Journal of proteome research

pubs.acs.org/jpr

Technical Note

LinX: A Software Tool for Uncommon Cross-Linking Chemistry

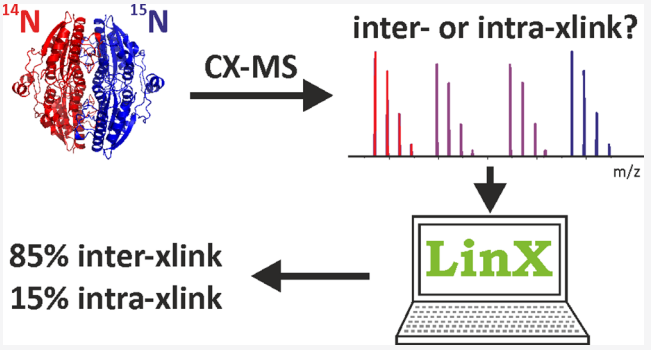
Zdeněk Kukačka,[‖] Michal Rosůlek,[‖] Jan Jelínek, Lukáš Slavata, Daniel Kavan, and Petr Novák*

Cite This: *J. Proteome Res.* 2021, 20, 2021−2027

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chemical cross-linking mass spectrometry has become a popular tool in structural biology. Although several algorithms exist that efficiently analyze data-dependent mass spectrometric data, the algorithm to identify and quantify intermolecular cross-links located at the interaction interface of homodimer molecules was missing. The algorithm in LinX utilizes high mass accuracy for ion identification. In contrast with standard data-dependent analysis, LinX enables the elucidation of cross-linked peptides originating from the interaction interface of homodimers labeled by $^{14}N/^{15}N$, including their ratio or cross-links from protein−nucleic acid complexes. The software is written in Java language, and its source code and a detailed user's guide are freely available at https://github.com/KukackaZ/LinX or https://ms-utils.org/LinX. Data are accessible via the ProteomeXchange server with the data set identifier PXD023522.



KEYWORDS: *mass spectrometry, high resolution, chemical cross-linking, proteins, homo oligomers, nucleic acids, data interpretation*

INTRODUCTION

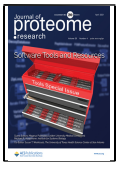
Chemical cross-linking mass spectrometry (CXMS) has developed into a powerful method to map low-resolution protein structures and to characterize molecular interfaces in protein−protein and protein−nucleic acid complexes.[1−5] Despite the excellence of CXMS as an analytical tool, the analysis of highly complex cross-linking data is a demanding task. Therefore, during the last two decades, several strategies to analyze cross-linked products have been reported. These available tools were mainly implemented in the following software tools: the Automated Spectrum Assignment Program (ASAP),[1] General Protein/Mass Analysis for Windows (GPMAW),[6] Kojak,[7] MS2links,[8] pLink2,[9] VIRTUALMSLAB,[10] xQuest,[11] X-Links,[12] X-Link Identifier,[13] xComb,[14] MS-Bridge, StavroX,[15] and, most recently, MeroX[16] and Mango[17] for MS/MS-cleavable cross-linkers. The detailed list of cross-linking software including their applicability and requirements was published in 2014.[18]

However, none of the software mentioned focuses directly on the challenging tasks of chemical cross-linking, the identification and quantitation of intermolecular cross-links originating from homodimer interactions. As was already published at the beginning of the century on interleukin-6,[19] the intermolecular cross-links can be elucidated using mixed-isotope cross-linking (MIX), where unlabeled (^{14}N) protein is mixed with ^{15}N-labeled protein in a 1:1 ratio before the cross-linking reaction. The cross-linked homodimer complex is subjected to proteolysis, and the resulting peptides are analyzed by mass spectrometry. Molecular ions from intermolecular cross-linked peptides are theoretically manifested in spectra as triplet/quadruplet MS peaks ($^{14}N/^{14}N$, $^{14}N/^{15}N$, $^{15}N/^{14}N$, and $^{15}N/^{15}N$) (Figure 1A), whereas molecular ions from intramolecular cross-linked peptides exclusively contain ^{14}N or ^{15}N and yield doublet peaks (Figure 1B). Unfortunately, in real cross-linking experiments, isotopic patterns of intra- and intermolecular cross-links often overlap, and two populations of restraints are formed (Figure 1C,D); thus relying only on MS/MS information is not sufficient to find out what population of cross-links is more dominant in its mixture. The cross-linked ion intensities must be taken into account.

