INSTITUTE FOR Systems Biology

Revolutionizing science. Enhancing life.

Three Integrative Computational Tools for a Chemical Protein Crosslinking Workflow using Isotope-Coded Crosslinkers and Mass Spectrometry: iXLINK, doXLINK and XlinkViewer

Jan Seebacher[1,2]; Parag Mallick[3]; Ning Zhang[1]; James S. Eddes[1]; Ruedi Aebersold[1,4]; Michael H. Gelb[2]

[1] Institute for Systems Biology, Seattle, WA; [2] University of Washington, Seattle, WA; [3] Louis Warschaw Prostate Cancer Center, Cedars-Sinai, Los Angeles, CA; [4] Federal Institute of Technology, Zürich, Switzerland

Overview:

We present here an integrative and high-throughput approach of using
- isotopically labeled crosslinkers
- MALDI-TOF/TOF tandem mass spectrometry
- a suite of new software tools: iXLINK, doXLINK and XLinkViewer

to identify proximally located crosslinks between protein segments in single subunit, multi-subunit proteins, and protein complexes.

The results obtained for crosslinking reaction between the Colicin E7 DNAse/Im7 heterodimer[1] and the crosslinkers DSS and DSG were compared to the published crystal structure.

Introduction:

Amino acid distance constraints for native proteins and protein complexes provide invaluable information for protein structure prediction. Chemical crosslinking is a powerful method for generating such distance constraints at medium resolution. However, the analysis of mass spectrometry (MS) data derived from crosslinked multimeric protein complexes is time-consuming and frequently a subjective procedure. The use of stable isotope labeling can be beneficial to such MS analyses, especially for the automatic detection of multiple types of linked species[2] (i.e. monolinks (Type 1), loplinks (Type 1) and crosslinks (Type 2), but requires appropriate software tools for data analysis. In this work, we describe a protocol using isotopically labeled crosslinking reagents in $[^{16/18}O]$water, LC-MALDI-tof/tof MS, all integrated with new computational MS-data analysis tools.

Methods:

Chemical crosslinking of the two subunits of the Colicin E7 DNAse/Im7 heterodimer with d_0/d_{12}-disuccinimidyl suberate (DSS) or d_0/d_6-disuccinimidyl glutarate (DSG) – both crosslinkers are reactive towards amino groups, i.e. lysines and amino termini- in phosphate buffer containing either $[^{18}O]$ and/or $[^{16}O]$ water were monitored by SDS PAGE. After completion, combined mixtures of isotopically light/heavy crosslinker-modified protein complex were digested, first with agarose-trypsin (Pierce), and then with Asp-N in solution. Digestion products were separated by nano-liquid chromatography (nano-LC, LC-Packings, water/acetonitrile gradient, flow rate 1.5 ul/min, inner capillary diameter 100 um). LC fractions were collected onto MALDI target plates and analyzed using a MALDI-tof/tof mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems). Integrated in-house-developed software tools (iXLINK, doXLINK, and XlinkViewer) were used to analyze MS and MS/MS data.

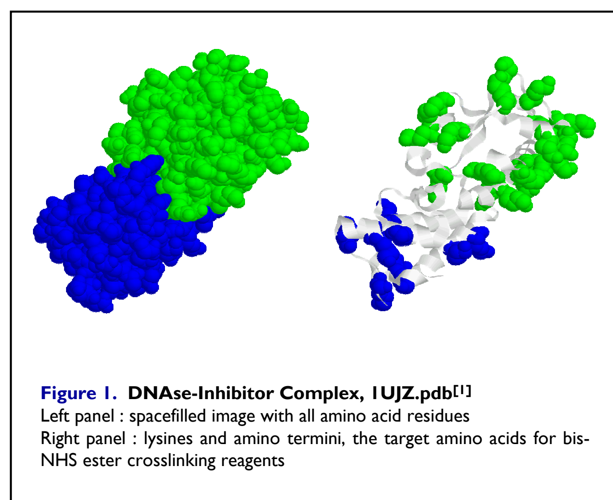

Figure 1. DNAse-Inhibitor Complex, 1UJZ.pdb[1]
Left panel : spacefilled image with all amino acid residues
Right panel : lysines and amino termini, the target amino acids for bis-NHS ester crosslinking reagents

