EuPA OPEN PROTEOMICS 3 (2014) 229–238

ScienceDirect
journal homepage: http://www.elsevier.com/locate/euprot

An integrated cross-linking-MS approach to investigate cell penetrating peptides interacting partners

Séverine Clavier[a,b], *Xiuxia Du*[c], *Sandrine Sagan*[a], *Gérard Bolbach*[a,b], *Emmanuelle Sachon*[a,b,*]

[a] *Laboratoire des Biomolécules, UMR7203 UPMC-ENS-CNRS-INSERM ERL1157, cc182, 4 place Jussieu, 75252 Paris Cedex 05, France*
[b] *Plateforme de spectrométrie de masse et protéomique, IBPS-UPMC, cc41, 7-9 quai saint Bernard, 75252 Paris Cedex 05, France*
[c] *Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, USA*

ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form
7 March 2014
Accepted 11 March 2014
Available online 24 March 2014

Keywords:
Chemical cross-linking
Cell penetrating peptides
Interacting partners
Mass spectrometry
Actin
Albumin

ABSTRACT

Cell penetrating peptides (CPPs) are attracting attention because of their ability to deliver biologically active molecules into cells. On their way they can interact with membrane and intracellular proteins. To fully understand and improve CPP efficiency as drug delivery tools, their partners need to be identified. To investigate CPP-protein complexes, chemical cross-linking coupled to mass spectrometry is a relevant method. With this aim, we developed an original approach based on two parallel strategies, an intact complex analysis and a bottom-up one, to have a global characterization of the cross-linked complexes composition as well as a detailed mapping of the interaction zones.

Biological significance: The robust and efficient cross-linking-MS workflow presented here can easily be adapted to any CPP-protein interacting system and could thus contribute to a better understanding of CPPs activity as cell-specific drug delivery tools. We validated the relevancy of this cross-linking-MS approach with two biologically active CPPs, $(R/W)_9$ and $(R/W)_{16}$, and two interacting protein partners, actin and albumin, previously reported using isothermal titration calorimetry (ITC) and NMR. Cross-linking-MS results obtained on these previous studies allowed us to go further by providing a detailed mapping of the interaction zones. The identified interaction zones between actin and CPPs $(R/W)_9$ and $(R/W)_{16}$ are biologically meaningful. Two cross-linked zones [46–57] and [202–210] of actin are indeed involved in the modulation of its dynamics. Moreover, [46–57] domain has also been described as one interaction domain for thymosin β4 whose actin binding can be displaced by competition with $(R/W)_{16}$ (NMR experiments).

* *Corresponding author at*: Laboratoire des Biomolécules, UMR7203 UPMC-ENS-CNRS-INSERM ERL1157 Plateforme de spectrométrie de masse et protéomique, IBPS-UPMC, France. Tel.: +33 144273234.
E-mail address: emmanuelle.sachon@upmc.fr (E. Sachon).
http://dx.doi.org/10.1016/j.euprot.2014.03.002

1. Introduction

Twenty years ago, the discovery of peptides able to ubiquitously cross cellular membranes commonly named cell penetrating peptides (CPPs), with very limited toxicity, launched a novel field in molecular delivery based on these non-invasive vectors. Most CPPs are positively charged peptides though the presence of few anionic or hydrophobic CPPs was also demonstrated. After a decade of debate on the trafficking routes of CPPs to the heart of cells, it is now more or less accepted that these peptides use concomitantly different internalization pathways, including pinocytosis and direct membrane translocation processes [1]. CPPs are generally considered as biologically inert intracellular delivery tools. However, some CPPs have intrinsically biological activity and are part of a recently described class of CPPs baptized bioportide [2].

For instance, previous studies showed, that only the two CPPs $(R/W)_9$ and $(R/W)_{16}$ (RRWWRRWRR and RRWRRWWRRWWRRWRR respectively) among other CPPs tested, are able to remodel the actin cytoskeleton in oncogen transformed NIH3T3/EWS-Fli cells once these CPPs had crossed the plasmamembrane [3]. In order to explain the actin-remodeling activity of the two CPPs, the hypothesis of a direct interaction with actin was tested. $(R/W)_9$ and $(R/W)_{16}$ peptides were actually found to directly interact *in vitro* with G-actin by NMR and ITC experiments [3] ($Kd \approx 10\,\mu M$ and $Kd = 0.4\,\mu M$, respectively). In addition, competitive binding experiments by NMR showed that $(R/W)_{16}$ was able to displace the actin sequestering protein thymosin β4 from G-actin [3].

It was also recently reported that arginine-rich CPPs interact with serum proteins like albumin, modifying their ability to internalize in cells [4]. Therefore our aim in this study was to analyze further the interaction of $(R/W)_9$ and $(R/W)_{16}$ with actin and albumin.

For this purpose, chemical cross-linking was chosen. Cross-linking reactions are conventionally based on the use of a bifunctional cross-linker, which is a carbon chain spacer bearing reactive sites at both ends that can either be identical (homobifunctional) or different (heterobifunctional). Reactive sites are mainly activated esters targeting either lysine residues (*e.g.* N-hydroxysuccinimide (NHS) ester) or cysteine residues (*e.g.* maleimide ester) although side reactions with tyrosine, threonine and serine have been reported [5].

By creating a covalent bond between two or more interacting partners, chemical cross-linking gives a snapshot of a molecule's environment and, combined with MS, constitutes a powerful tool to map protein-protein or peptide–protein interactions (distance constraints and interacting domains). However chemical cross-linking is often characterized by low reaction yields. In addition, a wide variety of cross-linked products are usually created. Therefore, cross-linkers bearing an affinity tag [6,7] allowing selective enrichment of the sample in cross-linked species have been developed. Cleavable cross-linkers [8,9], isotope labeled cross-linkers [10] or cross-linkers bearing a CHCA matrix moiety for MALDI analysis [11] improve detection and identification/characterization of the cross-linked species. Affinity purification using a tagged cross-linker or interacting partners [12] and SDS-PAGE are the most commonly used off-line techniques but chromatographic methods like strong cation-exchange (SCX) [13,14] and size exclusion chromatography (SEC) [15] constitute promising approaches to enrich samples in cross-linked species.

Analysis of low abundant cross-linked peptides requires high sensitivity for their detection and high mass accuracy (mass error < 10 ppm) for their identification since the number of combination of two peptides is enormous. Moreover tandem MS of cross-linked peptides is required for the characterization of the interaction zones.

A number of tailored softwares have been developed to deal with the selective acquisition or interpretation of these MS/MS spectra. Among others, FINDX [16] has been designed to selectively fragment inter-protein cross-links by LC–MALDI-TOF/TOF using $^{14}N/^{15}N$ mixed isotope strategy, xQuest [13] is dedicated to the search of isotopically tagged cross-linked peptides and CrossWork [17] or Xlink-Identifier [18] support label-free analyses of chemical cross-linking samples.

In this study, we set-up a general *in vitro* analytical workflow coupling cross-linking and mass spectrometry (cross-linking-MS), involving enrichment steps as well as manual or automated MS and MS/MS data processing to test potential interacting partners of any CPP sequence.