In this study, we report a new software called LinX, whose algorithm enables the evaluation of CXMS data from MIX, including calculations of ratios describing the occurrence of intermolecular cross-links in its mixture with intramolecular cross-links as well as an evaluation of data from protein−nucleic acid cross-linking. LinX is a Java language-based tool designed for the rapid assignment, evaluation, and validation of mass-spectrometric data sets. LinX has been developed primarily for accurate high-resolution mass spectrometry data (MS^1); however, it offers, in principle, analysis of any other deconvoluted MS data (MS^n) in case MS^1 data are separately

Special Issue: Software Tools and Resources 2021

Received: October 29, 2020
Published: March 4, 2021



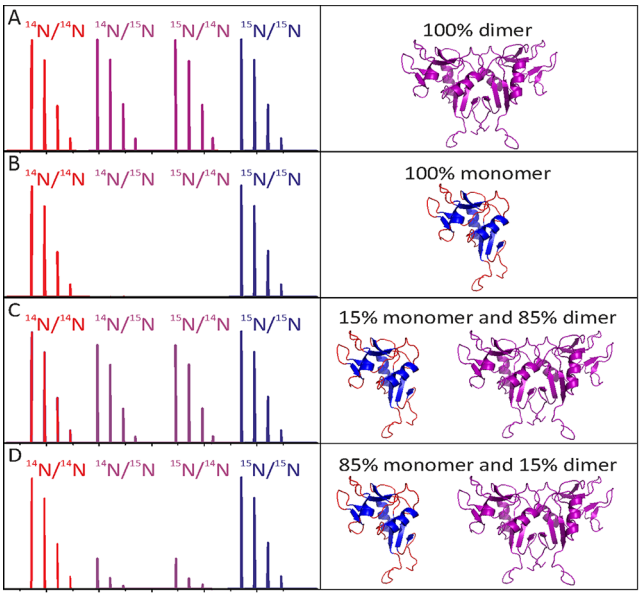


Figure 1. Theoretical mass spectra of type-2 products resulting from mixed-isotope cross-linking (MIX), where unlabeled (^{14}N) protein is mixed with ^{15}N-labeled protein in a 1:1 ratio before the cross-linking reaction. (A) Regular isotopic pattern for intermolecular type-2 products. (B) Isotopic pattern for intramolecular type-2 products. (C) Overlap of isotopic patterns for intermolecular and intramolecular type-2 products, where the intermolecular cross-link is preferentially formed, and the inter/intra ratio is 85:15. (D) Overlap of isotopic patterns for intermolecular and intramolecular type-2 products, where the intramolecular cross-link is preferentially formed, and the inter/intra ratio is 15:85

extracted from spectra. LinX is suitable for a large community of users because of its simple and intuitive graphical user interface (GUI), its possibilities for personalization, and the maximal independence of the type of mass spectrometer or vendor-specific software.

The LinX software was evaluated on three different biological systems: the Nkrp1b receptor of mouse natural killer cells,[20] lens epithelium derived growth factor/p75 (LEDGF/p75),[21] and the complex of transcription factor FOXO4 with its DNA response element.[22]

With regard to biological systems, natural killer cells represent a special type of toxic lymphocytes essential to innate immunity, which are regulated using receptors on target cells.[23] One of these receptors in mice is called Nkrp1b or Klrb1b. Here we studied data from its structural characterization by CXMS, where intermolecular cross-links were identified using MIX. Another system analyzed, LEDGF/p75, belongs to the group of transcription coactivators and forms crucial parts of the RNA polymerase II transcription system.[24] LEDGF/p75 was recently reported as an interaction partner of HIV1 integrase.[25] In this study, we focused on data from the CXMS analysis of the dimerization interface. The last protein tested, FOXO4, is a well-characterized human transcription factor, a member of the forkhead box class ("O" subfamily), which forms a complex with the DNA-binding element of Daf-16.[26] Data from the CXMS of the complex cross-linked by transplatin were used to examine the LinX tools for protein−nucleic acid cross-links.

EXPERIMENTAL SECTION

The experimental setup of protein cross-linking procedures followed previously published protocols.[20−22] A brief description of the experiments performed is provided as follows.

The 16 μM mixture of the Nkrp1b natural form and its ^{15}N-labeled form (at a 1:1 ratio) in 10 mM HEPES and 150 mM NaCl buffer (pH 7.4) was modified by 20 molar excess of disuccinimidyl glutarate (DSG) and disuccinimidyl suberate (DSS). The cross-linking reaction was quenched by 2 molar excess of ethanolamine after 2 h of incubation, and the resulting mixture was separated by SDS-PAGE in 12% polyacrylamide gel. Protein bands corresponding to the Nkrp1b homodimer were excised, reduced, alkylated, and digested via AspN and subsequently by trypsin protease overnight at 37 °C (in a 1:20 ratio). The peptides were extracted from the gel and characterized by analytical reversed-phase column online coupled to the Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer solariX XR.

The natural LEDGF/p75 protein was mixed with the 15 N form, and the mixture was subjected to gel filtration on a Superdex75 column. A corresponding fraction (20 μM) was cross-linked by a 10 molar excess of DSG at room temperature for 2 h and digested by trypsin in solution for 4 h at 37 °C (in a 1:10 ratio). The acquired peptides were separated on a reverse-phase column and measured using the FT-ICR mass spectrometer solariX XR.

The 20 μM complex of FOXO4 with the DNA-binding element in 150 mM ammonium acetate buffer (pH 6.85) or 10 mM HEPES, 50 mM NaCl(pH 7.4) was cross-linked by a 10 molar excess of *trans*-platinum(II)diammine dichloride (transplatin) for 14 h at 18 °C or by a 10 molar excess of DSG for 2 h at room temperature, respectively. The modified complex was then analyzed by native and SDS-PAGE electrophoresis. The digestion of the complex involved treatment with trypsin and nuclease BAL-31 complex and proceeded to liquid chromatography−MS/MS analysis in a data-independent mode. Data acquired from all experiments were processed using Data Analysis 4.4 (Bruker Daltonics, Billerica, MA), where the peptide signals were deisotoped and deconvoluted via the SNAP 2.0 algorithm to the text files and served as input for LinX.

