Biochimica et Biophysica Acta 1864 (2016) 123–129

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics☆

Bao Quoc Tran, David R. Goodlett, Young Ah Goo *

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 7 May 2015
Accepted 18 May 2015
Available online 27 May 2015

Keywords:
Cross-linking
Mass spectrometry
Protein-protein interaction
Protein structure
Informatics

ABSTRACT

For the analysis of protein-protein interactions and protein conformations, cross-linking coupled with mass spectrometry (CXMS) has become an essential tool in recent years. A variety of cross-linking reagents are used to covalently link interacting amino acids to identify protein-binding partners. The spatial proximity of cross-linked amino acid residues is used to elucidate structural models of protein complexes. The main challenges for mapping protein-protein interaction are low stoichiometry and low frequency of cross-linked peptides relative to unmodified linear peptides as well as accurate and efficient matches to corresponding peptide sequences with low false discovery rates for identifying the site of cross-link. We evaluate the current state of chemical cross-linking and mass spectrometry applications with the special emphasis on the recent development of informatics data processing and analysis tools that help complexity of interpreting CXMS data. This article is part of a Special Issue entitled:Physiological Enzymology and Protein Functions.

1. Introduction

Deciphering protein structure and protein-protein interactions has provided insight into protein conformation and often led to the characterization of once unknown protein functions. When proteins interact, conformational changes lead to formation of transient or stable protein complexes that control various cellular processes. Recently, significant effort has been made toward structural elucidation of protein complexes using mass spectrometry [1–4]. This effort has centered around the use of chemical cross-linking reagents to create covalent bonds that can subsequently be analyzed by mass spectrometry. Informatics tools to uncover the cross-linked amino acids have become more and more important to interpreting the complex nature of protein structures and interactions. A notable advancement has been made for the chemical cross-linking technology, by virtue of improved access to high mass accuracy and high resolution mass spectrometers, and more recently various computational tools enabling identification and validation of large number of cross-linking sites with high efficiency.

A typical cross-linking experimental step entails the following. 1) Protein cross-linking reaction where there are a number of different linker types available including the ones with stable isotopes to enable detection and quantification, 2) Isolating/separating the cross-linked proteins from non-cross-linked to reduce the false positive rate, 3) Digestion of proteins into peptides by enzymes or chemical reagents, 4) Cross-linked peptide enrichment by physio-chemical properties, 5) MS method optimization for cross-linked peptide detection, and 6) Data processing to map cross-linked sites (Fig. 1).

Chemical cross-linking coupled with mass spectrometry (CXMS) workflows require analytical methods that are specific and sensitive in detecting cross-linked peptides in the presence of a large pool of non-cross-linked peptides. The most popular and commonly used mass spectrometry technique for cross-linking analysis has been the “bottom-up” proteomic approach. In this approach a cross-linked protein complex is cleaved to peptides using enzymatic or chemical methods and then cross-linked peptides can be affinity enriched for the downstream analysis. While it is a well developed and widely used technique, the high complexity of the peptide population therein often requires adequate peptide separation to increase sensitivity and specificity for detecting cross-linked peptides from non-cross-linked ones. Another approach using “top-down” analysis detects the presence of cross-links in the intact protein complex by fragmenting it in the MS instrument by use of electron-based dissociation methods without prior enzymatic digestion. While it provides a much more direct means of analysis than the bottom-up approach to determine interaction sites, it suffers from poor ionization and fragmentation efficiency due to the large size and nature of protein complexes. In order to overcome the drawbacks of these techniques, a third “middle-down” approach has also been utilized that produces a limited cleavage of the complex by specific enzymatic or chemical reactions generating larger peptides than in the bottom-up approach and thus less complex mixtures where the chance of finding the cross-linked sites is greater. These larger peptides are then

☆ This article is part of a Special Issue entitled: Physiological Enzymology and Protein Functions.

* Corresponding author at: Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201 USA. Tel.: +1 410 706 3380; fax: +1 410 706 0886.
E-mail addresses: btran@rx.umaryland.edu (B.Q. Tran), dgoodlett@rx.umaryland.edu (D.R. Goodlett), ygoo@rx.umaryland.edu (Y.A. Goo).