To validate our cross-linking-MS workflow, we studied the systems described above: $(R/W)_9$ or $(R/W)_{16}$ interacting with actin or albumin. For this purpose, CPP analogs suitable for chemical cross-linking experiments, were synthesized with the following sequences: $Biot(O_2)$-G_4-K-RRWWRRWRR-NH_2 and $Biot(O_2)$-G_4-K-RRWRRWWRRWWRRWRR-NH_2, respectively. The K residue added at the N-terminus of the peptides sequences allowed the cross-linking reaction, the biotin tag ($Biot(O_2)$) was added for purification purpose and was separated from the biologically active motif by a four G residues spacer. This {$Biot(O)_2$-G_n-K-} group is easy to add at the N-terminus of peptides either manually or automatically during peptide synthesis whatever the peptide sequence.

The originality of our approach resides in the comprehensive study of the cross-linking reaction mixture from two angles: a global view of the interacting system with an intact complex analysis combined to the precise characterization of interacting zones by a bottom-up analysis. The intact complex analysis is based on the MALDI-TOF analysis in linear mode of the cross-linking reaction mixtures and on the modeling of the spectra obtained using the in-house SIMUL-XL program. The bottom-up analysis consists in the tryptic digestion of the cross-linking reaction mixtures followed by the affinity purification (biotin/streptavidin) of the biotinylated cross-linked peptides and their analysis by tandem MS (MALDI-TOF/TOF and/or nanoLC–ESI-Orbitrap). The MS/MS data are either manually interpreted with the help of GPMAW software [19] (MALDI-TOF/TOF spectra) or automatically searched using Xlink-Identifier software [18] (nanoLC–ESI-MS/MS spectra).

For practical reasons the intact complex analysis was developed only with $(R/W)_9$ and both actin and albumin proteins. In contrast, the bottom-up analysis was performed using both $(R/W)_9$ and $(R/W)_{16}$ and both actin and albumin proteins.

2. Experimental

2.1. Materials

Bovin serum albumin and actin from rabbit muscle were purchased from Sigma (St. Louis, Missouri, USA). Promix 1, proteins standard mixture, was from LaserBio Labs (Sophia Antipolis, France). Bis[sulfosuccinimidyl] suberate (BS^3) was from Thermo Scientific (Waltham Massachussetts, USA) and the K100 stabilization kit from CovalX (mix of three cross-linkers 1,1′-(suberoyldioxy)bisazabenzotriazole (SBAT), 1,1′-(suberoyldioxy)bisbenzotriazole (SBBT) and 1,1′-(glutaroyldioxy)bisazabenzotriazole (GBAT)) (Schlieren, Switzerland). Trypsin Gold, Mass spectrometry Grade, was from Promega (Fitchburg, Wisconsin, USA). Dynabeads M280 streptavidin-coated magnetic beads were from Invitrogen (Carlsbad, New-Mexico, USA). The CPPs $(R/W)_9$: $Biot(O_2)$-G_4-K-RRWWRRWRR-NH_2 (*m*/z 2127.12) and $(R/W)_{16}$: $Biot(O_2)$-G_4-K-RRWRRWWRRWWRRWRR-NH_2 (*m*/z 3309.76) were synthesized (Fmoc strategy) and purified in-house. Ziptip® C_4 pipette tips were from Millipore (Darmstadt, Germany).

2.2. Cross-linking reactions

Monomeric actin (G-actin) purchased lyophilized in Tris, ATP, and $CaCl_2$ was dialysed against a 20 mM HEPES, 150 mM NaCl, 0.5 mM DTT, 200 μM $CaCl_2$, 200 μM ATP and 0.005% NaN_3 pH 8 buffer (Actin non-denaturing buffer). Stock solutions of CPP and actin were diluted in the actin non-denaturing buffer to obtain the desired concentrations (typically 10 μM). For CPP and albumin a 20 mM HEPES, 150 mM NaCl buffer was used. For all experiments, after 15 min pre-incubation of the CPP and protein partners, 50 mM cross-linkers solutions prepared *extemporanely* in dimethylformamide (DMF) for the K100 kit, in water for BS^3, were added to have a 2 mM final concentration of cross-linkers (200 folds higher than proteins concentration) and the reactions were allowed to proceed for 120 min at room temperature (RT) under gentle stirring. Typical cross-linking reactions (*XL*) volumes were 50 μL. For experiments with low peptides concentrations (1 or 0.1 μM), reaction volumes were adapted (200 μL and 2000 μL respectively) to allow purifying enough cross-linked peptides. Control reaction mixtures noted *CT* were performed in the same conditions but without cross-linkers. Cross-linking reactions were quenched by adding Tris-base (final concentration 15 mM).

2.3. Enzymatic trypsin digestion

25 μL of samples were submitted to in-solution tryptic digestion. Disulfide bridges were reduced (5 mM dithiothreitol) and cysteines alkylated (20 mM iodoacetamide). Trypsin digestion was conducted at 37 °C overnight (1:30 (w:w) protease-to-protein ratio).

2.4. Biotinylated cross-linked peptides affinity purification

A volume of tryptic digest corresponding to 80 pmol of CPP was incubated with 200 μg of streptavidin-coated magnetic beads for 60 min at RT under gentle stirring. Beads were conditioned before use and washed after incubation [12]. Finally, peptides were eluted from the beads with 5 μL 0.1 M hydrochloric acid (HCl) under gentle stirring (30 min).

2.5. MS analysis and data treatment

MALDI-TOF spectra were obtained with a MALDI-TOF/TOF *AB4700 Proteomics Analyzer* mass spectrometer (Applied Biosystems) in positive ions linear or reflector mode and delayed extraction. CHCA (Sigma) was used as the matrix and solubilized at 5 mg/mL in 1/1 ACN/0.1% TFA.

Intact complexes analysis in linear mode MALDI-TOF: 0.5 μL of non-digested reaction mixtures purified by C_4 Ziptip® were mixed with an equal volume of matrix and deposited on the sample holder. A total of 20,000 laser shots per sample were acquired in the *m*/z range 10,000–100,000 (focus mass 60,000).

Modelings of the spectra obtained for intact cross-linked complexes and corresponding controls were achieved using SIMUL-XL, a program developed in-house (Visual Basic V6.0).

Bottom-up analysis in reflector mode MALDI-TOF/TOF: After affinity purification, 0.5 μL of the elution mixture was mixed with 0.5 μL matrix and 0.5 μL were deposited on the sample holder. A total of 10,000 laser shots per spot were acquired in the *m*/z range 500–5000. External calibrations were realized using the peptide calibration standard Pepmix4 (Laser BioLabs). A manual interpretation of the spectra was done comparing the cross-linking reaction mixture (*XL*) and the peptide and protein mixture without cross-linker (*CT*) spectra. Differential peaks identified in *XL* spectrum were confronted with the cross-linked peptides lists generated *in silico* using General Protein Mass Analysis for Windows (GPMAW) version 6.1 [19] (Lighthouse Data, Odense, Denmark) with a number of missed-cleavages for trypsin of 2.

