Bioinformatics, 31(12), 2015, 2043–2045
doi: 10.1093/bioinformatics/btv083
Advance Access Publication Date: 11 February 2015
Applications Note

OXFORD

Structural bioinformatics

Crossfinder-assisted mapping of protein crosslinks formed by site-specifically incorporated crosslinkers

Felix Mueller-Planitz

BioMedical Center, Ludwig-Maximilians-University, 80336 Munich, Germany

Associate Editor: Anna Tramontano

Received on November 24, 2014; revised on January 29, 2015; accepted on February 5, 2015

Abstract

Protein crosslinking has been used for decades to derive structural information about proteins and protein complexes. Only recently, however, it became possible to map the amino acids involved in the crosslinks with the advent of high resolution mass spectrometry (MS). Here, we present Crossfinder, which automates the search for crosslinks formed by site-specifically incorporated crosslinking amino acids in LC-MS-MS data.

Availability and Implementation: An executable version of Crossfinder for Windows machines (64-bit) is freely available to non-commercial users. It is bundled with a manual and example data.

Contact: felix.mueller-planitz@med.uni-muenchen.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Introduction

For decades, protein crosslinking has been a widely used technique to obtain information about the structural architecture of proteins and protein complexes. The full potential of the technique however was achieved only recently when it became technically feasible to map the precise amino acids that are involved in the crosslink by high accuracy mass spectrometry (MS) and dedicated bioinformatic tools that mine the MS datasets. Applications include probing protein conformations, mapping interaction surfaces between proteins and elucidating the structural architectures of protein complexes (Chen *et al.*, 2010; Forne *et al.*, 2012; Herzog *et al.*, 2012; Sinz and Wang, 2001).

Crosslinking can be induced for example by adding a chemical with two reactive groups to a protein solution. The chemical will then crosslink accessible amino acids. Several software solutions exist to map these kinds of crosslinks from MS data.

Alternatively, the crosslinker can be site-specifically installed into the protein, e.g. during chemical synthesis or by using genetically encoded crosslinking amino acids (Chin *et al.*, 2002). Site-specifically attached crosslinkers have several advantages (see Forne *et al.*, 2012, for discussion). Among them is the possibility to target the crosslinker to specific regions of the protein. Through this targeted approach, we could, e.g. validate a crystal interface in solution (Gazda *et al.*, 2013). Dedicated software tools to map crosslinks formed by site-specifically installed crosslinkers are scarce, unfortunately (Forne *et al.*, 2012; Gotze *et al.*, 2012).

To map such crosslinks, we previously developed a collection of Matlab scripts (Forne *et al.*, 2012). Programming skills were needed to use the scripts, however, limiting their practicality. Moreover, the scripts lacked a convenient workflow, had no intuitive user interface, and it was not possible to easily visualize and browse through the results. Collectively, these issues limited wide-spread application of crosslinking approaches in which the crosslinking moiety is site-specifically incorporated into the protein. We introduce here an easy-to-use executable version of Crossfinder that intuitively integrates all steps of the analysis.

Description of crossfinder

Crossfinder's scope is to map protein crosslinks formed between site-specifically incorporated crosslinking moieties and other, unknown amino acids in the same or a separate protein with amino acid resolution. Any crosslinking moiety can be used as long as its mass is known, and it forms crosslinks that are stable under MS conditions. The amino acid sequences of the protein(s) that potentially

take part in the crosslinking reaction must be known before running Crossfinder. These sequences can be readily determined by conventional MS search engines. The number of proteins Crossfinder can handle is only limited by the memory of the computer. Hundreds of proteins pose no problem on conventional PCs.

Crossfinder accepts LC-MS-MS data in three commonly used data formats: mzXML, MGF and DTA. Note that the charges of precursor ions must have been measured. It is advisable to exclude precursors with lower charge states during data collection as they mostly come from linear, not from crosslinked peptides (Rinner *et al.*, 2008).

All information necessary to run Crossfinder is compiled in a Microsoft Excel template file, which is directly opened by Crossfinder. The Excel template allows for a convenient overview over the settings.

A valuable feature of Crossfinder is its ability to perform the crosslinking analysis on multiple datasets at once, e.g. on samples and negative controls. Each dataset can consist of any number of data files. For instance, it may be practical to combine files containing replicates into a single dataset for the purpose of the analysis. Combining files into datasets drastically reduces the length and complexity of the output and thus significantly speeds up the downstream analysis. At the end of the analysis, the results obtained for individual datasets can also be filtered against one another, further simplifying the complexity of the output (see below).

Based on protein sequences, protease information, variable and constant modifications of amino acids and the mass and position of the crosslink-inducing amino acid, Crossfinder compiles a list of all theoretically possible peptides (Fig. 1). Any of these peptides may in principle be part of a crosslink. Crossfinder simulates crosslinking *in silico* by combining the peptides bearing the crosslinking moiety with all other possible peptides. Side reactions introduced by the crosslinking moiety, such as intra-peptide crosslinking or—in the case of benzophenones—abstraction of two hydrogen atoms from the target peptide (Dorman and Prestwich, 1994), are similarly simulated. Each theoretically possible peptide, crosslinked or not, is a candidate that Crossfinder will try to identify in the experimental data (Fig. 1).