RESULTS

Software Description

LinX is written in Java and runs on all devices installed with Java Runtime Environment version 8 or higher. The software is distributed as a compressed zip file containing all of the necessary components in a folder \LinX. Immediately after unpacking, the software is ready to use, with no need for further installations. LinX requires no additional large libraries; however, its functionality can be customized to specific applications by external add-ons called modules. The external module can be written in any programming language; the only restriction for the module is to read and return data in the LinX-compatible format. The syntax of the input/output data format is described in the documentation in detail. However, the working capacity of the system must be considered for new modules. LinX was designed for the analysis of single proteins or small complexes; currently, calculations for this type of data using a common desktop system (4−8 GB RAM) take a few seconds or minutes depending on the molecular weight of the protein, but a new module could increase the computation

https://dx.doi.org/10.1021/acs.jproteome.0c00858
J. Proteome Res. 2021, 20, 2021−2027

time. For example, the average speed of the algorithm is 8.2 ms per scan for the following parameters: studied system: 8 proteins; molecular weight:110 kDa; modifications: variable oxidation of methionine, fixed carbamidation of cysteines; cross-linking agent: DSG; protease: trypsin and AspN protease, four missed cleavages; mass accuracy: 2 ppm; computation system: 8 GB RAM system, processor: i6 1.8 GHz, operating system: Windows 10.

The simple LinX GUI is divided into three general sections. Input data are defined or uploaded in the first section. An intuitive one-way flow guides the user through several panels where protein sequences, proteases, post-translational modifications (PTMs), cross-linkers, and input MS data are selected for the assignment process. A menu bar also allows the user to upload settings from previous analyses, save current settings to the output file, or modify libraries of elements, PTMs, or cross-linkers. An informative console in the second section provides process state information. The last result section enables handling (sorting, Filtering, and Grouping) with a list of assigned masses. Additional external modules can be called from the results window. For example, the Draw module distributed with the software enables users to align the list of peptides to the protein sequence to immediately verify the sequence coverage and peptide distribution.

Software Workflow

The general workflow of LinX is shown schematically in Figure 2. Input 1 consists of the amino acid sequence(s) of the protein(s) in a single letter code, for example, a fasta file. Compared with the other cross-linking software, LinX allows multiple selection as well as the selection of only one sequence from the list of sequences in one fasta file. Input 1 also includes

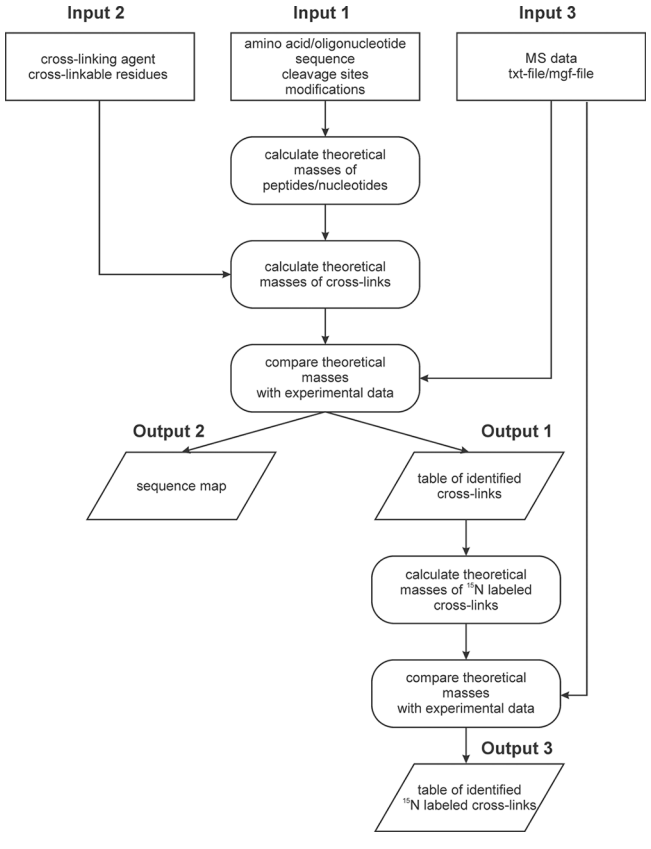


Figure 2. General scheme of LinX workflow.