NanoLC–ESI-Orbitrap: The complete characterization of the cross-linked peptides was performed using an Ultimate 3000 Nano-HPLC system (Dionex) coupled with a LTQ-Orbitrap-XL mass spectrometer (Thermo Scientific). Samples were injected by the autosampler and concentrated on a trapping column (Pepmap, C_{18}, 300 μm × 5 mm, 5 μm, 100 Å, Dionex) with water containing 2% ACN and 0.1% formic acid (solvent A). After 10 min, the peptides were eluted onto the separation column (Pepmap, C_{18}, 75 μm × 150 mm, 2 μm 100 Å, Dionex) equilibrated with 98% solvent A. Peptides were separated using the gradient 0–50 min 2–40% solvent B (98% ACN + 0.1% formic acid), 50–60 min 40–60% solvent B, and 60–70 min 60% solvent B at a flow rate of 200 nL/min. The LTQ-Orbitrap mass spectrometer is outfitted with a nano ESI interface. Electrospray emitters were 360/20 μm o.d. × 10 μm i.d. fused-silica tips (PicoTip Emitter, Standard Coated SilicaTip, New Objective). The heated capillary temperature and spray voltage were 200 °C and 1.5 kV, respectively. Orbitrap spectra (automated gain control (AGC) 2×10^5) were collected from *m*/z 300–2000 at a resolution of 30,000 in the profile mode followed by data dependent sequential CID and HCD MS/MS spectra of the three

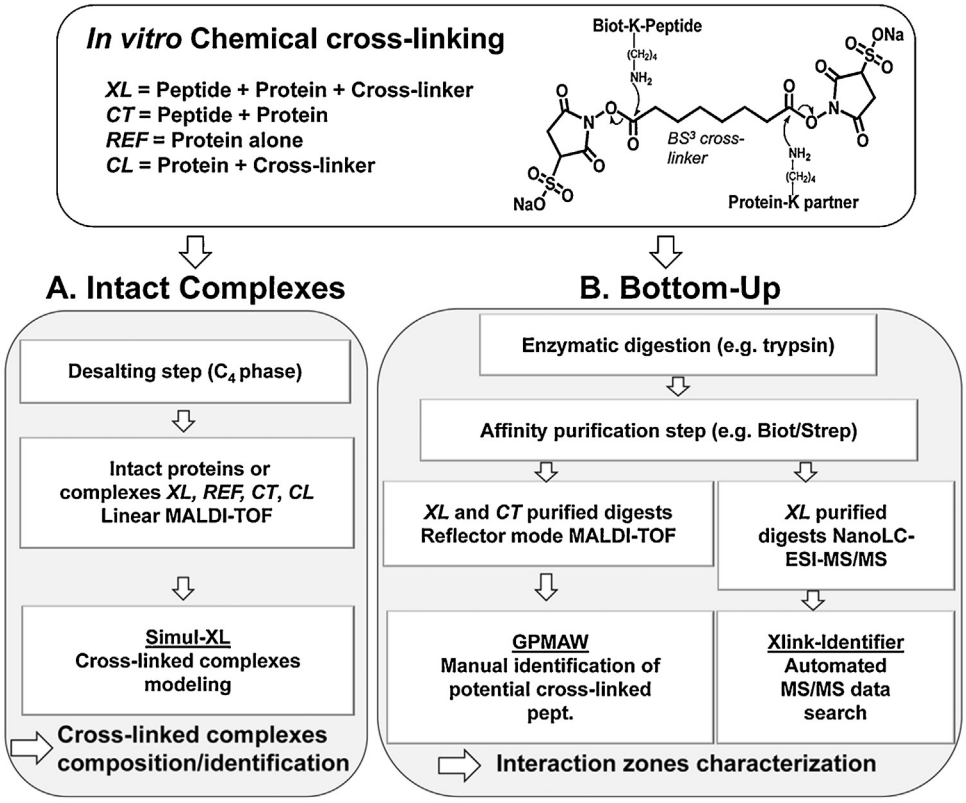

Fig. 1 – General analytical workflow to *in vitro* identify and characterize protein partner(s) of a known CPP. The chemical cross-linking reaction between a biotinylated CPP, the BS^3 cross-linker and a potential protein partner is presented (top scheme). The cross-linked complexes were characterized using an intact complex analysis (A) and a bottom-up one (B).

most intense ions with a normalized energy of 35 for both fragmentation modes. A dynamic exclusion time of 60 s was used to discriminate against previously analyzed ions.

MS/MS spectra were automatically searched using the Xlink-Identifier software [18]. Xlink-Identifier is a search engine for identifying and characterizing cross-linked peptides from label-free experiments (neither the peptide nor the cross-linker is isotopically labeled). It takes the MS/MS spectra in .dta or .mgf format and the protein sequences in FASTA format. Search parameters include the fragmentation technique, the nature of the dynamic modifications (carbamidomethylation (C), oxidation (M)), the maximum number of missed-cleavages, the precursor and fragment mass tolerance (respectively 10 ppm and 0.6 Da) and finally the definition of the cross-linker spacer arm mass involved in the formation of the cross-linking products (96.0211 Da for GBAT, 138.0618 Da for SBAT, SBBT or BS^3).

As mentioned, Xlink-Identifier is a fully automated search engine specifically for cross-linking analysis. It is equipped with a visualization module allowing researchers to examine the annotated MS/MS spectra and details of each matched peak.

3. Results and discussion

In this study, we synthesized analogs of $(R/W)_9$ and $(R/W)_{16}$ CPPs ($Biot(O_2)$-G_4-K-RRWWRRWRR-NH_2 and $Biot(O_2)$-G_4-K-RRWRRWWRRWWRRWRR-NH_2, respectively) adapted for cross-linking experiments. To study the interaction of these CPPs with protein partners, a cross-linking-MS approach integrating two parallel strategies, an intact complex analysis and a bottom-up one, was developed and optimized (Fig. 1).

3.1. *Chemical cross-linking reaction*

Actin is a globular 42 kDa protein that binds an ATP molecule and a divalent cation (Mg^{2+} or Ca^{2+}). These cofactors are necessary to maintain the integrity of the protein. Commercial actin was available as a lyophilized powder containing Tris buffer, which is not compatible with chemical cross-linking because of its reactive primary amine group. The sample was thus dialysed against an HEPES buffer also containing appropriate concentrations of salts and actin cofactors. Initial concentration of 10 μM was chosen for the two partners as the specificity of chemical cross-linking with activated esters was demonstrated in the low μM range [20].

