large-scale cross-linking. Amongst the three main obstacles listed in the introduction, identifying inter-protein cross-links in the background of intra-protein cross-links is the most general and

difficult problem to solve. The bigger dynamic range of detection and the shorter scan times become on the modern mass spectrometers, the more inter-protein cross-links will be identified in the large-scale PPI studies. However, in the case protein of homo-dimers and oligomers without available X-ray structure novel approaches to validate inter-molecular cross-links will be needed.

References

1. Hormozdiari F, Salari R, Bafna V, Sahinalp SC (2010) Protein-protein interaction network evaluation for identifying potential drug targets. *J Comput Biol* 17: 669-684.
2. Cui T, Zhang L, Wang X, He ZG (2009) Uncovering new signaling proteins and potential drug targets through the interactome analysis of *Mycobacterium tuberculosis*. *BMC Genomics* 10: 118.
3. Kotlyar M, Fortney K, Jurisica I (2012) Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods* 57: 499-507.
4. Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93: 13-20.
5. Huynen MA, Snel B, von Mering C, Bork P (2003) Function prediction and protein networks. *Curr Opin Cell Biol* 15: 191-198.
6. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751-753.
7. Enright AJ, Iliopoulos I, Kyripides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86-90.
8. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol* 1: 93-108.
9. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 23: 324-328.
10. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96: 4285-4288.
11. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 10: 2763-2788.
12. Gavin AC, Bösch M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141-147.
13. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623-627.
14. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast interactome. *Proc Natl Acad Sci U S A* 98: 4569-4574.
15. Lievens S, Lemmens I, Tavernier J (2009) Mammalian two-hybrids come of age. *Trends Biochem Sci* 34: 579-588.
16. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, et al. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 30: 303-305.
17. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, et al. (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res* 37: D767-D772.
18. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, et al. (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 38: D525-D531.
19. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38: W214-W220.
20. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 39: D698-D704.
21. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci U S A* 99: 7821-7826.

22. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555.
23. Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jørgensen C, et al. (2007) Systematic discovery of *in vivo* phosphorylation networks. *Cell* 129: 1415-1426.
24. Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Phys Rev E Stat Nonlin Soft Matter Phys* 84: 066122.
25. Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103: 8577-8582.
26. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A* 105: 1118-1123.
27. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.
28. King AD, Przulj N, Jurisica I (2004) Protein complex prediction via cost-based clustering. *Bioinformatics* 20: 3013-3020.
29. Li XL, Tan SH, Foo CS, Ng SK (2005) Interaction graph mining for protein complexes using local clique merging. *Genome Inform* 16: 260-269.
30. Geraci J, Liu G, Jurisica I (2012) Algorithms for systematic identification of small subgraphs. *Methods Mol Biol* 804: 219-244.
31. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, et al. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417: 399-403.
32. Causier B (2004) Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom Rev* 23: 350-367.
33. Taylor IW, Wrana JL (2012) Protein interaction networks in medicine and disease. *Proteomics* 12: 1706-1716.
34. Padilla-Parra S, Tramier M (2012) FRET microscopy in the living cell: different approaches, strengths and weaknesses. *Bioessays* 34: 369-376.
35. Hallworth R, Currall B, Nichols MG, Wu X, Zuo J (2006) Studying inner ear protein-protein interactions using FRET and FLIM. *Brain Res* 1091: 122-131.
36. Steinhoff HJ (2004) Inter- and intra-molecular distances determined by EPR spectroscopy and site-directed spin labeling reveal protein-protein and protein-oligonucleotide interaction. *Biol Chem* 385: 913-920.
37. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* 68: 803-812.
38. Ideker T, Krogan NJ (2012) Differential network biology. *Mol Syst Biol* 8: 565.
39. Hwang T, Park T (2009) Identification of differentially expressed subnetworks based on multivariate ANOVA. *BMC Bioinformatics* 10: 128.
40. Bisson N, James DA, Ivosev G, Tate SA, Bonner R, et al. (2011) Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol* 29: 653-658.
41. La D, Kong M, Hoffman W, Choi YI, Kihara D (2012) Predicting permanent and transient protein-protein interfaces. *Proteins*.
42. Cho KI, Lee K, Lee KH, Kim D, Lee D (2006) Specificity of molecular interactions in transient protein-protein interaction interfaces. *Proteins* 65: 593-606.
43. Xu D, Tsai CJ, Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 10: 999-1012.
44. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N (2001) Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins* 43: 89-102.
45. Fernandez A, Scott RL, Scheraga HA (2003) Amino-acid residues at protein-protein interfaces: Why is propensity so different from relative abundance? *J Phys Chem B* 107: 9929-9932.
46. Zellner H, Staudigel M, Trenner T, Bittkowski M, Wolowski V, et al. (2012) PresCont: predicting protein-protein interfaces utilizing four residue properties. *Proteins* 80: 154-168.
47. Ahmed MH, Spyraakis F, Cozzini P, Tripathi PK, Mozzarelli A, et al. (2011) Bound water at protein-protein interfaces: partners, roles and hydrophobic bubbles as a conserved motif. *PLoS One* 6: e24712.
48. Brooks DJ, Fresco JR, Lesk AM, Singh M (2002) Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Mol Biol Evol* 19: 1645-1655.
49. Chakrabarti P, Janin J (2002) Dissecting protein-protein recognition sites. *Proteins* 47: 334-343.
50. Suchanek M, Radzikowska A, Thiele C (2005) Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* 2: 261-267.
51. Leszyk J, Grabarek Z, Gergely J, Collins JH (1990) Characterization of zero-length cross-links between rabbit skeletal muscle troponin C and troponin I: evidence for direct interaction between the inhibitory region of troponin I and the NH2-terminal, regulatory domain of troponin C. *Biochemistry* 29: 299-304.
52. Hwang YJ, Granelli J, Lyubovitsky J (2012) Effects of zero-length and non-zero-length cross-linking reagents on the optical spectral properties and structures of collagen hydrogels. *ACS Appl Mater Interfaces* 4: 261-267.
53. Paramelle D, Miralles G, Subra G, Martinez J (2012) Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* 13: 438-456.
54. Paoletti AC, Zybailov B, Washburn MP (2004) Principles and applications of multidimensional protein identification technology. *Expert Rev Proteomics* 1: 275-282.
55. Zybailov BL, Florens L, Washburn MP (2007) Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol Biosyst* 3: 354-360.
56. Fritzsche R, Ihling CH, Götze M, Sinz A (2012) Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun Mass Spectrom* 26: 653-658.
57. Leitner A, Reischl R, Walzthoeni T, Herzog F, Bohn S, et al. (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics* 11: M111.
58. Ahmed FE (2008) Utility of mass spectrometry for proteome analysis: part I. Conceptual and experimental approaches. *Expert Rev Proteomics* 5: 841-864.
59. Singh P, Shaffer SA, Scherl A, Holman C, Pfuetzner RA, et al. (2008) Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. *Anal Chem* 80: 8799-8806.
60. Ye X, Luke B, Andresson T, Blonder J (2009) 18O stable isotope labeling in MS-based proteomics. *Brief Funct Genomic Proteomic* 8: 136-144.
61. Huang X, Tian C, Liu M, Wang Y, Tolmachev AV, et al. (2012) Quantitative proteomic analysis of mouse embryonic fibroblasts and induced pluripotent stem cells using 16O/18O labeling. *J Proteome Res* 11: 2091-2102.
62. Gao Q, Xue S, Shaffer SA, Doneanu CE, Goodlett DR, et al. (2008) Minimize the detection of false positives by the software program DetectShift for 18O-labeled cross-linked peptide analysis. *Eur J Mass Spectrom (Chichester, Eng)* 14: 275-280.
63. Gao Q, Xue S, Doneanu CE, Shaffer SA, Goodlett DR, et al. (2006) Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. *Anal Chem* 78: 2145-2149.
64. Liu H, Wang F, Xu W, May K, Richardson D (2013) Quantitation of asparagine deamidation by isotope labeling and liquid chromatography coupled with mass spectrometry analysis. *Anal Biochem* 432: 16-22.
65. Du Y, Wang F, May K, Xu W, Liu H (2012) Determination of deamidation artifacts introduced by sample preparation using 18O-labeling and tandem mass spectrometry analysis. *Anal Chem* 84: 6355-6360.
66. Li X, Courmoyer JJ, Lin C, O'Connor PB (2008) Use of 18O labels to monitor deamidation during protein and peptide sample processing. *J Am Soc Mass Spectrom* 19: 855-864.
67. Sevinsky JR, Brown KJ, Cargile BJ, Bundy JL, Stephenson JL Jr (2007) Minimizing back exchange in 18O/16O quantitative proteomics experiments by incorporation of immobilized trypsin into the initial digestion step. *Anal Chem* 79: 2158-2162.
68. Fernandez-de-Cossio J (2011) Mass spectrum patterns of 18O-tagged peptides labeled by enzyme-catalyzed oxygen exchange. *Anal Chem* 83: 2890-2896.
69. Valkenburg D, Burzykowski T (2011) A Markov-chain model for the analysis of high-resolution enzymatically 18O-labeled mass spectra. *Stat Appl Genet Mol Biol* 10.
70. Petrotchenko EV, Serpa JJ, Borchers CH (2010) Use of a combination of isotopically coded cross-linkers and isotopically coded N-terminal modification