information on static and variable modifications and protease cleavage sites. Unlike similar strategies to analyze cross-linking products,[1,11] the proteases site panel in LinX is formed by a specific semiautomatic script, and thus the search enables users to cleave from the N- and C-terminus in one analysis. Using theoretical proteolysis products resulting from Input 1 and considering the properties of the cross-linker (Input 2), all possible cross-links are calculated. The cross-links library generated is compared with the experimental data (Input 3), with user-specified mass accuracy. Input 3 must contain plain text values in delimiter (e.g., tabulator or space)-separated columns and can be loaded as a .txt or .mgf (Mascot generic file) file with predefined parameters. Additionally, the program allows the user to compare more than one mass list file in the same assignment process. Outputs of the program represent a table of positively assigned peptides and cross-links (Output 1) and protein maps showing protein sequence coverage, including modified peptides (Output 2) visualized by the Draw external module function. Output 1 represents the final table of LinX (Figure S1), which consists of two parts. The upper part is a window with parameters for the calculation. The second part is the final list of assigned unmodified peptides and peptides modified by cross-linking agents (dead-end cross-links/monolinks: type-0 products; intrapeptide cross-links: type-1 products; and interpeptide cross-links: type-2 products).[2] Multiple peptide modifications by cross-linking agents can be selected, and the final list can simultaneously contain peptides forming type-0 and type-1 or type-2 products. The results can be saved as a .sen file. In addition, the table contains several functions for more efficient work with annotated experimental data.

If LinX is used to identify intermolecular type-2 products in MIX homodimers, then Output 1 represents input for new calculations, where theoretical masses of ^{15}N-labeled cross-linked peptides are computed and compared with experimental data (Input 3). Unlike Output 1, the new list of type-2 products generated (Output 3) contains four extra columns called ^{14}N/^{14}N, ^{14}N/^{15}N, ^{15}N/^{14}N, and ^{15}N/^{15}N (Figure S2). Masses in these columns are highlighted in blue or red. Blue indicates that the charge mass for the precalculated mass (written in the cell) was positively assigned; therefore, it can be assumed that the corresponding form of the monomer has been present in solution. On the contrary, red means that the corresponding mass was not found in the experimental data. The shade of red/blue correlates with the intensity of the mass. The color becomes darker with increasing intensity of the mass.

In addition, the list of type-2 products contains the column labeled inter/intra ratio, which shows whether inter- and intramolecular cross-linked peptides are present in the sample at the same time. The ratio is calculated from the intensities of the isotopic patterns written in the experimental data (Output 3). This feature makes LinX a unique tool to evaluate data from MIX cross-linking, and even if some other software packages (Kojak, pLink2) can also be used to identify cross-links that originated from ^{15}N labeling,[27] they cannot distinguish whether the type-2 product identified is strictly intermolecular or if it is a mixture of inter- and intramolecular cross-links and which type dominates. Output 3 provides information on the original mixture of unlabeled (^{14}N) and ^{15}N-labeled protein as well. Information is represented by a correction factor that is computed from the intensities of all assigned peptides for each protein written in the Input 1 file.

https://dx.doi.org/10.1021/acs.jproteome.0c00858
J. Proteome Res. 2021, 20, 2021−2027

