Methods xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Methods

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

Formaldehyde cross-linking and structural proteomics: Bridging the gap

Savita Srinivasa [a,b], Xuan Ding [a,b,d,1], Juergen Kast [a,b,c,*]

[a] The Biomedical Research Centre, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[b] Department of Chemistry, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[c] Centre for Blood Research, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[d] Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, United States

ARTICLE INFO

Article history:
Received 1 January 2015
Received in revised form 30 April 2015
Accepted 6 May 2015
Available online xxxx

Keywords:
Formaldehyde
Cross-linking
Mass spectrometry
Protein structure
Protein interactions
Biological applications

ABSTRACT

Proteins are dynamic entities constantly moving and altering their structures based on their functions and interactions inside and outside the cell. Formaldehyde cross-linking combined with mass spectrometry can accurately capture interactions of these rapidly changing biomolecules while maintaining their physiological surroundings. Even with its numerous established uses in biology and compatibility with mass spectrometry, formaldehyde has not yet been applied in structural proteomics. However, formaldehyde cross-linking is moving toward analyzing tertiary structure, which conventional cross-linkers have already accomplished. The purpose of this review is to describe the potential of formaldehyde cross-linking in structural proteomics by highlighting its applications, characteristics and current status in the field.

1. Introduction

Structural proteomics is focused on building a comprehensive understanding of protein topology [1]. Chemical cross-linking along with mass spectrometry (MS) can elucidate protein geometry by introducing covalent linkages as distance constraints [2]. As one of the oldest cross-linkers, formaldehyde's well-known compatibility with physiological environments stems from its initial use to sustain clinical tissues and freeze protein function. Therefore, formaldehyde cross-linking could be a promising tool to probe protein structures, while maintaining biological context [3].

Formaldehyde's relatively small size (2.3 Å) allows it to quickly permeate through the cellular membrane and into intracellular compartments without external manipulation, and cross-link to stabilize proteins within close proximity of each other, unlike many larger cross-linkers and/or membrane-impermeable cross-linkers. Since it can cross-link rapidly, formaldehyde has the inherent ability to capture transient protein assemblies. The native environment of protein complexes is maintained, since formaldehyde cross-links form under physiological conditions. Finally, unlike most cross-linkers that target a defined, small number of reactive amino acid residues, formaldehyde's semi-specific reactivity permits cross-linking in the second step of the reaction to occur with several amino acid residues [4].

While formaldehyde chemistry is progressing towards enabling the translation of formaldehyde cross-links into structural models, bridging the existing gap has not been achieved [3,4]. Here, we describe the advancement of formaldehyde cross-linking and what obstacles remain to fully uncover its potential for structural proteomics.

2. Importance of formaldehyde cross-linking: common applications and key features

Formaldehyde is one of the oldest and widely applied cross-linkers in biological studies. Over a hundred years ago, formaldehyde was discovered to be a suitable tissue fixative. Its widespread use has generated thousands, if not millions of tissue samples with their structural and functional properties preserved with formaldehyde. Histopathological studies performed on these formalin-fixed tissues has unlocked valuable information to

Abbreviations: ChIP, chromatin immunoprecipitation; MS, mass spectrometry; MS/MS, tandem mass spectrometry; MALDI, Matrix-assisted laser desorption/ionization; LC, high performance reverse-phase chromatography; tcTPC, time-controlled transcardiac perfusion cross-linking; SCX, strong cation exchange chromatography; SEC, size exclusion chromatography; CID, collision-induced dissociation.

* Corresponding author at: 2222 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada. Fax: +1 604 822 7815.
E-mail address: kast@brc.ubc.ca (J. Kast).
[1] Present address

http://dx.doi.org/10.1016/j.ymeth.2015.05.006
1046-2023/

characterize and diagnose diseases [5]. Formaldehyde binding inactivates enzymes and destroys bacteria responsible for tissue degradation, allowing tissues to be sustained for a long period of time [6].

In the early 1920s, Ramon recognized formaldehyde's utility to develop vaccines because it inactivates toxins and viral proteins without destroying and possibly stabilizing antigens under mild formaldehyde reaction conditions [7–9]. Essentially, this demonstrated that formaldehyde can suspend the function of proteins without permanently damaging sites it modifies. Formaldehyde's minimal effect on antigens has enabled its conjunction with immunoprecipitation methods relying on accessible antigen sites for antibody binding to facilitate the investigation of protein–protein interactions.

The first application of formaldehyde cross-linking to capture specific interactions between biological species in their native environment was examining protein–DNA binding [10,11]. Unlike tissue fixation, which utilizes long reaction times and high concentrations of formaldehyde to extensively cross-link and maintain tissues long-term, these experiments call for selective cross-linking using short reaction times (~10–30 min) and low concentrations (~1%) of formaldehyde. In chromatin immunoprecipitation (ChIP), formaldehyde is directly introduced to cells to cross-link and maintain the spatial context of protein–DNA complexes, which are then isolated with an appropriate antibody-antigen interaction to obtain genomic binding sites [12,13]. These successful experiments promoted the expansion of the formaldehyde cross-linking approach to enhance the detection of true, specific protein–protein interactions by maintaining their cellular environment and spatial constraints, using similar experimental conditions [3,4].