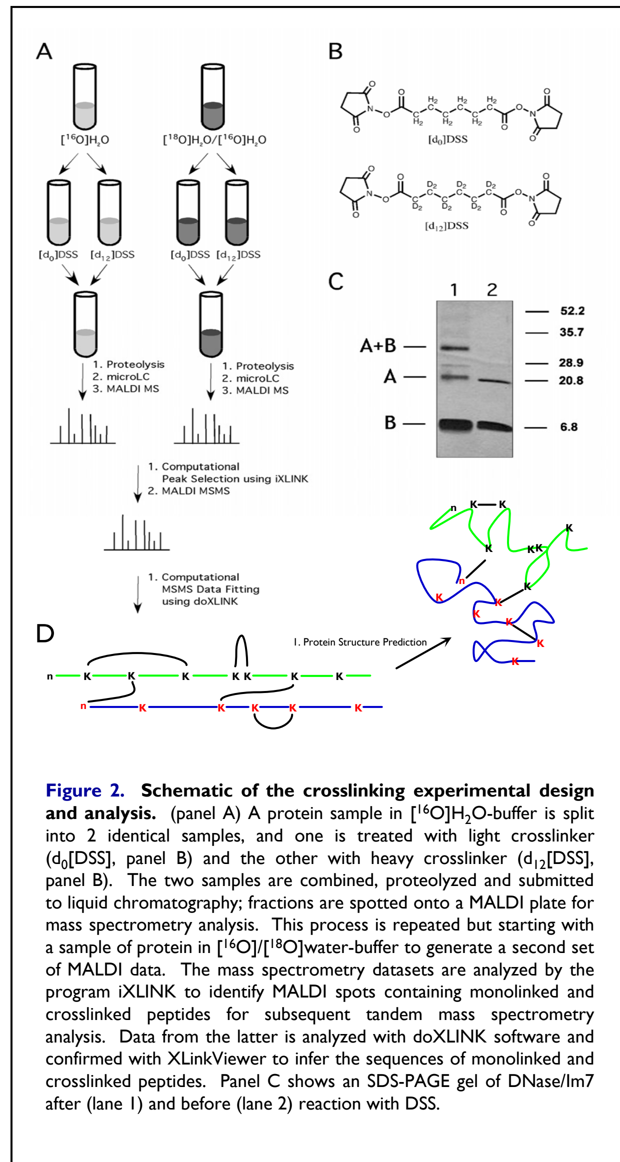

Figure 2. Schematic of the crosslinking experimental design and analysis. (panel A) A protein sample in $[^{16}O]H_2O$-buffer is split into 2 identical samples, and one is treated with light crosslinker (d_0[DSS], panel B) and the other with heavy crosslinker (d_{12}[DSS], panel B). The two samples are combined, proteolyzed and submitted to liquid chromatography; fractions are spotted onto a MALDI plate for mass spectrometry analysis. This process is repeated but starting with a sample of protein in $[^{16}O]/[^{18}O]$water-buffer to generate a second set of MALDI data. The mass spectrometry datasets are analyzed by the program iXLINK to identify MALDI spots containing monolinked and crosslinked peptides for subsequent tandem mass spectrometry analysis. Data from the latter is analyzed with doXLINK software and confirmed with XLinkViewer to infer the sequences of monolinked and crosslinked peptides. Panel C shows an SDS-PAGE gel of DNase/Im7 after (lane 1) and before (lane 2) reaction with DSS.

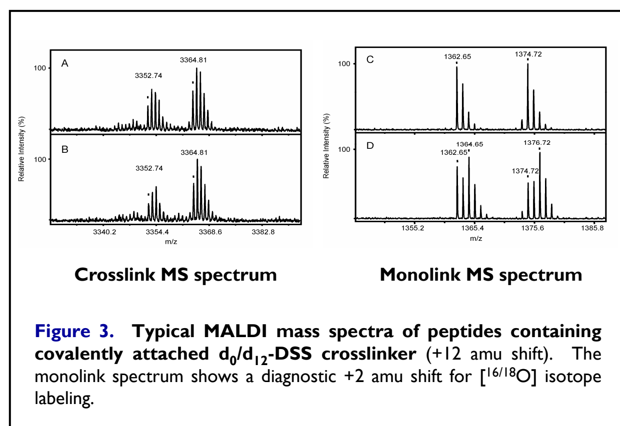

Figure 3. Typical MALDI mass spectra of peptides containing covalently attached d_0/d_{12}-DSS crosslinker (+12 amu shift). The monolink spectrum shows a diagnostic +2 amu shift for $[^{16/18}O]$ isotope labeling.