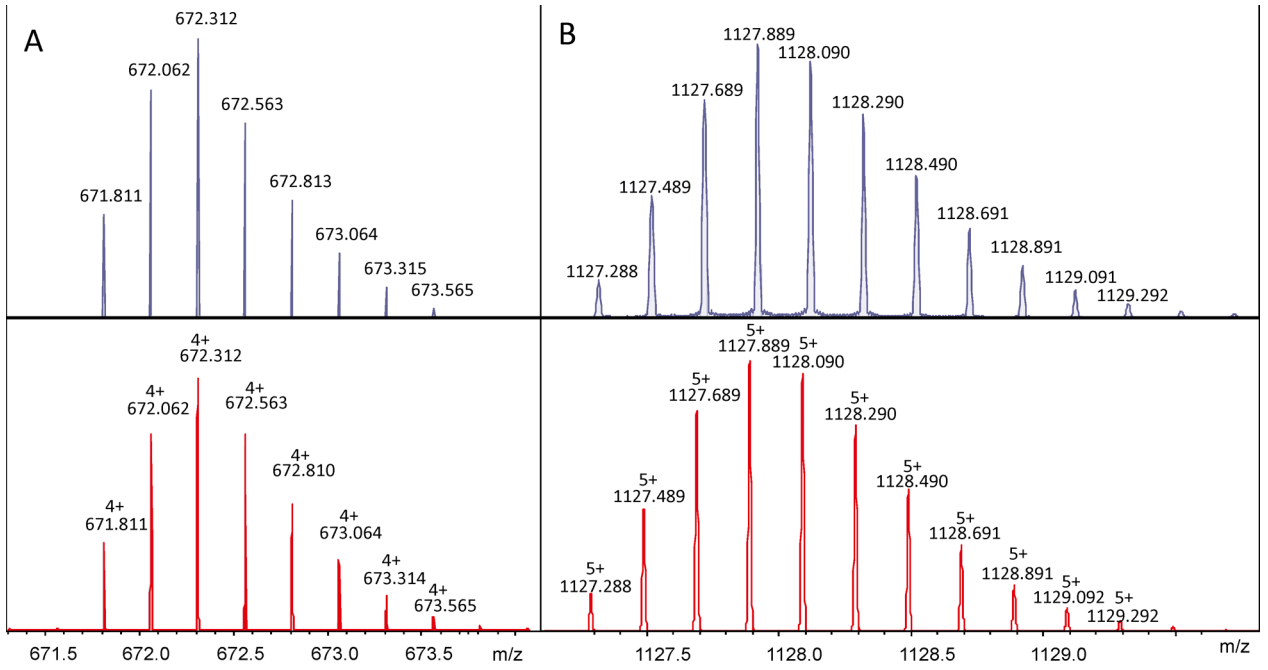


Figure 3. (A) Example of the protein−protein cross-link identified using LinX software. Peptide 74−88 from the FOXO4 protein cross-linked with peptide 152−159 by transplatin. Theoretical simulation of the isotopic envelope for the protein−protein cross-link (blue) compared with experimental data (red). (B) Example of the protein−nucleic acid cross-link identified using LinX software. Peptide 74−88 from the FOXO4 protein cross-linked with oligonucleotide CTTGTTTACCCAA by transplatin. Theoretical simulation of the isotopic envelope protein−DNA cross-link (blue) compared with experimental data (red).

The correction factor serves as a double check of whether unlabeled (^{14}N) and ^{15}N-labeled proteins are equimolar. If not, LinX warns the user(s) that the isotopic patterns in raw spectra will be affected and recommends that a manual verification of all identified intermolecular cross-links in the raw data must be performed.

Alternatives, Grouping, and Filtering

To simplify and improve work with results, several functions called Alternative, Grouping, and Filtering can be used. The grouping and filtering utilities enable users to join several lines into one line, according to the specified modifications, bonds, or masses, or to highlight data of specific interest and hide the rest, respectively. Several math operators, such as "intensity higher than 10^X" and "sequence contains...", can be selected and used as part of the filtering query. The Alternative search serves as a validation mechanism (target decoy search strategy) for the selected masses assignment. The search is executable from the results window by right-clicking the selected line(s) and choosing the option "Find alternatives" in the context menu. Although proteins are digested according to the rules selected in a standard search, the Alternative search employs a nonspecific digest of all sequences. All of the other parameters in the following panels remain active and are included in the new calculation. A detailed description of these functions is provided in the user's guide in the Supporting Information.

User-Defined Building Blocks

Free access to general settings enables users to define/redefine basic parameters of the script and modify them for the user's specific needs. Additional elements or isotopes that are commonly incorporated in isotopically coded cross-linkers, such as ^{13}C carbons[28] and ^{2}H hydrogens,[29,30] can be added. These isotopes must be defined as a new list item with a unique letter/character and corresponding atomic mass. Similar to elements, the basic block chain can be redefined; therefore, amino acids can be easily replaced by nucleotides, and LinX can be used for the analysis of nucleic acid−nucleic acid cross-links or even for protein−nucleic acid cross-links. This simple change requires the user to define nucleotides in the general settings with different letters from those used for amino acids and modified using phosphate.

Software Evaluation

LinX was validated on three different biological systems: Nkrp1b, LEDGF/p75, and the complex of transcription factor FOXO4 with its DNA response element. Using the Links tool, in-house-built software based on the ASAP,[1] we had previously identified nine intramolecular cross-links within the Nkrp1b molecule and six intermolecular cross-links between monomeric subunits. The data reevaluation using LinX revealed another six intermolecular restrains in addition to those previously described[20] (Figure S2; Table S1). Those additional distance constraints satisfied the structural models of Nkrp1b[20] and the X-ray structural model of the Nktp1b−ClrB complex.[31] The DSG distance constraints do not exceed 20 Å. Moreover, the quantitation feature of the LinX algorithm enabled us to calculate the inter/intra ratio, which identified intramolecular cross-links involved mainly in dimer formation.

Another example demonstrates the robustness of LinX to define intra/intermolecular restraints within the LEDGF/p75 homodimer. In a previous report, the inter/intra ratio was manually calculated from the raw data. The LinX analysis of an identical data set determined that the inter/intra ratio for all intermolecular cross-links identified is in agreement with previous results (Table S2) and the NMR structural model.[21] The LinX deviation did not exceed the manual quantification by more than 4%.

Furthermore, the analysis of the FOXO4/Daf16 complex by LinX revealed 37 protein−protein cross-links, which were

https://dx.doi.org/10.1021/acs.jproteome.0c00858
J. Proteome Res. 2021, 20, 2021−2027

identical to the previous report (Figure 3A)[22] where the DSG cross-linking data set was analyzed by StavroX (3.6.0.1).[15] The speed of the data interpretation was comparable (8 s for StavroX; 10 s for LinX). Moreover, LinX is capable of calculating oligonucleotides and thus enables the identification of protein−nucleic acid cross-links (Figure 3B). Seven peptide−oligonucleotide cross-links were identified and subsequently validated by the presence of a platinum atom, which has a specific isotopic pattern (Figure 3B). Those additional distance constraints nicely correlate with the reported structural models of the FOXO4/Daf16 complex.[22]

