Journal of Chromatography A, 1348 (2014) 34–46

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Isolation of cross-linked peptides by diagonal strong cation exchange chromatography for protein complex topology studies by peptide fragment fingerprinting from large sequence databases

Hansuk Buncherd[a,1], Winfried Roseboom[a], Behrad Ghavim[a], Weina Du[b], Leo J. de Koning[a], Chris G. de Koster[a], Luitzen de Jong[a,*]

[a] Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 HX Amsterdam, The Netherlands
[b] Van't Hoff Institute of Molecular Sciences, University of Amsterdam, Science Park 904, 1098 HX Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form 22 April 2014
Accepted 24 April 2014
Available online 2 May 2014

Keywords:
Mass spectrometry
Cross-linking
Diagonal chromatography
Nuclear extract
Structural proteomics

ABSTRACT

Knowledge of spatial proximity of amino acid residues obtained by chemical cross-linking and mass spectrometric analysis provides information about protein folding, protein–protein interactions and topology of macromolecular assemblies. We show that the use of bis(succinimidyl)-3-azidomethyl glutarate as a cross-linker provides a solution for two major analytical problems of cross-link mapping by peptide fragment fingerprinting (PFF) from complex sequence databases, i.e., low abundance of protease-generated target peptides and lack of knowledge of the masses of linked peptides. Tris(carboxyethyl)phosphine (TCEP) reduces the azido group in cross-linked peptides to an amine group in competition with cleavage of an amide bond formed in the cross-link reaction. TCEP-induced reaction products were separated by diagonal strong cation exchange (SCX) from unmodified peptides. The relation between the sum of the masses of the cleavage products and the mass of the parent cross-linked peptide enables determination of the masses of candidate linked peptides. By reversed phase LC–MS/MS analysis of secondary SCX fractions, we identified several intraprotein and interprotein cross-links in a HeLa cell nuclear extract, aided by software tools supporting PFF from the entire human sequence database. The data provide new information about interacting protein domains, among others from assemblies involved in splicing.

1. Introduction

Chemical cross-linking to fix cells or tissues for microscopic analysis or to identify interacting proteins has been a valuable approach for a long time to understand biological structure and function. Recent developments in mass spectrometry, along with the emerging of large sequence databases and the availability of residue-specific bifunctional reagents, like bis(N-hydroxysuccinimidyl)esters, reacting specifically with amino groups, have enabled mapping of cross-links at the amino acid level [1–4]. Identification of chemically cross-linked residues in peptides by mass spectrometry (CXMS) can reveal protein interaction sites in complex mixtures [5,6] or living cells [7], and the obtained distance constraints defined by the length of the spacer of the cross-linker can be used for experimental validation of models of the 3-D structure of protein complexes [8]. Challenging analytical problems are encountered when applying CXMS to complex protein samples. Here we develop a new approach to meet these problems using bis(succinimidyl)-3-azidomethyl glutarate as a primary amine-specific cross-linker.

In general, mapping linked amino acid residues is accomplished by MS/MS analysis and database searching, also called peptide fragment fingerprinting (PFF) [9] of a tryptic digest of the cross-linked proteins. Three types of cross-linked peptides can be distinguished, designated type 0, type 1 and type 2 cross-links [10]. Type 0 cross-links, also called dead-end cross-links or monolinks, result from reaction of the cross-linker with a lysine residue, while the other reactive group is rendered inactive by hydrolysis. Type 1 cross-links connect two lysine residues within the same tryptic peptide. Type 2 cross-links are built up from two tryptic peptides. From a 3-D structural point of view, type 2 cross-links are the most informative ones of the three, since lysine residues far away from each other in the sequence of a particular protein (intraprotein cross-links) or lysines from different proteins can be connected in type 2 cross-links. This

* Corresponding author. Tel.: +31 20 5255691; fax: +31 20 5256971.
E-mail address: l.dejong@uva.nl (L. de Jong).
[1] Present address: Faculty of Medical Technology, Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand.

http://dx.doi.org/10.1016/j.chroma.2014.04.083
0021-9673/

implies that the spatial distance of connected residues at the time of reaction must have been less than the length of the spacer of the cross-linking agent.

To identify the connected peptides and the linked residues, a peptide database calculated from a sequence database is interrogated by a search engine with tandem mass spectrometric (MS_1MS_2) data in which MS_1 gives the mass and charge of a peptide ion selected for collision-induced dissociation (CID), and MS_2 results in the mass spectrum of the fragment ions generated by CID of a selected peptide ion. A cross-linked peptide is identified if the experimental MS_1MS_2 data sufficiently match a theoretical spectrum. However, application of CXMS with conventional bis(N-hydroxysuccinimidyl)esters as cross-linkers in complex samples is hampered by inefficient identification of type 2 cross-links.