We worked with a mixture of three cross-linkers (K100 stabilizing kit): SBAT, SBBT and GBAT (Fig. S-1). This mixture of cross-linkers is recommended to stabilize complexes with a molecular weight below 100 kDa and allows testing two spacer lengths at the same time (11.4 Å for SBAT and SBBT and 7.7 Å for GBAT). The K100 kit cross-linkers were solubilized in DMF. To make sure that the small proportion of this organic

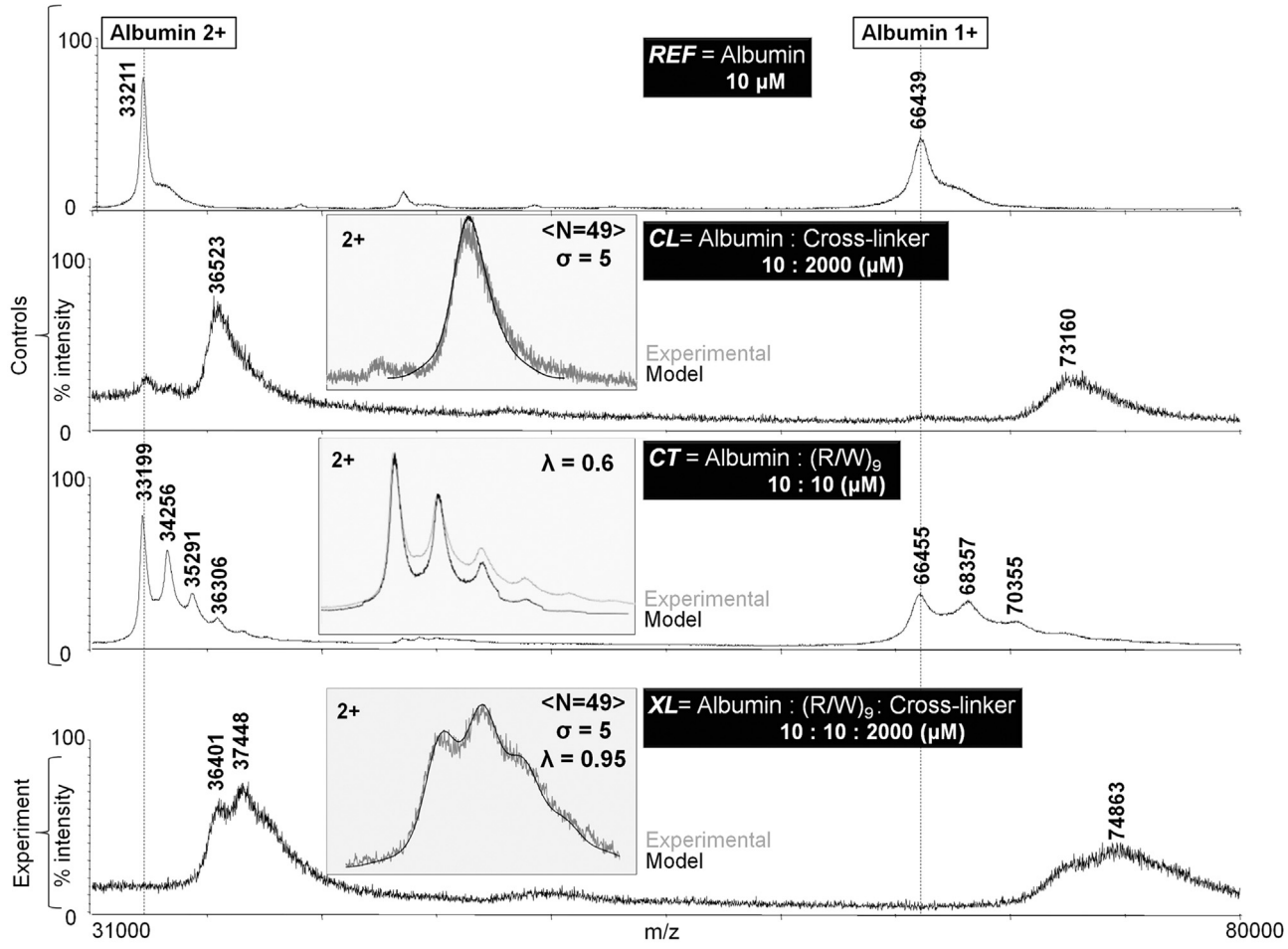

Fig. 2 – Positive ions linear mode MALDI-TOF of intact controls and chemical cross-linking reaction mixture. The controls (*REF*, *CL* and *CT*) are presented on the three top spectra, the cross-linking experiment (*XL*) in the lower spectrum. Mass spectra were modeled using SIMUL-XL program. Modelings (black) that best fit the experimental data (gray) are presented in a box (doubly charged species). σ: standard deviation, λ: average number of $(R/W)_9$, $\langle N \rangle$: average number of cross-linker.

solvent is not affecting the protein or peptide conformation and their interactions and thus the cross-linking reactions results, experiments were repeated with the BS^3 water soluble cross-linker which has the same spacer arm as SBAT or SBBT.

Supplementary Fig. 1 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

3.2. *Intact complex analysis: determination of $(R/W)_9$ CPP – protein ratio in intact cross-linked complexes*

Linear mode MALDI-TOF: MALDI-TOF is usually not a method of choice to study non-covalent complexes due to imperfect preservation of complexes during target preparation and unspecific non-covalent multimers desorption due to spatial proximity and high concentration in the matrix crystals. However it is well adapted to the study of covalent complexes formed using chemical cross-linking. Linear mode MALDI-TOF analyses of the intact complexes were carried out after a simple desalting step of the samples on a C_4 ZipTip®. The aim was to assess the efficiency of the cross-linking reaction by determining the average number of $(R/W)_9$ CPPs covalently attached per protein, for different reaction conditions.

To assess the contribution of the different components to the observed signal for a cross-linking reaction (*XL*), controls are necessary: the protein alone (*REF*), the protein plus the cross-linker only (*CL*) (to assess the mono-links, intra and inter cross-linked proteins), and the $(R/W)_9$ CPP and the protein mixture without the cross-linker (*CT*) (to assess non-covalent interactions that could remain even with the acidic CHCA dissociative matrix). Many precautions need to be taken in the preparation of these controls and in mass spectra acquisition (same laser fluence and focus mass as for *XL* mixture). To quantitatively describe the different mass spectra obtained (number of mono-links, number of peptides covalently or non-covalently attached) a dedicated program SIMUL-XL was developed in our laboratory (Visual Basic V6.0) and is available as an open source on request at gerard.bolbach@upmc.fr.

Modeling of intact interacting species: The modeling is based on the experimental peak profile of the protein alone (singly and multiply charged ions) (*REF* spectrum, Fig. 2). To each point of this profile a hypothetical distribution of $(R/W)_9$ or mono-links or $(R/W)_9$ + mono-links is added. The average number of

mono-links is noted $\langle N \rangle$ and of $(R/W)_9$ CPPs is λ. Mono-links for which the average number is much greater than one are added according to a gaussian law with a chosen standard deviation value σ, while $(R/W)_9$ CPPs are added according to a poisson law since the average number λ is less than one.