To define the reliability of the LinX algorithm, the ion signal assignment was performed on the hybrid fasta library consisting of the protein of interest, redundant proteins, trypsin, and their inverted sequences using variable mass accuracy and the nitrogen discriminator. In particular, the Nkrp1B data set was searched against the database that consists of the Nkrp1b protein sequence, the inverted Nkrp1b sequence, the LEDGF/p75 protein sequence, the inverted LEDGF/p75 sequence, the FOXO4 sequence, the inverted FOXO4 sequence, the trypsin protease sequence, and the inverted trypsin sequence. The parameters for the search were as follows: modifications: variable oxidation of methionine; variable type: 0 product of lysine, serine, threonine, and tyrosine, fixed carbamidation of cysteines; cross-linker: DSG or DSS; protease: trypsin, three missed cleavages; mass accuracy: 1−5 ppm. Any false-positive assignments were not observed when analyzing ^{14}N and ^{15}N samples (Nkrp1b and LEDGF/p75) up to mass accuracy 5 ppm. This target decoy analysis indicates the 0.0% false discovery rate (FDR) value for cross-linked peptides using the nitrogen discriminator. More false-positive assignments were reported not using the nitrogen discriminator: 3.0% FDR at mass accuracy <1 ppm, 5.0% FDR at mass accuracy <2 ppm, and 10.0% FDR at mass accuracy <3 ppm. Therefore, it is recommended to keep the mass accuracy to <1 ppm when the nitrogen discriminator is not in use, and it might be increased up to 3 ppm with the ^{15}N function on.

Benefits of LinX for ^{15}N-Labeled Protein Analysis

As previously mentioned, a few software packages, namely, DXMSMS Match,[32] Kojak,[7] pLink2,[9] and MeroX,[16] can also be used for MIX experimental data evaluation. It is difficult to compare these because they have objectives different from those of *LinX*. Because these software packages were developed to go through highly complex data sets from large proteomic studies and their subsequent quick interpretation, they focus mainly on the reading and analysis of fragmentation spectra; however, this focus could partially lead to the loss of information hidden in MS^1 spectra. The aim of LinX is quite the opposite. The idea behind the program was to design a tool for the analysis of a simple system (a single protein or a small complex whose sequence is known), considering the quantitative manner of MS data. LinX is based on reading MS^1 spectra, where users can elucidate the inter/intra ratio previously described, which illustrates whether the type-2 products identified from MIX are strictly intermolecular or a mixture of inter- and intramolecular cross-links and which type dominates. Taken together, compared with other software packages, LinX provides a different type of information on systems studied by MIX, and thus it can be used as a complementary tool for other searches and vice versa.

CONCLUSIONS

LinX is a free and open-source software for the evaluation of MS data from CXMS experiments. Originally, it was designed to provide a tool for the interpretation of high-resolution accurate mass data (1.0 ppm and lower); however, the simplicity of the input parameters enables users to use LinX to process data from any mass spectrometer in case MS^1 data are separately extracted from spectra. Nevertheless, LinX is not suitable for data analysis when the mass accuracy is >2.0 ppm. Although LinX was primarily intended for protein cross-linking experiments, its unique scripting design significantly extends the range of possible applications to nucleic acid−nucleic acid and protein−nucleic acid cross-linking. To our knowledge, it is currently the only available alternative tool in a situation when the novel covalent bond introduced by the cross-linker is labile and does not survive data-dependent acquisition. The major strengths of LinX are its easy-to-use GUI, the choice to easily improve it by external modules, and the highly automated analysis of CXMS data for the identification of intermolecular cross-links originating from homodimer interactions, including the calculation of the ratio that describes the occurrence of intermolecular cross-links in its mixture with intramolecular cross-links. Because the software was developed for single-protein or small-complex analysis, its only limitation might be the analysis of large biological complexes, where the calculation time would increase because of the random-access memory available. The software, including the source code and user guide, can be requested by e-mail to zdenek.kukacka@biomed.cas.cz or obtained by free download from the Web site https://github.com/KukackaZ/LinX or directly from https://ms-utils.org/LinX.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00858.

Figures S1 and S2. Examples of Output 1 and Output 3, respectively. Tables S1 and S2. Intra- and intermolecular cross-links identified from all proteins studied. Manual for the LinX software (PDF)

AUTHOR INFORMATION

Corresponding Author

Petr Novák − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic; Department of Biochemistry, Faculty of Science, Charles University, 12800 Prague 2, Czech Republic;* orcid.org/0000-0001-8688-529X; Phone: +420 325 873 610; Email: pnovak@biomed.cas.cz; Fax: +420 241 062 156

Authors

Zdeněk Kukačka − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic;* orcid.org/0000-0001-7569-843X

Michal Rosůlek − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic; Department of Biochemistry, Faculty of Science, Charles University, 12800 Prague 2, Czech Republic*

Jan Jelínek − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic; Department of Software Engineering, Faculty of Mathematics*

https://dx.doi.org/10.1021/acs.jproteome.0c00858
J. Proteome Res. 2021, 20, 2021−2027

and Physics, Charles University, 12000 Prague 2, Czech Republic

Lukáš Slavata − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic; Department of Biochemistry, Faculty of Science, Charles University, 12800 Prague 2, Czech Republic*

Daniel Kavan − *Institute of Microbiology, v.v.i., The Czech Academy of Sciences, Prague 14220, Czech Republic; Department of Biochemistry, Faculty of Science, Charles University, 12800 Prague 2, Czech Republic*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00858

Author Contributions

[‖]Z.K. and M.R. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

The software, including the source code and user guide, can be requested by e-mail to zdenek.kukacka@biomed.cas.cz or obtained by free download from the Web site https://github.com/KukackaZ/LinX or directly from https://ms-utils.org/LinX. Data are accessible via the ProteomeXchange server with the data set identifier PXD023522.