The first attempt to identify type 2 cross-links in complex samples by PFF from large sequences databases has been published in 2008, concerning a study of a protein extract from a prokaryotic organism, i.e., *Escherichia coli*, [6]. In 2012 results of a similar study concerning a eukaryotic organism, *Caenorhabditis elegans* [5] have been published. Only a limited number of cross-links were identified at a false discovery rate of 5% in these studies. There is general agreement that further fractionation of peptides aimed to enrich the cross-linked species, besides one dimensional SCX chromatography used in these studies, is required to identify either more cross-links in these extracts or cross-links at all from an organism having a more complex sequence database than *C. elegans*. Only then the relative concentration of cross-linked peptides can become high enough to enable their data-dependent selection for MS_2. At the same time the amount of most normal linear peptides can become reduced to a level below the threshold for MS_2 selection. This will result in an increase of identified cross-linked peptides and in a decrease of false positives caused by the presence of many highly abundant normal peptides of which the MS_2 spectrum may accidentally resemble that of a cross-linked species. Here we use two-dimensional SCX chromatography, usually called diagonal chromatography, to isolate cross-linked peptides, and our experimental approach has enabled for the first time identification of numerous cross-links at an extremely low FDR in a complex protein extract by searching the entire human sequence database.

The inefficiency of type 2 cross-link mapping in complex digests is related both to low abundance of target peptides, and to limitations of available PFF approaches for these species. Low abundance is unavoidable, since cross-linking needs to be only partial both to prevent protease resistance in the treated proteins and to prevent cross-linking-induced conformational changes [11]. These requirements allow introduction of only a few cross-links per protein molecule, usually distributed over a large number of different sites, leading to substoichiometric amounts of cross-linked peptides upon digestion. To meet the limitations of low abundance, enrichment based on charge [6,12] and size [13] of type 2 cross-linked peptides has been described. However, enrichment by these methods is partial, due to charge and size overlap between target peptides and unmodified peptides. Affinity purification using cross-linkers provided with an affinity tag in the spacer, reviewed in [14], is another approach to attempt to circumvent the problem of low abundance. However, these reagents have been used sparsely to map cross-links in complex protein mixtures.

In one PFF approach, a database consisting of cross-linked peptides calculated from a protein database was used [6,15,16]. However, this strategy is limited to systems of relatively low complexity like isolated protein complexes, due to the quadratic increase in database size as a function of the number of tryptic peptides taken into account. The larger the search space, the larger is the risk for false positives. In a second option, a peptide database calculated from a protein database, instead of a database of cross-linked peptides, is interrogated with MS_1MS_2 data of the digest of the cross-linked protein sample [5,6]. The use of stable isotope-labeled cross-linkers facilitates this approach [6]. It enables both recognition of mass signals of target peptides in MS_1 spectra and distinction in MS_2 spectra of composing peptide-derived fragments with or without the cross-link. Also the use of secondary CID fragments for PFF has been described [5]. Despite recent improvements [17,18], these approaches are limited by lack of prior knowledge of the masses of composing peptides in a cross-link. To provide a solution for this problem, cross-linkers have been described with a spacer containing one or two gas phase cleavable bonds to enable composing peptide mass determination in an MS_2 step and PFF in an MS_3 step [19–22] or directly in the MS_2 step [23,24]. However, yet none of these reagents, besides a cross-linker named protein interaction reporter (PIR) provided with an affinity tag [7], have been applied to complex proteomes. Unfortunately, at the expense of resolving power, relatively long spacers are required for PIRs to accommodate the affinity tag and the two cleavage sites.

Here we explore the use of bis(succinimidyl)-2-azidomethyl glutarate (BAMG) [11] to meet both the limitation of available PFF approaches for cross-link mapping in complex samples and the problem of low abundance of target peptides. BAMG reacts with ε-amine groups of lysine residues and with α-amino groups of N-termini, thereby connecting proximate target residues via two amide bonds by a spacer with a length of 5 carbon atoms. BAMG-cross-linked peptides react with tris(carboxyethyl)phosphine (TCEP) along two competing pathways [11], resulting in products that form the basis of a new analytical strategy, including isolation of target peptides by diagonal strong cation exchange chromatography, for mapping of cross-links in complex samples by PFF from large protein databases.

We apply the new strategy to a high molecular weight fraction of a nuclear extract from HeLa cells. This preparation contains protein complexes involved in maintenance, duplication and expression of the genetic material and in other functions. The functional organization and dynamics of large assemblies is often poorly understood. Here we test the effectiveness of our method to identify cross-linked peptides in this complex preparation.

2. Materials and methods

2.1. Software tools

To facilitate mass spectrometric identification of cross-linked peptides we developed three software tools named Biner, hXLDB and Yeun Yan, described in detail in Sections 2.8, 2.9 and 2.10, respectively. Codes for software tools were written in Visual Basic language for applications. Software tools were operated on a Microsoft Visual Basic platform.

2.2. Proteins

Nuclear extract from HeLa cells [25] was obtained from Cilbiotec, Mons, Belgium (Cilbiotech.be). A protein fraction with a size distribution of approx. 150 kDa to 1 MDa was obtained by gel filtration of the nuclear extract from 4×10^8 HeLa cells (approx. 6 mg protein) on a Superose 6 10/300 GL column (GE Healthcare) operated on an Amersham Biosciences Akta FPLC platform in a medium containing 20 mM HEPES pH 7.9, 300 mM KCl, 0.2 mM EDTA, 0.1 mM DTT and 20% glycerol [26] (gel filtration buffer) at a flow rate of 0.5 ml min^{-1}. The absorbance at 280 nm of the effluent was continuously measured. Blue dextran (2 MDa, void marker), bovine thyroglobulin (669 kDa), horse apoferritin (443 kDa), yeast alcohol dehydrogenase (150 kDa), bovine serum albumin (67 kDa) and horse cytochrome *c* (12 kDa) were used as calibrants.