- reagents for selective identification of inter-peptide crosslinks. *Anal Chem* 82: 817-823.
71. Zhang S, Liu X, Kang X, Sun C, Lu H, et al. (2012) iTRAQ plus 180: a new technique for target glycoprotein analysis. *Talanta* 91: 122-127.
72. Aggarwal K, Choe LH, Lee KH (2006) Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct Genomic Proteomic* 5: 112-120.
73. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, et al. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3: 1154-1169.
74. Kalkhof S, Sinz A (2008) Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem* 392: 305-312.
75. Bich C, Maedler S, Chiesa K, DeGiacomo F, Bogliotti N, et al. (2010) Reactivity and applications of new amine reactive cross-linkers for mass spectrometric detection of protein-protein complexes. *Anal Chem* 82: 172-179.
76. Lauber MA, Reilly JP (2010) Novel amidinating cross-linker for facilitating analyses of protein structures and interactions. *Anal Chem* 82: 7736-7743.
77. Li H, Zhao Y, Phillips HI, Qi Y, Lin TY, et al. (2011) Mass spectrometry evidence for cisplatin as a protein cross-linking reagent. *Anal Chem* 83: 5369-5376.
78. Karadzic I, Maupin-Furlow J, Humbard M, Prunetti L, Singh P, et al. (2012) Chemical cross-linking, mass spectrometry, and *in silico* modeling of proteasomal 20S core particles of the haloarchaeon *Haloflex volcanii*. *Proteomics* 12: 1806-1814.
79. Mouradov D, King G, Ross IL, Forwood JK, Hume DA, et al. (2008) Protein structure determination using a combination of cross-linking, mass spectrometry, and molecular modeling. *Methods Mol Biol* 426: 459-474.
80. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, et al. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A* 97: 5802-5806.
81. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, et al. (2008) Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5: 315-318.
82. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, et al. (2012) Identification of cross-linked peptides from complex samples. *Nat Methods* 9: 904-906.
83. Schelté P, Boeckler C, Frisch B, Schuber F (2000) Differential reactivity of maleimide and bromoacetyl functions with thiols: application to the preparation of liposomal dipeptide constructs. *Bioconjug Chem* 11: 118-123.
84. Zara JJ, Wood RD, Boon P, Kim CH, Pomato N, et al. (1991) A carbohydrate-directed heterobifunctional cross-linking reagent for the synthesis of immunoconjugates. *Anal Biochem* 194: 156-162.
85. Komatsu T, Oguro Y, Teramura Y, Takeoka S, Okai J, et al. (2004) Physicochemical characterization of cross-linked human serum albumin dimer and its synthetic heme hybrid as an oxygen carrier. *Biochim Biophys Acta* 1675: 21-31.
86. Jacobsen RB, Sale KL, Ayson MJ, Novak P, Hong J, et al. (2006) Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. *Protein Sci* 15: 1303-1317.
87. Brown WE, Wold F (1973) Alkyl isocyanates as active-site-specific reagents for serine proteases. Identification of the active-site serine as the site of reaction. *Biochemistry* 12: 835-840.
88. El-Shafey A, Tolic N, Young MM, Sale K, Smith RD, et al. (2006) "Zero-length" cross-linking in solid state as an approach for analysis of protein-protein interactions. *Protein Sci* 15: 429-440.
89. Schilling B, Row RH, Gibson BW, Guo X, Young MM (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom* 14: 834-850.
90. Dubinsky L, Krom BP, Meijler MM (2012) Diazirine based photoaffinity labeling. *Bioorg Med Chem* 20: 554-570.
91. Brunner J (1993) New photolabeling and crosslinking methods. *Annu Rev Biochem* 62: 483-514.
92. Dormán G, Prestwich GD (1994) Benzophenone photophores in biochemistry. *Biochemistry* 33: 5661-5673.
93. Gomes AF, Gozzo FC (2010) Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J Mass Spectrom* 45: 892-899.
94. Wu YI, Frey D, Lungu OI, Jaehrig A, Schlichting I, et al. (2009) A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* 461: 104-108.
95. Beck-Sickinger AG, Budisa N (2012) Genetically encoded photocrosslinkers as molecular probes to study G-protein-coupled receptors (GPCRs). *Angew Chem Int Ed Engl* 51: 310-312.
96. Grunbeck A, Huber T, Sachdev P, Sakmar TP (2011) Mapping the ligand-binding site on a G protein-coupled receptor (GPCR) using genetically encoded photocrosslinkers. *Biochemistry* 50: 3411-3413.
97. Yanagisawa T, Hino N, Ibrah F, Mukai T, Sakamoto K, et al. (2012) Wide-range protein photo-crosslinking achieved by a genetically encoded N(ε)-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Mol Biosyst* 8: 1131-1135.
98. Bousquet-Dubouch MP, Baudelet E, Guérin F, Matondo M, Uttenweiler-Joseph S, et al. (2009) Affinity purification strategy to capture human endogenous proteasome complexes diversity and to identify proteasome-interacting proteins. *Mol Cell Proteomics* 8: 1150-1164.
99. Sutherland BW, Toews J, Kast J (2008) Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. *J Mass Spectrom* 43: 699-715.
100. Toews J, Rogalski JC, Kast J (2010) Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal Chim Acta* 676: 60-67.
101. Klockenbusch C, O'Hara JE, Kast J (2012) Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal Bioanal Chem* 404: 1057-1067.
102. Toews J, Rogalski JC, Clark TJ, Kast J (2008) Mass spectrometric identification of formaldehyde-induced peptide modifications under *in vivo* protein cross-linking conditions. *Anal Chim Acta* 618: 168-183.
103. Hou J, Liu X, Shen J, Zhao G, Wang PG (2012) The impact of click chemistry in medicinal chemistry. *Expert Opin Drug Discov* 7: 489-501.
104. Chowdhury SM, Du X, Tolić N, Wu S, Moore RJ, et al. (2009) Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. *Anal Chem* 81: 5524-5532.
105. Luo J, Fishburn J, Hahn S, Ranish J (2012) An integrated chemical cross-linking and mass spectrometry approach to study protein complex architecture and function. *Mol Cell Proteomics* 11: M111.
106. Zhang H, Tang X, Munske GR, Tolic N, Anderson GA, et al. (2009) Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol Cell Proteomics* 8: 409-420.
107. Evaristo GP, Verhaert PD, Pinkse MW (2012) PTM-driven differential peptide display: survey of peptides containing inter/intra-molecular disulfide bridges in frog venoms. *J Proteomics* 77: 215-224.
108. Anusiewicz I, Berdys-Kochanska J, Simons J (2005) Electron attachment step in electron capture dissociation (ECD) and electron transfer dissociation (ETD). *J Phys Chem A* 109: 5801-5813.
109. Wu SL, Jiang H, Lu Q, Dai S, Hancock WS, et al. (2009) Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. *Anal Chem* 81: 112-122.
110. Agarwal A, Diedrich JK, Julian RR (2011) Direct elucidation of disulfide bond partners using ultraviolet photodissociation mass spectrometry. *Anal Chem* 83: 6455-6458.
111. Tang Y, Chen Y, Lichti CF, Hall RA, Raney KD, et al. (2005) CLPM: a cross-linked peptide mapping algorithm for mass spectrometric analysis. *BMC Bioinformatics* 6: S9.
112. Panchaud A, Singh P, Shaffer SA, Goodlett DR (2010) xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J Proteome Res* 9: 2508-2515.
113. Bennett KL, Kussmann M, Björk P, Godzwon M, Mikkelsen M, et al. (2000) Chemical cross-linking with thiol-cleavable reagents combined with