http://dx.doi.org/10.1016/j.bbapap.2015.05.015
1570-9639/

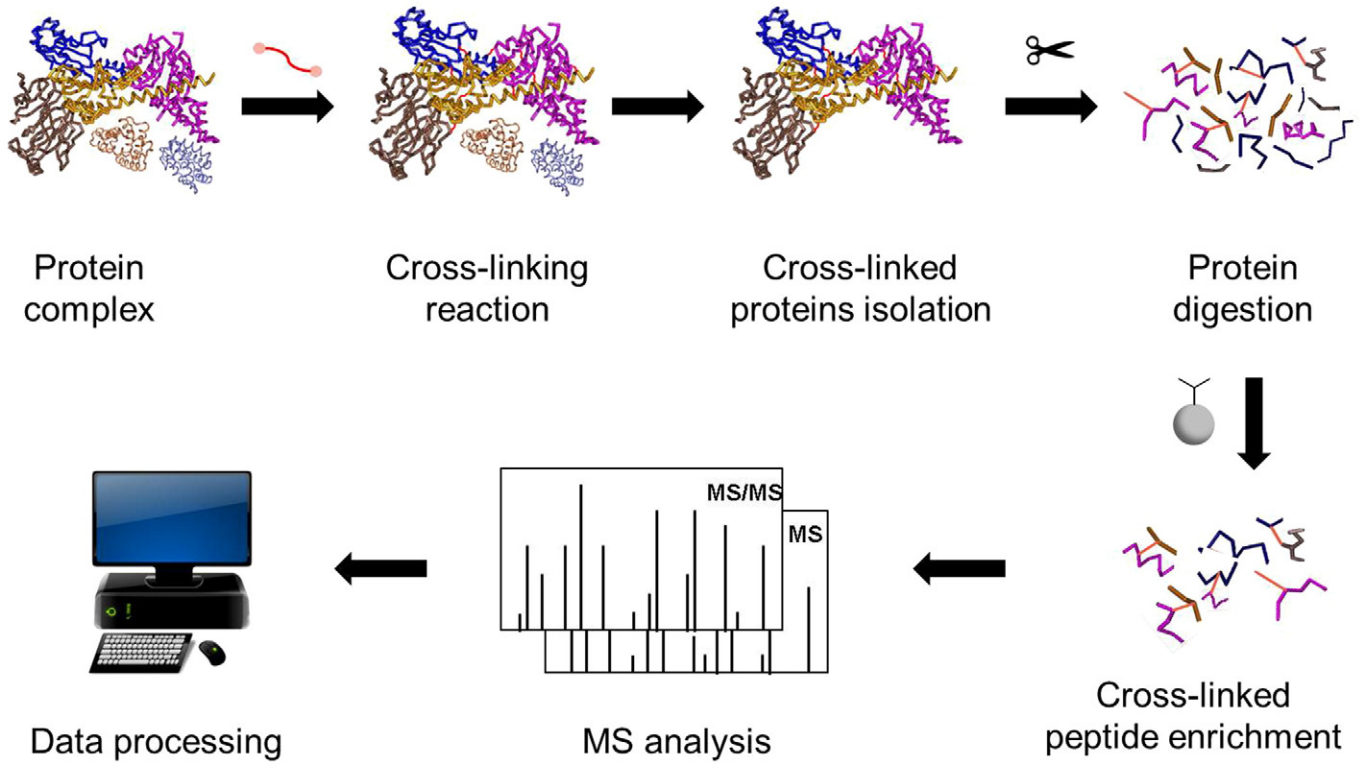

Fig. 1. A schematic representation of a typical cross-linking experimental workflow.

subject to the same fragmentation methods as in the bottom-up and top-down methods. To enhance the mass spectrometry detection, cross-linked peptides may incorporate functional labeling groups that produce a mass difference between heavy- and light-stable isotope forms [2,5–11], or add a specific moiety that is cleavable during MS/MS fragmentation thus facilitating detection [6,12–14].

In most cases bioinformatics that facilitate detection of cross-link peptides have been limited to a small number of protein sequence database and often used for targeted analysis in looking for cross-linked sites with prior knowledge about what protein complexes are interacting. However more recently, a few tools that can detect cross-links from complex samples using larger protein databases by either use of isotopically tagged cross-linkers or cleavable cross-linkers have demonstrated feasibility of automated high throughput identification of cross-link sites and peptides [8,15–17].

This perspective provides an overview of the current state and challenges associated with CXMS with special emphasis on the recent development of mass spectrometry techniques for CXMS and computational data processing and analysis tools.

2. Chemical cross-linking reagents

Cross-linking reagents play a key role in designing CXMS experiments because the choice of reagent will affect the specificity of the cross-linking reaction and the sensitivity of downstream analysis for cross-linked products. Depending on the experiments to be conducted cross-linking reagents may be activated by chemical or light-induced methods without altering the native conformation of the proteins. A large number of cross-linking reagents have been developed and studied in-depth in the past [4,18–20]. A basic cross-linker comprises two cross-linking reactive groups joined by a spacer, which is usually a carbon chain with a defined length. The reactive groups are either identical or different producing homo or hetero bifunctional cross-linkers, respectively. Use of spacers with defined lengths in cross-linkers provides useful information for the proximity between two cross-linked sites for the structural elucidation [21]. A combination of cross-linkers with different length could be useful to characterize the quaternary structure of complex assembly and three-dimensional (3D) structure as shown in proteasomal 20S core particles of the haloarchaeon *Haloferax volcanii* [22]. Popular for their simplicity carbodiimides, like 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) [23,24], are known as zero length cross-linkers because after the crosslink forms there is no trace of the reagent remaining, e.g. these are used to form an amide bond between carboxyl chain containing amino acid and a primary amine. However, due to obstructions within the protein interior, a cross-linker theoretically designed to span a known distance might not always form a covalently bonded crosslink [18].

The specificity of a cross-linker is one of the most important properties to consider when performing cross-linking experiments as it may impact on the level of cross-linking product diversity. Cross-linkers with less specificity, such as sulfonyl halides [25] or non-specific photo-inducible aryl azide compounds [4], are preferred for low resolution 3D protein structure study. For structural studies of small or mid-size proteins that have fewer possible cross-linked sites, use of a less specific cross-linker can be helpful to assign the cross-linking sites. In contrast, a cross-linker with high specificity could provide less ambiguity for cross-link assignments. High specificity is particularly important for large proteins or protein complexes because of the tremendous number of possible cross-linking combinations. Depending on the physio-chemical properties of a given cross-linker, the cross-linking specificity could be controlled by altering experimental conditions such as pH, protein/cross-linker ratio or irradiation time. For instance, maleimides are highly specific in reacting to sulfhydryl groups being roughly 1000 times faster than to a primary amine group at pH 7 [3,26–28]. The choice of cross-linking reactive group is highly dependent upon the availability of active sites on the protein where a covalent bond is formed. The most popular reactive cross-linkers can be categorized into four major groups [3]. First, reaction with primary amine groups such as amine-reactive cross-linkers like N-hydroxysuccinimide (NHS) [29,30],