The final peak profile is build up by superimposition of all these profiles (Fig. S-2). The comparison of this convoluted profile with the experimental data allows adjusting the hypothetical distribution of $(R/W)_9$ or cross-linkers or $(R/W)_9$ + cross-linkers. The initial profile of the protein alone is large including the isotopic pattern, initial axial velocity effects and matrix adducts. In comparison, the isotopic pattern of the $(R/W)_9$ peptide and of the cross-linker are negligible and they are thus not taken into account. Simple gaussian and Poisson distributions were found to fit satisfactory the experimental data.

Supplementary Fig. 2 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

This work was performed for both interacting systems, $(R/W)_9$-albumin and $(R/W)_9$-actin, for two ratios 1:1 and 6:1. The 1:1 ratio is giving us an indication on the cross-linking reaction "yield" and the 1:6 ratio (large excess of peptide) was chosen to see if the CPP could have more than one interaction sites on the proteins. Spectra and modelings obtained for the $(R/W)_9$-albumin system in a 1:1 ratio are presented in Fig. 2.

The *CL* mixture was satisfactory fitted with a gaussian distribution of cross-linkers centered on $\langle N = 49 \rangle$ with $\sigma = 5$. This suggests a very good accessibility of the lysine residues (total 52). The same distribution was found to fit all the charge states demonstrating that the covalent attachment is the major interaction and that non-covalent attachments are negligible [21] even in the presence of a large excess of cross-linkers.

The *XL* mixture spectrum was correctly fitted for all the charge states (+1 to +4) using the previous cross-linker gaussian distribution ($\langle N = 49 \rangle$ and $\sigma = 5$) and a Poisson distribution of $(R/W)_9$ CPP with a mean value of $\lambda = 0.95$. Increasing the ratio $(R/W)_9$-albumin to 6:1, the mean value is $\lambda = 3$ indicating the existence of several binding sites on the albumin protein (Fig. S-3).

Supplementary Fig. 3 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

For *CT* samples, we observed that peaks can be fitted with a CPP Poisson distribution with a mean value strongly dependent on the charge state: 0.7 for +1 and 0.5 for +3. Similar results were found with both CPPs–protein ratios 1:1 and 6:1. This observation is in good agreement with in-source dissociation of the non-covalent complexes in MALDI-TOF [21].

All these results clearly show that a high cross-linking yield is observed for the $(R/W)_9$-albumin interacting system and that artifacts due to non-covalent interactions, if any, are negligible for *XL* reaction mixture.

For the $(R/W)_9$-actin interacting system in a 1:1 ratio, *CL* was fitted with $\langle N = 8 \rangle$ mono-links and $\sigma = 2$, indicating that some of the 17 lysines are likely less exposed than others to the cross-linkers. The best fit for the *XL* mixture was obtained with $\langle N = 6 \rangle$ mono-links and $\lambda = 0.3$ CPP (data not shown). Hence the cross-linking yield was low for the $(R/W)_9$-actin system. For a 6:1 ratio, the S/N was too low and mass spectra were thus not interpreted using SIMUL-XL program.

The results of this strategy devoted to intact complexes show an *in vitro* direct interaction of $(R/W)_9$ CPP with both albumin and actin. It also easily gives access to the efficiency of the cross-linking reaction, by providing an estimation of the average number of accessible lysine residues and covalently attached CPPs. Cross-linked complexes obtained with a relatively high yield could correspond to a rather homogeneous or heterogeneous population. This was further studied using the bottom-up analysis.

3.3. *Bottom-up analysis: characterization of the interacting zones between $(R/W)_9$ or $(R/W)_{16}$ CPPs and actin or albumin proteins*

MALDI-TOF analysis of the purified cross-linking digests revealed a robust, reproducible and efficient method to evaluate and optimize the cross-linking experiments workflow (evaluation of the efficiency of the affinity purification step, the specificity of the cross-linking reactions, the influence of the CPP to protein ratio etc.).

Purification of the biotinylated tryptic cross-linked peptides. Among the possible products of the cross-linking reactions, we had to distinguish between the mono-linked species on one of the partner (CPP or protein) of the complex, the intra- or inter-protein cross-linked products and the cross-linked products involving both the CPP and the protein, the latter being those of interest. After in-solution trypsin digestion of the cross-linking (*XL*) and control (*CT*) mixtures, species containing the tryptic biotinylated CPP were selectively enriched through a biotin/streptavidin affinity purification step using streptavidin-coated magnetic beads. An improvement of the initial protocol [12] was achieved by adding a 10 min incubation step of the beads in 100% ACN to remove non-specifically adsorbed species. The elution step was also adapted to the cross-linking-MS workflow to allow the elution of the cross-linking products in an aqueous phase compatible with the nanoLC–MS/MS system. Three different acidic elution solutions were compared: 0.1% TFA, 10% formic acid (FA) and 0.1 M HCl, the latter turned to be the most efficient. MALDI-TOF analysis of the streptavidin beads directly spotted onto the sample holder after treatment with 0.1 M HCl revealed that the great majority or even the totality of the cross-linked species were released from the beads.

Fig. S-4 emphasizes the absolute necessity of a purification/enrichment step for the observation of subpicomole cross-linked products. Moreover this enrichment is highly selective as the MALDI-TOF analysis of the purified samples showed only few hydrophobic albumin tryptic peptides non-specifically adsorbed on the beads. It is important to notice that biotin/streptavidin affinity purification performed on intact complexes *prior* to trypsin digestion revealed unsuccessful. In this case, the recognition of the biotin tag by the streptavidin molecules is likely hindered when the CPPs (2.1 or 3.3 kDa) are associated to a bulky protein (42 kDa for actin and 66 kDa for albumin).

Supplementary Fig. 4 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

Manual bottom-up MALDI-TOF data analysis. For *XL* experiments of both CPP-actin and CPP-albumin interacting systems, a visual comparison of the MALDI-TOF spectra of