Importantly, protein interaction analyses demonstrate the utilization of MS to identify interacting proteins preserved by formaldehyde [3]. Indeed, MS is a critical component in many current structural proteomics studies. Notably, the development of a polyacrylamide gel silver-staining procedure designed for subsequent MS analyses involved the incorporation of formaldehyde, suggesting formaldehyde-treated material can be subjected to MS [4,14]. Hence, MS is compatible with formaldehyde-induced chemical changes, albeit not requiring their actual detection.

Dimethyl labeling, a routine MS-based technology, exploits short reaction times and low formaldehyde concentrations to produce Schiff-base modifications. These are immediately reduced to dimethyl substituents with $NaCNBH_3$ instead of forming methylene cross-link bridges (Fig. 1) [15]. Stable isotope dimethyl labeling has successfully quantified protein expression levels in cells by being applied to both peptide digests and intact proteins [16]. Minimal side reactions and the conservation of charge states in formaldehyde modified peptides and proteins were reported, indicating that these modifications should not significantly disrupt chemical or physical properties of proteins. Dimethyl substituents were observed almost exclusively on lysines, illustrating the high specificity of the formaldehyde modification in the first step of the reaction at low formaldehyde concentrations and short incubation times [16,17].

As shown, formaldehyde cross-linking is suitable for *in vivo* biological applications and has been effectively utilized with MS, which are essential qualities for structural proteomics. A brief review of the existing structural proteomics approaches and their use of MS should help highlight areas in which formaldehyde may be used.

3. Purpose of structural proteomics and its common approaches

The fundamental aim of structural proteomics is defining three dimensional protein structures. Protein conformation and reactivity is dictated by its sequence. Protein function is governed by the interactions of protein structures in stable multi-subunit complexes as well as more transient assemblies. Such interactions are responsible for carrying out a multitude of tasks in cells from binding to small molecules during storage, transport and cellular signaling to serving as molecular switches, structural supports and catalysts [18]. Irregularities in protein structures at every level caused by mutating amino acids, denaturing or aggregating structures, or non-specific binding can indicate the presence of disease [19–21].

MS provides the means to study proteins as constantly changing molecules under physiological conditions, making it the primary technology implemented in structural proteomics. Bottom-up and top-down methodologies based on the MS analysis of peptide digests and intact proteins, respectively, have both been employed [22]. Reverse-phase high performance liquid chromatography MS (LC–MS) coupled with affinity purification analyses identify components of protein complexes and networks [23]. Cutting-edge instrumentation in tandem MS (MS/MS) maps proteins and their modifications at the amino acid residue level for high resolution geometry [24]. Native ion mobility MS characterizes protein conformation by measuring its cross-sectional area [25,26]. Imaging MALDI-MS can investigate the spatial arrangement of protein structures in intact tissues.

Approaches combining these MS technologies with chemical methods such as limited proteolysis, chemical surface modification and hydrogen–deuterium exchange, monitor the solvent accessibility of regions of a protein to observe conformational changes [1,27,28]. The enzymatic cleavage, modification, or deuterium exchange at particular amino acids directly correlates to the exposure of that respective region to the solvent and these can be used to determine structural changes [27,28]. Chemical cross-linkers form covalent bonds in proteins, which preserve their cellular context and introduce distance constraints to map their structure. Over the last decade, this low-resolution mapping technique along with MS has complemented molecular modeling experiments to verify topologies and contribute dynamic information of proteins [2,4,22,29–32]. However, formaldehyde has yet to match conventional cross-linkers engineered to produce straightforward MS analyses of three dimensional protein structures. The question remains, why has formaldehyde's potential for structural proteomics not been fully unleashed despite its compatibility with MS and the successful MS discovery of protein interactions *in vivo*?

4. Potential of formaldehyde cross-linking for structural proteomics

4.1. Formaldehyde cross linking in vivo: MS analyses of formalin fixed tissues and protein–protein interactions

Formaldehyde cross-linking in biological applications relies on one of two distinct strategies. The long-term preservation of protein structures in tissues employs high concentrations (>4%) of formaldehyde and long incubation times (several hours to days) to form non-specific cross-links with a higher yield. However, to access proteins for functional, and possibly structural, information, the yield of cross-linking should be reduced to be compatible with MS analysis [33]. In contrast, less extensive formaldehyde cross-linking conditions (0.05–1% formaldehyde and 5–20 min incubation) have already routinely been implemented to capture relevant interactions through *in vivo* formaldehyde cross-linking of protein complexes [3,4,34]. Such analysis is not dependent on detecting cross-linked species and relies on utilizing unmodified peptides to identify interacting proteins. Thus, maximizing formaldehyde cross-linking yield to detect cross-linked species with MS without sacrificing specificity has yet to be achieved *in vivo*.

Please cite this article in press as: S. Srinivasa et al., Methods (2015), http://dx.doi.org/10.1016/j.ymeth.2015.05.006