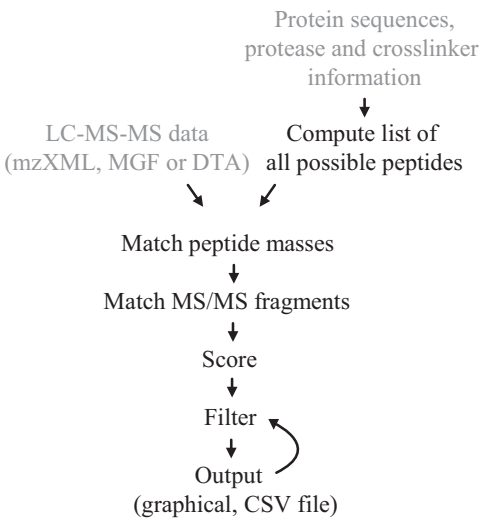

Fig. 1. Crossfinder's strategy to map crosslinks. See text for details. Black: steps performed by Crossfinder. Gray: input information. CSV: comma separated values

Depending on the specificity of the crosslinking moiety, multiple amino acids within one peptide may be the target of the crosslinking reaction. Each possible position is considered, creating a family of isobaric crosslink candidates. Crossfinder later determines which of these candidates best explains the data to obtain the attachment site with amino acid resolution.

Crossfinder assigns candidates to spectra by letting all candidates, including uncrosslinked ones, compete for any given spectrum. If multiple candidates can be assigned to one spectrum, a calculated score is used to rank the matches (Forne *et al.*, 2012). A cutoff relative to the score of the best match can be freely chosen to display only a fraction of the possible matches and reduce the complexity of the output (Supplementary Fig. S1A). Internal ions, i.e. ions that are produced by double fragmentation of the peptide backbone, provide additional support for correct identification (Supplementary Fig. S1B) and can help to distinguish between candidates with ambiguous scores.

Precursor and fragment masses are matched within a specified experimental error. Precise error limits for both MS dimensions may not be known the first time the data are analyzed, however. Crossfinder can help the user to determine the error limits first. Errors can be specified in ppm or absolute masses. We note that crosslinks can be assigned with a much higher confidence if actual errors are in the ppm range in both MS dimensions. Indeed, Crossfinder is optimized to work with high accuracy MS-MS data.

The results of the analysis are saved as text files that can be directly imported into Excel. In addition, the user can conveniently browse through the annotated spectra in the graphical user interface (Supplementary Fig S1A).

Crossfinder provides the opportunity to re-filter the data after the analysis. For example, a dataset can be filtered against another one by retaining only those candidates that are present in or absent from the second dataset. The data can also be filtered for candidates that could explain at least a certain fraction of the total ion intensity in the MS-MS spectrum, that achieved a certain minimal score or that contained a sufficient number of matched fragment ions of various sorts (Supplementary Fig. S1A). More detailed explanations of the filters are provided in the manual. The re-filtered data can be saved or exported as a comma separated text file that can be directly opened by Excel.

Supplementary Figure 1 shows an example analysis of previously published and additional unpublished data. As expected, all previously identified crosslinks could be readily mapped.

In conclusion, Crossfinder streamlines the identification of crosslinks formed by site-specifically incorporated crosslinkers. It offers an intuitive interface, allows for sophisticated filtering, displays the data in a compact and organized fashion and readily exports the results to Excel for further downstream analyses.

Acknowledgements

The author would like to thank Ignasi Forné and Nadine Harrer for numerous discussions, comments on the manuscript and testing of the software.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [MU3613/1-1, MU3613/3-1 and SFB 1064 TP A7].

Conflict of Interest: none declared.

References

Chambers,M.C. *et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, **30**, 918–920.

Chen,Z.A. *et al.* (2010) Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.*, **29**, 717–726.

Chin,J.W. *et al.* (2002) Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc. Natl. Acad. Sci. USA*, **99**, 11020–11024.

Dorman,G. and Prestwich,G.D. (1994) Benzophenone photophores in biochemistry, *Biochemistry*, **33**, 5661–5673.

Forne,I. *et al.* (2012) Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. *Mol Cell Proteomics*, **11**, M111 012088.

Gazda,L. *et al.* (2013) The myosin chaperone UNC-45 is organized in tandem modules to support myofilament formation in C. *elegans*. Cell, **152**, 183–195.

Gotze,M. *et al.* (2012) StavroX–a software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom*, **23**, 76–87.

Herzog,F. *et al.* (2012) Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science*, **337**, 1348–1352.

Rinner,O. *et al.* (2008) Identification of cross-linked peptides from large sequence databases. *Nat. Methods*, **5**, 315–318.

Sinz,A. and Wang,K. (2001) Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin-calmodulin complexes. *Biochemistry*, **40**, 7903–7913.