Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Chemical cross-linking with mass spectrometry: a tool for systems structural biology

Juan D Chavez and James E Bruce

Biological processes supporting life are orchestrated by a highly dynamic array of protein structures and interactions comprising the interactome. Defining the interactome, visualizing how structures and interactions change and function to support life is essential to improved understanding of fundamental molecular processes, but represents a challenge unmet by any single analytical technique. Chemical cross-linking with mass spectrometry provides identification of proximal amino acid residues within proteins and protein complexes, yielding low resolution structural information. This approach has predominantly been employed to provide structural insight on isolated protein complexes, and has been particularly useful for molecules that are recalcitrant to conventional structural biology studies. Here we discuss recent developments in cross-linking and mass spectrometry technologies that are providing large-scale or systems-level interactome data with successful applications to isolated organelles, cell lysates, virus particles, intact bacterial and mammalian cultured cells and tissue samples.

Address
Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA

Corresponding author: Bruce, James E (jimbruce@uw.edu)

Current Opinion in Chemical Biology 2019, **48**:8–18

This review comes from a themed issue on **Omics**

Edited by **Ileana M Cristea** and **Kathryn S Lilley**

https://doi.org/10.1016/j.cbpa.2018.08.006

Introduction

Life relies on a vast ensemble of highly dynamic yet organized and regulated set of protein conformations and interactions to carry out biological functions. Visualizing dynamic structural and interaction changes of proteins in their native environments has the potential to greatly advance understanding of function in all areas of biology. Structural biology has primarily relied on three techniques, X-ray crystallography, NMR and Cryo-EM, to provide detailed structural information on proteins and protein complexes. Together these techniques have provided a wealth of information on the molecular structure and function of proteins, with over 140 000 structures currently deposited in the Protein Data Bank (PDB) [1]. A caveat is that these techniques generally rely on producing samples containing large amounts of highly purified proteins of interest, removed from their native environment. Furthermore, the majority of structures in the PDB are partial structures, representing static snapshots of a protein, void of any information on the conformational dynamics the protein undergoes while functioning within the living organism. Chemical cross-linking with mass spectrometry has emerged as a technique able to bridge structural biology and systems biology and provide structural information on protein complexes as they exist within a biological system.

Chemical cross-linking has long been recognized as a valuable tool for preserving the structure of biological systems through the formation of covalent bonds [2]. Several decades ago cross-linking experiments were used to provide low resolution structural information on proteins and protein complexes [3,4]. With the emergence of biological mass spectrometry came an analytical means to identify the specific cross-linked amino acid residues in a high throughput manner [5], and the combination of cross-linking with mass spectrometry (XL-MS, also CL-MS) has been established as a new technique for structural biology alongside the traditional methods [6•]. Although ultimately a relatively low resolution structural technique, the primary advantages of XL-MS are the ability to probe protein complexes of unlimited size, gain information on the native ensemble of protein conformations, and provide insight into protein structures and interactions as they exist *in vivo* [6•,7,8]. Continued advances made in recent years to mass spectrometry instrumentation, methodology and informatics now enable XL-MS to be applied to many diverse biological questions and gain new insight on the interactome. The technique of XL-MS encompasses a wide range of options that can be tailored to address a breadth of scientific enquiries. From selection from a variety of molecular cross-linkers, sample processing, LC–MS methodology and data processing and visualization tools, researchers planning to utilize XL-MS need to familiarize themselves with these options to select the best combination to suit their particular research questions. For this purpose, we refer readers to a number of recent review articles cover various aspects of XL-MS in detail that will not be covered here [6•,9–14]. This review focuses primarily on the application of XL-MS to complex samples

(i.e. intact cells, tissues, and so on) to gain structural and interaction information on a systems level.