■ ACKNOWLEDGMENTS

We thank Alan Kadek for his help during the initiation of the project. This work was mainly supported by the Czech Science Foundation (grant number 19-16084S), the Ministry of Education of the Czech Republic (projects LH15010 and LD15089; program "NPU II" project LQ1604), COST Action (BM1403), European Regional Development Funds (CZ.1.05/1.1.00/02.0109 BIOCEV), and European Commission H2020 (European Network of Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry Centers - project agreement no. 731077).

■ ABBREVIATIONS

CXMS, chemical cross-linking coupled to mass spectrometry; ASAP, Automated Spectrum Assignment Program; DSG, disuccinimidyl glutarate; GPMAW, General Protein/Mass Analysis for Windows; GUI, graphical user interface; LEDGF/p75, lens epithelium derived growth factor/p75; MGF, mascot generic file; MS, mass spectrometry; MIX, mixed-isotope cross-linking; PTM, post-translational modification

■ REFERENCES

(1) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. High Throughput Protein Fold Identification by Using Experimental Constraints Derived from Intramolecular Cross-Links and Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (11), 5802−5806.

(2) Sinz, A. Chemical Cross-Linking and Mass Spectrometry for Mapping Three-Dimensional Structures of Proteins and Protein Complexes. *J. Mass Spectrom.* **2003**, *38* (12), 1225−1237.

(3) Fabris, D.; Yu, E. T. Elucidating the Higher-Order Structure of Biopolymers by Structural Probing and Mass Spectrometry: MS3D. *J. Mass Spectrom.* **2010**, *45* (8), 841−860.

(4) Zhang, H.; Tang, X.; Munske, G. R.; Tolic, N.; Anderson, G. A.; Bruce, J. E. Identification of Protein-Protein Interactions and Topologies in Living Cells with Chemical Cross-Linking and Mass Spectrometry. *Mol. Cell. Proteomics* **2009**, *8* (3), 409−420.

(5) Petrotchenko, E. V.; Borchers, C. H. Crosslinking Combined with Mass Spectrometry for Structural Proteomics. *Mass Spectrom. Rev.* **2010**, *29* (6), 862−876.

(6) Peri, S.; Steen, H.; Pandey, A. GPMAW - a Software Tool for Analyzing Proteins and Peptides. *Trends Biochem. Sci.* **2001**, *26* (11), 687−689.

(7) Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; Maccoss, M. J.; Davis, T. N.; Moritz, R. L. Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. *J. Proteome Res.* **2015**, *14* (5), 2190−2198.

(8) Kellersberger, K.; Yu, E. T.; Kruppa, G. H.; Young, M. M.; Fabris, D. Top-down Characterization of Nucleic Acids Modified by Structural Probes Using High-Resolution Tandem Mass Spectrometry and Automated Data Interpretation. *Anal. Chem.* **2004**, *76* (9), 2438−2445.

(9) Chen, Z.-L.; Meng, J.-M.; Cao, Y.; Yin, J.-L.; Fang, R.-Q.; Fan, S.-B.; Liu, C.; Zeng, W.-F.; Ding, Y.-H.; Tan, D.; et al. A High-Speed Search Engine PLink 2 with Systematic Evaluation for Proteome-Scale Identification of Cross-Linked Peptides. *Nat. Commun.* **2019**, *10* (1), 3404.

(10) De Koning, L. J.; Kasper, P. T.; Back, J. W.; Nessen, M. A.; Vanrobaeys, F.; Van Beeumen, J.; Gherardi, E.; De Koster, C. G.; De Jong, L. Computer-Assisted Mass Spectrometric Analysis of Naturally Occurring and Artificially Introduced Cross-Links in Proteins and Protein Complexes. *FEBS J.* **2006**, *273* (2), 281−291.

(11) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-Linked Peptides from Large Sequence Databases. *Nat. Methods* **2008**, *5* (4), 315−318.

(12) Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C.; Bruce, J. E. Informatics Strategies for Large-Scale Novel Cross-Linking Analysis. *J. Proteome Res.* **2007**, *6* (9), 3412−3421.

(13) Du, X.; Chowdhury, S. M.; Manes, N. P.; Wu, S.; Mayer, M. U.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. Xlink-Identifier: An Automated Data Analysis Platform for Confident Identifications of Chemically Cross-Linked Peptides Using Tandem Mass Spectrometry. *J. Proteome Res.* **2011**, *10* (3), 923−931.

(14) Panchaud, A.; Singh, P.; Shaffer, S. A.; Goodlett, D. R. XComb: A Cross-Linked Peptide Database Approach to ProteinGêÆProtein Interaction Analysis. *J. Proteome Res.* **2010**, *9* (5), 2508−2515.

(15) Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. StavroX-A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 76−87.