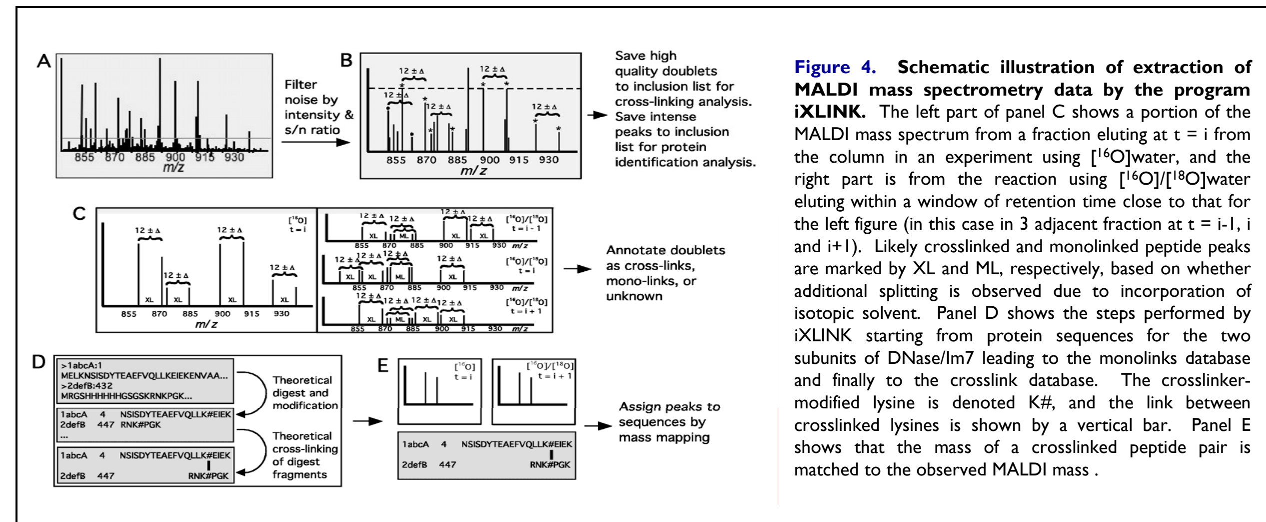

Figure 4. Schematic illustration of extraction of MALDI mass spectrometry data by the program iXLINK. The left part of panel C shows a portion of the MALDI mass spectrum from a fraction eluting at t = i from the column in an experiment using $[^{16}O]$water, and the right part is from the reaction using $[^{16}O]/[^{18}O]$water eluting within a window of retention time close to that for the left figure (in this case in 3 adjacent fraction at t = i-1, i and i+1). Likely crosslinked and monolinked peptide peaks are marked by XL and ML, respectively, based on whether additional splitting is observed due to incorporation of isotopic solvent. Panel D shows the steps performed by iXLINK starting from protein sequences for the two subunits of DNase/Im7 leading to the monolinks database and finally to the crosslink database. The crosslinker-modified lysine is denoted K#, and the link between crosslinked lysines is shown by a vertical bar. Panel E shows that the mass of a crosslinked peptide pair is matched to the observed MALDI mass .