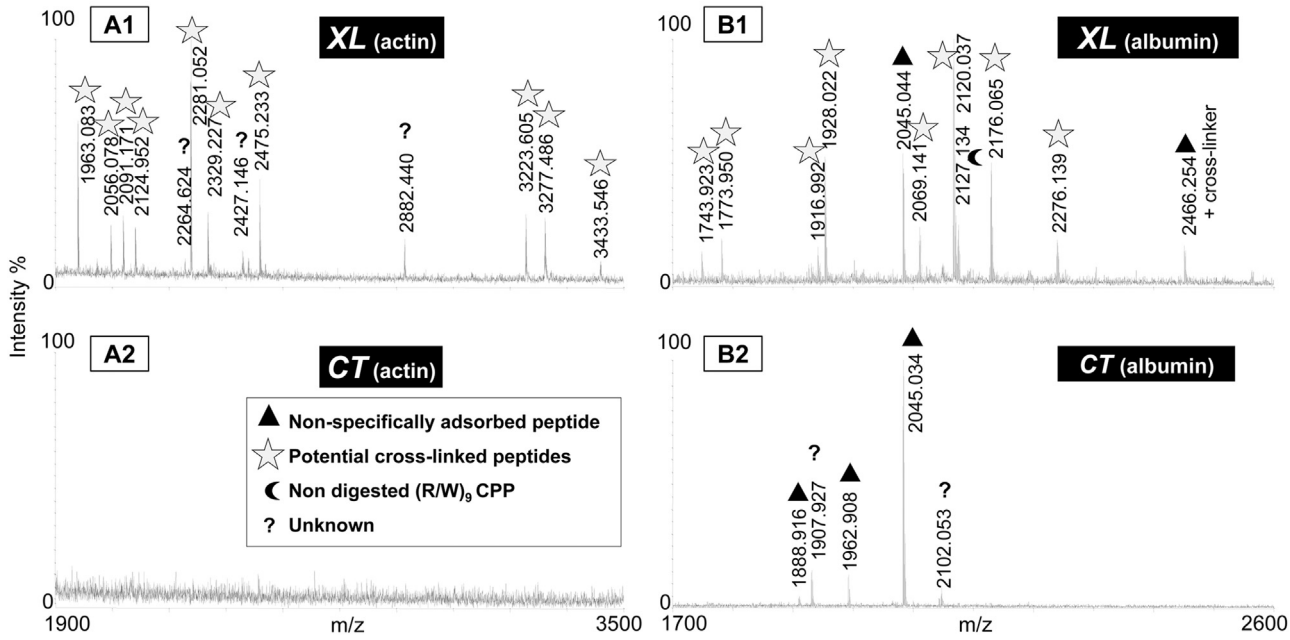

Fig. 3 – Positive ions reflector mode MALDI-TOF of reaction mixtures after biotin/streptavidin affinity purification for actin-$(R/W)_9$ and albumin-$(R/W)_9$ interacting systems. *XL* stands for cross-linking experiments with actin or albumin (A1, B1), *CT* are the corresponding controls without cross-linking agents (A2, B2).

XL and *CT* reactions, allowed to identify discriminating peaks appearing only in the *XL* reaction. Fig. 3 shows XL and CT MALDI spectra obtained for the $(R/W)_9$ CPP-actin or $(R/W)_9$ CPP-albumin systems. Same peaks were obtained with the $(R/W)_{16}$ CPP corresponding systems.

The manual confrontation of this list of discriminating peaks with the GPMAW list of predicted cross-links led to the identification of cross-linked peptides (Table 1). Sequences of cross-linked peptides are presented in Table S-1.

Supplementary Table S-1 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

As external calibration of each spectrum was realized on the nearest position, the cross-linked peptides were identified with a very satisfying mass accuracy (<20 ppm) providing a good confidence in the results. For the $(R/W)_9$ or $(R/W)_{16}$-actin interacting system a total of nine cross-linked peptides were reproducibly (five independent experiments) identified, which correspond to four different zones (six lysines) of the protein and for $(R/W)_9$ or $(R/W)_{16}$-albumin, eight cross-linked peptides were identified as five different zones (seven lysines) of the protein.

Specificity of the cross-linking reaction: To look at the specificity of the reaction, we achieved a competition experiment with both actin and albumin. Therefore, 10 μM of $(R/W)_9$ CPP were incubated with 10 μM of each protein and 2 mM of BS^3 cross-linker. The MALDI-TOF spectra of the purified digests showed peaks corresponding to cross-linked peptides with actin and with albumin. Even if the intensity of the peaks could not be directly related to their abundance in MS, the close relative intensity of the peaks corresponding to cross-linked peptides with both proteins, strongly suggested that association constants of $(R/W)_9$ CPP for actin and albumin were within the same range of values.

The specificity of the cross-linking reaction was also checked by incubating $(R/W)_9$ CPP (10 μM) with Promix1, a commercial mixture of three proteins: insulin (4 μM), cytochrome C (12 μM) and myoglobin (16 μM) in the presence of the K100 kit of cross-linkers. The comparison of the MALDI-TOF spectra (reflector positive ions mode) of the cross-linking (*XL*) and control (*CT*) purified digests revealed no discriminating peaks. This is in good agreement with the absence of formation of complexes between $(R/W)_9$ CPP and these proteins, that was observed with the intact complex analysis (Fig. S-5). Actually, these three proteins insulin (pI = 7.6), cytochrome C (pI = 9.6) and myoglobin (pI = 7.2) have either neutral or basic pI whereas actin (pI = 5.2) and albumin (pI = 5.8) both have acidic pI (UniprotKB database/ProtParam tool pI values). Therefore interactions between actin/albumin and the polycationic $(R/W)_9$ or $(R/W)_{16}$ CPPs are likely electrostatic, at least partly.

Supplementary Fig. 5 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

MALDI-TOF/TOF characterization of the cross-linked peptides. MALDI-TOF/TOF fragmentation of these ions gave mostly weak intensity and poor quality spectra. However in most spectra, identical series of peaks at *m/z* 259.1, 316.1, 373.1, 430.1 were observed corresponding to the N-terminal ions of the biotinylated tag added to the CPPs sequences (Biot(O_2), Biot(O_2)-G, Biot(O_2)-GG and Biot(O_2)-GGG respectively). This fragmentation highly competes with the formation of other fragments, which partly explains the poor quality of the high collision energy MS/MS spectra. However, the formation of these diagnostic ions constitutes a relevant marker of the presence of the biotinylated CPP in the cross-linked species.

The limited number of fragments obtained by MALDI-TOF/TOF did not allow characterization of the branched

Table 1 – Manual interpretation of cross-linked peptides between $(R/W)_9$ and actin (A) or albumin (B) using GPMAW software. Experimental *versus* simulated *m/z* values of identified cross-linked peptides between $(R/W)_9$ and actin or albumin obtained after trypsin digestion are presented in the two left columns. The delta mass is in the third column. Corresponding tryptic peptides of the protein and $(R/W)_9$ are in the two last columns. Sequences of the cross-linked peptides underlined were confirmed by MS/MS. Same results were obtained with $(R/W)_{16}$.

Exp. *m/z* of potential cross-linked peptides	Theo. *m/z* value (GPMAW)	Δ mass (ppm)	Potential tryptic peptides involved in the complex	Potential tryptic $(R/W)_9$ involved in the complex
A. Actin-$(R/W)_9$ interacting system				
1963.067	1963.079	6.1	322–332	1–7
2056.065	2056.086	10.2	202–210	1–7
2091.153	2091.174	10.0	322–333	1–7
2124.952	2124.95	0.9	46–57	1–6
2281.035	2281.051	7.0	46–57	1–7
2329.208	2329.225	7.3	311–323	1–7
2475.201	2475.24	15.7	308–321	1–7
3223.61	3223.653	13.3	91–111	1–7
3277.543	3277.503	12.2	35–56	1–7
3433.56	3433.605	13.1	35–57	1–7
B. Albumin-$(R/W)_9$ interacting system				
1743.916	1743.917	0.5	452–459	1–7
1773.943	1773.932	6.2	242–248	1–7
1916.992	1916.986	3.1	210–218	1–7
1928.016	1928.017	0.5	233–241	1–7
2069.158	2069.142	7.7	548–557	1–7
2120.031	2120.03	0.5	24–34	1–6
or	2120.03	0.5	25–34	1–7
2176.058	2176.049	4.1	35–44	1–7
2276.131	2276.131	0.0	24–34	1–7
or	2276.131	0.0	25–34	1–7

peptides. Moreover, manual identification of cross-linked peptide can be envisaged for simple and well defined interacting system (e.g. two known partners) but is not suitable for more complex mixtures. Therefore we developed an online and automated separation (nanoLC system), data acquisition (ESI-LTQ-Orbitrap) and data search (Xlink-Identifier) adapted to more complex systems involving a known biotinylated peptide or protein with one or several unknown partners. This automated procedure was validated with the same CPPs-protein (actin or albumin) interacting systems.