Molecular features of chemical cross-linkers

The vast majority of XL-MS experiments carried out to date utilize homo-bifunctional cross-linking reagents that predominately react with primary amines at the ε-amino group of Lys side chain and protein N-termini. Primary amines are excellent targets for cross-linking due to their reactivity and relatively high abundance on the surface of proteins. The widespread commercial availability of N-hydroxysuccinimide (NHS) esters has made them by far the most common amine-reactive groups used in homo-bifunctional cross-linkers, although other esters including N-hydroxyphthalimide, hydroxybenzotriazole, and 1-hydroxy-7-azabenzotriazole have been used and demonstrate better reaction efficiency and kinetics [15]. Imidates are another class of amine targeting reactive groups utilized in cross-linking, however they suffer from side reactions unless the cross-linking reaction is carried out under alkaline conditions (pH > 10), limiting their use for cross-linking in physiological conditions [16,17]. This prompted Lauber et al. to utilize the suberthioimidate group as an alternative which reacts with amines at physiological pH, avoids the side reactions with imidates, and maintains the charge on the protein surface [17]. Application of the suberthioimidate cross-linker diethylsuberthioimidate (DEST) was demonstrated by Lauber *et al.* in a structural study of the *Escherichia coli* ribosome [18•]. Because suberthioimidates maintain the charge on the Lys, they can be beneficial for downstream SCX enrichment, fragmentation in the mass spectrometer and potentially less disruptive to the native protein structure making them attractive alternatives to NHS [16,17,19]. Beyond primary amines, cross-linking strategies targeting the acidic side chains in Asp and Glu have been developed. Initial demonstrations of cross-linking acidic residues utilized dihydrazide compounds which require a large excess of coupling agents to drive the reaction, limiting their usefulness for application to cellular systems [20,21]. Utilizing a double activation strategy and a diamine compound, Fioramonte *et al.* demonstrated a multiplexed cross-linking strategy linking acid groups to each other or acid groups to Lys and Ser [22]. Recently, diazo compounds been shown to cross-link acidic residues at neutral pH without the need for activating agents [23,24]. The resulting ester linkages generated from diazo based cross-linkers have also been shown to be cleavable at low collisional activation energy within the mass spectrometer, potentially making them useful for large scale acidic group cross-linking in complex cellular samples [24]. Cross-linkers targeting hydroxyl groups [25] and thiol reactive cross-linkers targeting Cys [26] have also been demonstrated and offer potential for increased depth of interactome analyses. Photo-reactive, aldehyde-based and other cross-linker chemistries which are more promiscuous in their targets, make identification of the site of cross-linking difficult and currently limit their use in structural interactome studies. Although photo-reactive cross-linkers can be utilized in combination with affinity purification mass spectrometry to identify interacting partners from complex mixtures such as cell lysates [27], without the ability to identify the specific residue sites of cross-linking limited structural information is obtained. Similarly, while formaldehyde and glutaraldehyde are some of the oldest and widely applied cross-linkers in biology and offer the beneficial characteristics of small size, high solubility, cell permeability and high reactivity the identification of the sites of cross-linking from aldehyde cross-linkers remains challenging particularly from very complex biological samples beyond model systems [28]. These issues can be overcome in part by working with relatively simple systems consisting of a small number of purified proteins, where using a combination of cross-linkers with different reactivities have found use in determining the structure of purified proteins by increasing the density of observed cross-links [29].

In addition to the variety of reactive groups, cross-linkers also contain differing spacer arm lengths which range from zero-length to tens of angstroms. The spacer arm length ultimately limits the distance between which two residues can be linked, and thus provides structural information in terms of an upper bound distance constraint. Zero-length cross-linkers provide the tightest distance restraints (generally ~10 Å Cα–C

α) as they covalently link two side chains without adding additional atoms. On the other hand, zero-length cross-linked species are generally more difficult to identify than peptides linked with longer cross-linkers that incorporate additional functionalities [30]. It would seem that choosing a smaller spacer arm length would lead to tighter restraints and thus higher resolution structural information. However, in practice it has been observed by simulation [31] and experimentally that a large percentage of cross-link distances when mapped against protein crystal structures actually exceed the theoretical spacer arm length [32]. This observation suggests that features of the protein structure itself, such as local chemical environments influencing residue reactivity, peptide backbone flexibility and conformational dynamics are more critical determinants of what residues get linked. This also explains why a cross-linker with a spacer arm of near 30 Å can produce a similar distribution of observed cross-link distances to linkers with a spacer arm near 11.4 Å, such as DSS/BS3 [33••]. In fact, cross-links exceeding the expected cross-linker distance