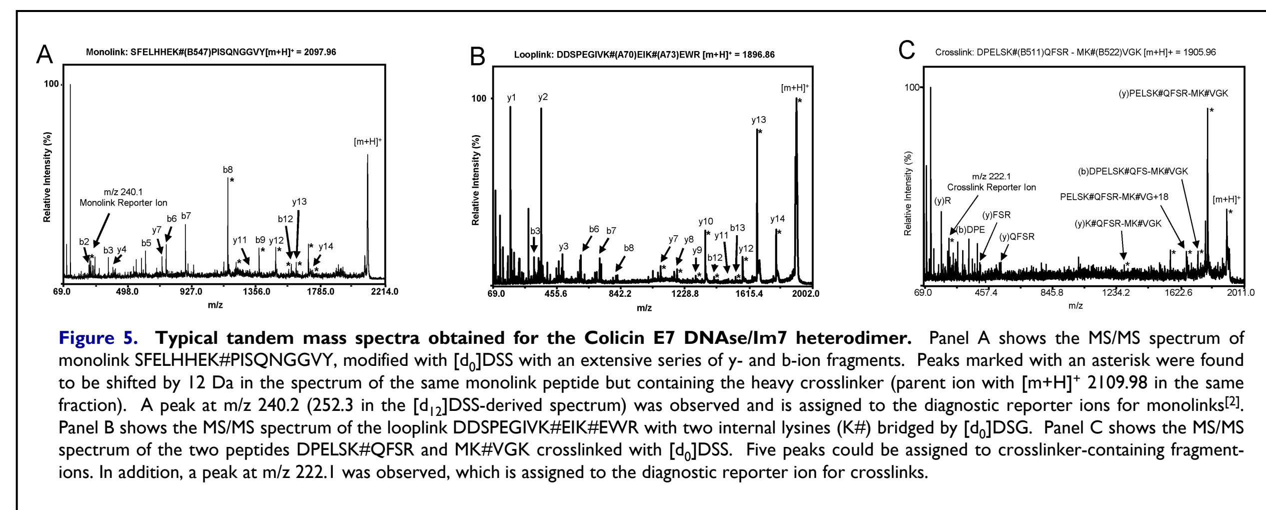

Figure 5. Typical tandem mass spectra obtained for the Colicin E7 DNAse/Im7 heterodimer. Panel A shows the MS/MS spectrum of monolink SFELHHEK#PISQNGGVY, modified with $[d_0]$DSS with an extensive series of y- and b-ion fragments. Peaks marked with an asterisk were found to be shifted by 12 Da in the spectrum of the same monolink peptide but containing the heavy crosslinker (parent ion with $[m+H]^+$ 2109.98 in the same fraction). A peak at m/z 240.2 (252.3 in the $[d_{12}]$DSS-derived spectrum) was observed and is assigned to the diagnostic reporter ions for monolinks[2]. Panel B shows the MS/MS spectrum of the looplink DDSPEGIVK#EIK#EWR with two internal lysines (K#) bridged by $[d_0]$DSG. Panel C shows the MS/MS spectrum of the two peptides DPELSK#QFSR and MK#VGK crosslinked with $[d_0]$DSS. Five peaks could be assigned to crosslinker-containing fragment-ions. In addition, a peak at m/z 222.1 was observed, which is assigned to the diagnostic reporter ion for crosslinks.