(16) Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3d-Structure Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (1), 83−97.

(17) Mohr, J. P.; Perumalla, P.; Chavez, J. D.; Eng, J. K.; Bruce, J. E. Mango: A General Tool for Collision Induced Dissociation-Cleavable Cross-Linked Peptide Identification. *Anal. Chem.* **2018**, *90* (10), 6028−6034.

(18) Tinnefeld, V.; Sickmann, A.; Ahrends, R. Catch Me If You Can: Challenges and Applications of Cross-Linking Approaches. *Eur. J. Mass Spectrom.* **2014**, *20* (1), 99−116.

(19) Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. Characterization of an Antagonist Interleukin-6 Dimer by Stable Isotope Labeling, Cross-Linking, and Mass Spectrometry. *J. Biol. Chem.* **2002**, *277* (48), 46487−46492.

(20) Hernychová, L.; Rosůlek, M.; Kádek, A.; Mareška, V.; Chmelík, J.; Adámková, L.; Grobárová, V.; Šebesta, O.; Kukačka, Z.; Skála, K.; et al. The C-Type Lectin-like Receptor Nkrp1b: Structural Proteomics Reveals Features Affecting Protein Conformation and Interactions. *J. Proteomics* **2019**, *196*, 162−172.

(21) Lux, V.; Brouns, T.; Čermáková, K.; Srb, P.; Fábry, M.; Mádlíková, M.; Hořejší, M.; Kukačka, Z.; Novák, P.; Kugler, M.; et al.

Molecular Mechanism of LEDGF/P75 Dimerization. *Structure* **2020**, *28* (12), 1288−1299.e7.
(22) Slavata, L.; Chmelík, J.; Kavan, D.; Filandrová, R.; Fiala, J.; Rosůlek, M.; Mrázek, H.; Kukačka, Z.; Vališ, K.; Man, P.; et al. MS-Based Approaches Enable the Structural Characterization of Transcription Factor/DNA Response Element Complex. *Biomolecules* **2019**, *9* (10), 535.
(23) Vivier, E.; Tomasello, E.; Baratin, M.; Walzer, T.; Ugolini, S. Functions of Natural Killer Cells. *Nat. Immunol.* **2008**, *9* (5), 503−510.
(24) Ge, H.; Si, Y.; Roeder, R. G. Isolation of CDNAs Encoding Novel Transcription Coactivators P52 and P75 Reveals an Alternate Regulatory Mechanism of Transcriptional Activation. *EMBO J.* **1998**, *17* (22), 6723−6729.
(25) Maertens, G.; Cherepanov, P.; Pluymers, W.; Busschots, K.; De Clercq, E.; Debyser, Z.; Engelborghs, Y. LEDGF/P75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells. *J. Biol. Chem.* **2003**, *278* (35), 33528−33539.
(26) Furuyama, T.; Nakazawa, T.; Nakano, I.; Mori, N. Identification of the Differential Distribution Patterns of MRNAs and Consensus Binding Sequences for Mouse DAF-16 Homologues. *Biochem. J.* **2000**, *349* (2), 629−634.
(27) Pettelkau, J.; Thondorf, I.; Theisgen, S.; Lilie, H.; Schröder, T.; Arlt, C.; Ihling, C. H.; Sinz, A. Structural Analysis of Guanylyl Cyclase-Activating Protein-2 (GCAP-2) Homodimer by Stable Isotope-Labeling, Chemical Cross-Linking, and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (12), 1969−1979.
(28) Brodie, N. I.; Makepeace, K. A. T.; Petrotchenko, E. V.; Borchers, C. H. Isotopically-Coded Short-Range Hetero-Bifunctional Photo-Reactive Crosslinkers for Studying Protein Structure. *J. Proteomics* **2015**, *118*, 12−20.
(29) Müller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **2001**, *73* (9), 1927−1934.
(30) Seebacher, J.; Mallick, P.; Zhang, N.; Eddes, J. S.; Aebersold, R.; Gelb, M. H. Protein Cross-Linking Analysis Using Mass Spectrometry, Isotope-Coded Cross-Linkers, and Integrated Computational Data Processing. *J. Proteome Res.* **2006**, *5* (9), 2270−2282.
(31) Balaji, G. R.; Aguilar, O. A.; Tanaka, M.; Shingu-Vazquez, M. A.; Fu, Z.; Gully, B. S.; Lanier, L. L.; Carlyle, J. R.; Rossjohn, J.; Berry, R. Recognition of Host Clr-b by the Inhibitory NKR-P1B Receptor Provides a Basis for Missing-Self Recognition. *Nat. Commun.* **2018**, *9* (1), 4623.
(32) Petrotchenko, E. V.; Serpa, J. J.; Makepeace, K. A. T.; Brodie, N. I.; Borchers, C. H. 14N15N DXMSMS Match Program for the Automated Analysis of LC/ESI-MS/MS Crosslinking Data from Experiments Using 15N Metabolically Labeled Proteins. *J. Proteomics* **2014**, *109*, 104−110.

https://dx.doi.org/10.1021/acs.jproteome.0c00858
J. Proteome Res. 2021, 20, 2021−2027