- differential mass spectrometric peptide mapping—a novel approach to assess intermolecular protein contacts. *Protein Sci* 9: 1503-1518.
114. Lee YJ, Lackner LL, Nunnari JM, Phinney BS (2007) Shotgun cross-linking analysis for studying quaternary and tertiary protein structures. *J Proteome Res* 6: 3908-3917.
115. Xu H, Hsu PH, Zhang L, Tsai MD, Freitas MA (2010) Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. *J Proteome Res* 9: 3384-3393.
116. Du X, Chowdhury SM, Manes NP, Wu S, Mayer MU, et al. (2011) Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. *J Proteome Res* 10: 923-931.
117. Kim PM, Korbel JO, Gerstein MB (2007) Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc Natl Acad Sci U S A* 104: 20274-20279.
118. Simkó GI, Gyurkó D, Veres DV, Nánási T, Csermely P (2009) Network strategies to understand the aging process and help age-related drug design. *Genome Med* 1: 90.
119. Abdollahi A, Schwager C, Kleeff J, Esposito I, Domhan S, et al. (2007) Transcriptional network governing the angiogenic switch in human pancreatic cancer. *Proc Natl Acad Sci U S A* 104: 12890-12895.

This article was originally published in a special issue, **Applications of Mass Spectrometry in Epigenetics** handled by Editor(s), Dr. Alan Tackett, University of Arkansas for Medical Sciences, USA