NanoLC–ESI-Orbitrap automatic data analysis. NanoLC–ESI-MS/MS analyses of the purified digests were performed in order to confirm cross-linked peptides sequences. We noticed that these analyses of low abundant peptides had to be done rapidly after affinity purification to avoid sample adsorption on the plastic tubes. If the nanoLC–MS/MS analysis was done within the 2 days following the purification step then the sensitivity of the LTQ-Orbitrap mass spectrometer was adequate to detect cross-linked peptides and to trigger MS/MS on their precursor ions.

MS/MS data treatment via Xlink-Identifier. The MS/MS spectra generated were searched using Xlink-Identifier [18]. Xlink-Identifier generates a list of cross-linked peptides for a given mass tolerance. For our experiments the mass tolerance on the precursor ions was set at 10 ppm. For each cross-linked peptides identified, Xlink-Identifier indicates the position of the cross-linking sites and provides matching scores including XlinkScore and the mass accuracy for the precursor ions selected (Fig. S-6A and B). A link is also available to visualize the corresponding MS/MS spectrum with identified fragments from both peptide chains (labeled a or b) and highlighted in different colors according to their type (b, y etc.).

Supplementary Fig. 6 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

NanoLC–ESI-MS/MS analysis of the cross-linked peptides allowed confirming the sequence of 7 out of the 10 cross-linked peptides previously identified for the CPPs-actin interacting systems in MALDI-TOF while all except one cross-linked peptides sequences were confirmed for the CPPs-albumin complex. Moreover three additional cross-linked peptides were found for the CPPs-albumin systems searching NanoLC–ESI-MS/MS data with Xlink-Identifier.

The identifications were mainly obtained using CID spectra and confirmed with HCD spectra (Fig. S-7). CID spectra were generally sufficient to validate the sequences.

Supplementary Fig. 7 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

All the cross-linked peptides identified were the result of a covalent bond formation between the CPPs and actin or albumin *via* the longer version of the cross-linker (SBAT or SBBT). The similar results obtained using the BS^3 cross-linker showed that the leaving group type did not seem to modify the number or the nature of the cross-linked peptides obtained.

The vast majority of the cross-linked species identified between the biotinylated CPPs and the protein partner (albumin or actin) contained the biotinylated CPP under the Biot(O_2)-G_4-K-R form (residues 1–7) since the covalent link on the lysine prevents the trypsin cleavage after the K residue.

It is worth noticing that the covalent attachment of the Biot(O_2)-G_4-K-R group could be considered as a

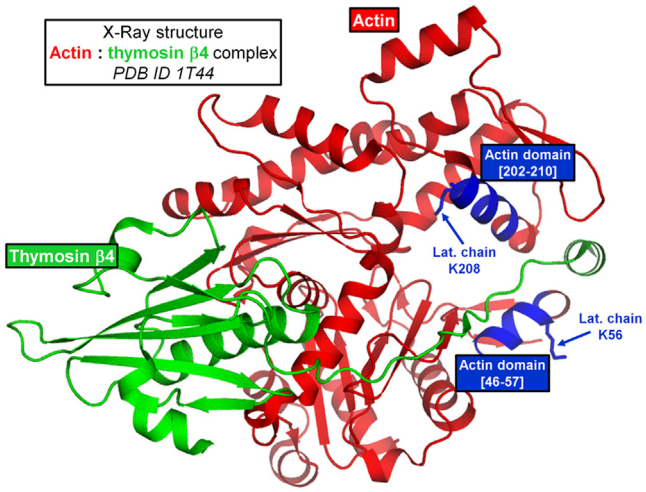

Fig. 4 – X-Ray structure of the complex actin: thymosin β4 (PDB ID) [22]. The actin protein is colored in red, thymosin β4 in green. Interaction zones characterized between the CPPs $(R/W)_9$ or $(R/W)_{16}$ and actin [202–210] from sub-domain 4 and [35–57] from sub-domain 2 are in blue (the region [40–50] is not visible on the crystal structure since it has not been crystallized). The lateral chain of the K residues K56 and K208 of actin, involved in the cross-linking are indicated.

post-translational modification of the peptides issued from the protein digestion and searched in this way with conventional search engine such as Mascot. However, contrary to Xlink-Identifier, with this type of search we can only have MS/MS information on the protein part and scoring of these cross-linked peptides is not properly made.

Data rationalization in a biological context. Cross-linking experiments were repeated several times for each interacting systems, giving reproducible results and leading to the identification of the same interaction zones for each system. For $(R/W)_9$ or $(R/W)_{16}$-actin systems four different interaction zones were found: [35–57] containing the cross-linked K45 and K56, [91–111] containing the cross-linked K108, [202–210] containing the cross-linked K208 and [308–333] containing the cross-linked K310, K321 and K323 (Fig. S-8). For $(R/W)_9$ or $(R/W)_{16}$-albumin interacting systems five different interaction zones were identified: [24–44], [210–218], [233–248], [452–459] and [548–557]. In addition, using the protein, actin or albumin, in a large excess (10:1 protein:CPP) gave the same results indicating that there would not be any preferential interaction site among the sites identified or any order in the colonization of these sites.

Supplementary Fig. 8 can be found, in the online version, at doi:10.1016/j.euprot.2014.03.002.

Using PyMOL free software program (DeLano Scientific LLC), it was possible to localize the identified interaction zones on the 3D structure (X-ray crystallography) of the protein for both systems. Demonstration of a direct interaction of both CPPs with the serum protein albumin is of interest for *in vivo* applications, since it might affect positively (protect from degradation) or negatively (sequestration of the CPP) the cell delivery of conjugated cargoes into cells [4]. This interaction could involve electrostatic, hydrophobic or π/cation interactions, which could explain the dispersion and multiplicity of the interaction zones identified in albumin.