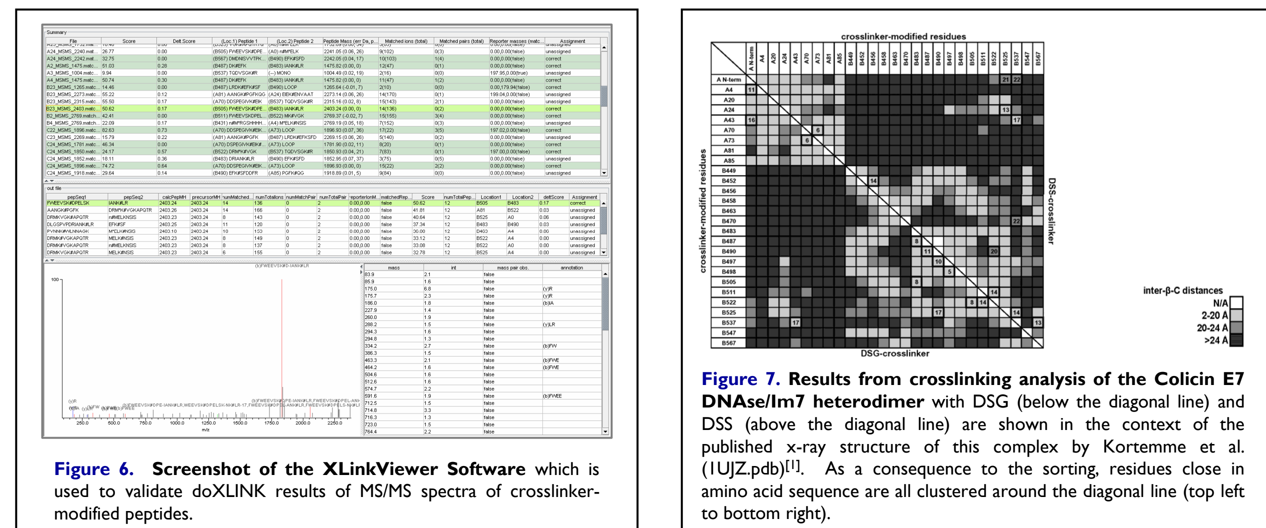

Figure 6. Screenshot of the XLinkViewer Software which is used to validate doXLINK results of MS/MS spectra of crosslinker-modified peptides.

Figure 7. Results from crosslinking analysis of the Colicin E7 DNAse/Im7 heterodimer with DSG (below the diagonal line) and DSS (above the diagonal line) are shown in the context of the published x-ray structure of this complex by Kortemme et al. (1UJZ.pdb)[1]. As a consequence to the sorting, residues close in amino acid sequence are all clustered around the diagonal line (top left to bottom right).

Results:

Using isotopically labeled versions of the bis-NHS esters DSS and DSG in reaction buffer containing $[^{18}O]$water, we were able to detect crosslinked protein subunits of the Colicin E7 DNAse/Im7 heterodimer by SDS PAGE with an estimated 30% yield. After digestion with Asp-N and/or trypsin, the reaction products were subjected to nano-LC-MALDI tandem MS (MS/MS). The new software tools for MS data analysis are designed to be used in an integrated workflow for crosslinking multiple proteins with isotopically labeled bis-NHS esters in buffer containing heavy and/or normal water, allowing the use of multiple proteases and common peptide modifications. The program iXLINK easily processes MS-MALDI data, extracting mass pairs for potentially crosslinker-modified peptide species, including peptide mass mapping to a peptide candidate database. Due to the diagnostic quadruplet mass pattern of peptides modified with hydrolyzed crosslinker (monolinks), when $[^{18}O]$water was used during the crosslinking reaction, iXLINK easily manages to distinguish mass signals generated from monolinks and from crosslinked petides. An inclusion list generated by iXLINK, containing LC-fraction and precursor mass of potentially crosslinker-modified peptides, is then submitted to tandem MS acquisition. MS/MS data analysis can automatically be performed using the program doXLINK, and the preliminary results can easily be user-evaluated with the XlinkViewer program. 23 crosslinked lysine-lysine or lysine-N-terminus pairs were identified for the crosslinked Colicin E7 DNAse/Im7 protein complex with DSS and DSG. All were found to be close in space when examined in the context of the known x-ray structure of the heterodimer. The new software will be made publicly available after publication including a detailed user manual.

Conclusions

- 23 amino acid residues (lysines and N-termini) of the Colicin E7 DNAse/Im7 protein complex were found to be crosslinked with DSS, and DSG respectively
- they are located within <22.1 Å.
- This suggests that no major distortion of the native protein structure, or "random" crosslinking was observed.
- DSG and DSS crosslinking showed overlapping results but the observed discrepancy between these results suggest the use of multiple crosslinkers in comparative experiments.
- The experimental and analytical workflow can be automated and allow the use of a variety of bis-NHS ester crosslinking reagents, including the commercially available d_0/d_4-isotopically labeled reagents BS2 and BS3 (both from Pierce).
- Therefore, the results shown in this study illustrate the potential of using protein crosslinking together with isotope labeling, and MALDI mass spectrometry for high-throughput protein structure analysis.

Acknowledgement

This work was supported with federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health under contract No. N01-HV-28179. Special thanks to Richard Bonneau, Nichole King, Patrick Pedrioli and Brian Smart for technical help with this study.

References

[1] Kortemme, T.; Joachimiak, L. A.; Bullock, A. N.; Schuler, A. D.; Stoddard, B. L.; Baker, D. Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **2004**, *11*, 371-379.

[2] Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834-850.