In the case of actin, three out of the four interaction zones containing the cross-linked lysines K45/K56, K108, K310/K321/K323, are located on the outer surface of the protein and are accessible to the CPPs (Fig. S-8). By contrast, the cross-linked lysine K208 is located in the close proximity of the nucleotide binding cleft and is less exposed to the solvent. It was previously reported that both $(R/W)_{16}$ and $(R/W)_9$ interact with G-actin [3]. We identified in this study identical domains of interaction for the two CPPs in actin. $(R/W)_{16}$ was shown previously to compete with the G-actin sequestering protein thymosin-ß4 for binding to actin [3]. Interestingly, among the cross-linked regions identified in this study, the domains [35–57] (K45 and K56) and [202–210] (K208) are localized within the binding domain of thymosin-ß4 [22] (Fig. 4), a spatial proximity that could explain the previous results of competition experiments. Actin polymerization/depolymerization dynamics is a complex process that involves numerous actin binding proteins (capping, nucleation, elongation, severing or bundle proteins). Kang and co-workers recently identified two discrete cation-binding sites within F-actin that they described as “polymerization” (driving actin filament assembly) and “stiffness” (modulating filament bending rigidity) sites [23]. Knowing that $(R/W)_9$ and $(R/W)_{16}$ CPPs can induce formation of stress fibers in cells that genetically lack F-actin formation, it is of particular interest to find herein that these CPPs bind G-actin within these two cation-binding regions that have been described to modulate actin dynamics in cells.

4. Conclusions

We set up and optimized a robust and efficient cross-linking-MS workflow allowing a complete characterization of *in vitro* CPP-protein interacting systems. This workflow originally designed for the study of CPP interaction partners can be applied to various interacting systems composed of two or more potential protein partners as soon as one of the two partners is known and can be labeled for enrichment purpose (*e.g.*: biotin tag). While the intact complex analysis gives reliable information on the cross-linking reaction yield and the cross-linked complexes composition according to the reaction conditions, the bottom-up analysis allows identification and characterization of the interaction zones.

Our study based on $(R/W)_9$ and $(R/W)_{16}$ CPPs illustrates the necessity of the use of cross-linking-MS for the *in vitro* study of interaction partners, and its complementarity with analytical techniques such as ITC and NMR.

Other CPPs can now be analyzed using the same strategy, in particular Tat that was shown to remodel actin cytoskeleton in actin-encapsulated giant vesicles [24]. Biologically speaking, we can anticipate that many, if not all, CPPs are not so inert and that similar cross-linking-MS strategies will be helpful to characterize membrane and intracellular partners for these peptides. Mapping the domains of interaction by cross-linking approaches will also help understanding, at least partly, the biological activity of CPPs.

Conflict of interest

There is no conflict of interest.

Transparency document

The Transparency document associated with this article can be found in the online version.

REFERENCES

[1] Bechara C, Sagan S. Cell-penetrating peptides: 20 years later, where do we stand? FEBS Lett 2013;587:1693–702.
[2] Howl J, Matou-Nasri S, West DC, Farquhar M, Slaninová J, Ostenson C-G, et al. Bioportide: an emergent concept of bioactive cell-penetrating peptides. Cell Mol Life Sci 2012;69:2951–66.
[3] Delaroche D, Cantrelle F-X, Subra F, Van Heijenoort C, Guittet E, Jiao C-Y, et al. Cell-penetrating peptides with intracellular actin-remodeling activity in malignant fibroblasts. J Biol Chem 2010;285:7712–21.
[4] Kosuge M, Takeuchi T, Nakase I, Jones AT, Futaki S. Cellular internalization and distribution of arginine-rich peptides as a function of extracellular peptide concentration, serum, and plasma membrane associated proteoglycans. Bioconjug Chem 2008;19:656–64.
[5] Mädler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. J Mass Spectrom 2009;44:694–706.
[6] Chowdhury SM, Du X, Tolić N, Wu S, Moore RJ, Mayer MU, et al. Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. Anal Chem 2009;81:5524–32.
[7] Sohn CH, Agnew HD, Lee JE, Sweredoski MJ, Graham RLJ, Smith GT, et al. Designer reagents for mass spectrometry-based proteomics: clickable cross-linkers for elucidation of protein structures and interactions. Anal Chem 2012;84:2662–9.
[8] Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein–protein interactions. Anal Chem 2005;77:311–8.
[9] Müller MQ, Zeiser JJ, Dreiocker F, Pich A, Schäfer M, Sinz A. A universal matrix-assisted laser desorption/ionization cleavable cross-linker for protein structure analysis. Rapid Commun Mass Spectrom 2011;25:155–61.
[10] Ihling C, Schmidt A, Kalkhof S, Schulz DM, Stingl C, Mechtler K, et al. Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex. J Am Soc Mass Spectrom 2006;17:1100–13.
[11] Paramelle D, Cantel S, Enjalbal C, Amblard M, Forest E, Heymann M, et al. A new generation of cross-linkers for selective detection by MALDI MS. Proteomics 2009;9:5384–8.
[12] Sachon E, Bolbach G, Chassaing G, Lavielle S, Sagan S. Cgamma H2 of Met174 side chain is the site of covalent attachment of a substance P analog photoactivable in position 5. J Biol Chem 2002;277:50409–14.
[13] Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, et al. Identification of cross-linked peptides from large sequence databases. Nat Methods 2008;5:315–8.
[14] Fritzsche R, Ihling CH, Götze M, Sinz A. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. Rapid Commun Mass Spectrom 2012;26:653–8.
[15] Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. Mol Cell Proteomics 2010;9:1634–49.
[16] Söderberg CAG, Lambert W, Kjellström S, Wiegandt A, Wulff RP, Månsson C, et al. Detection of crosslinks within and between proteins by LC–MALDI-TOFTOF and the software FINDX to reduce the MSMS-data to acquire for validation. PLoS ONE 2012;7:e38927.
[17] Rasmussen MI, Refsgaard JC, Peng L, Houen G, Højrup P. CrossWork: software-assisted identification of cross-linked peptides. J Proteomics 2011;74:1871–83.
[18] Du X, Chowdhury SM, Manes NP, Wu S, Mayer MU, Adkins JN, et al. Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. J Proteome Res 2011;10:923–31.
[19] Nielsen T, Thaysen-Andersen M, Larsen N, Jørgensen FS, Houen G, Højrup P. Determination of protein conformation by isotopically labelled cross-linking and dedicated software: application to the chaperone, calreticulin. Int J Mass Spectrom 2007;268:217–26.
[20] Nazabal A, Wenzel RJ, Zenobi R. Immunoassays with direct mass spectrometric detection. Anal Chem 2006;78:3562–70.
[21] Sachon E, Clodic G, Blasco T, Bolbach G. Protein desolvation in UV matrix-assisted laser desorption/ionization (MALDI). J Am Soc Mass Spectrom 2007;18:1880–90.
[22] Irobi E, Aguda AH, Larsson M, Guerin C, Yin HL, Burtnick LD, et al. Structural basis of actin sequestration by thymosin-?4: implications for WH2 proteins. EMBO J 2004;23:3599–608.
[23] Kang H, Bradley MJ, McCullough BR, Pierre A, Grintsevich EE, Reisler E, et al. Identification of cation-binding sites on actin that drive polymerization and modulate bending stiffness. Proc Natl Acad Sci USA 2012;109:16923–7.
[24] Mishra A, Lai GH, Schmidt NW, Sun VZ, Rodriguez AR, Tong R, et al. Translocation of HIV TAT peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. Proc Natl Acad Sci 2011;108:16883–8.