



RESEARCH ARTICLE

StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies

Michael Götze,¹ Jens Pettelkau,² Sabine Schaks,² Konstanze Bosse,²
Christian H. Ihling,² Fabian Krauth,² Romy Fritzsche,² Uwe Kühn,¹ Andrea Sinz²

¹Institute of Biochemistry and Biotechnology, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany

²Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin-Luther University Halle-Wittenberg, Wolfgang-Langenbeck-Str. 4, 06120 Halle (Saale), Germany

Abstract

Chemical crosslinking in combination with mass spectrometry has matured into an alternative approach to derive low-resolution structural information of proteins and protein complexes. Yet, one of the major drawbacks of this strategy remains the lack of software that is able to handle the large MS datasets that are created after chemical crosslinking and enzymatic digestion of the crosslinking reaction mixtures. Here, we describe a software, termed StavroX, which has been specifically designed for analyzing highly complex crosslinking datasets. The StavroX software was evaluated for three diverse biological systems: (1) the complex between calmodulin and a peptide derived from Munc13, (2) an N-terminal β -laminin fragment, and (3) the complex between guanylyl cyclase activating protein-2 and a peptide derived from retinal guanylyl cyclase. We show that the StavroX software is advantageous for analyzing crosslinked products due to its easy-to-use graphical user interface and the highly automated analysis of mass spectrometry (MS) and tandem mass spectrometry (MS/MS) data resulting in short times for analysis. StavroX is expected to give a further push to the chemical crosslinking approach as a routine technique for protein interaction studies.

Key words: Chemical crosslinking, Protein complexes, Mass spectrometry, Enzymatic digestion, Data analysis, Software

Introduction

Chemical crosslinking in combination with mass spectrometry (MS) has emerged as an alternative strategy to derive 3D structural information of proteins [1–9], which is reflected by the abbreviation “MS3D” to describe this fruitful combination of both techniques [3]. Chemical crosslinking relies on the introduction of a covalent bond between functional groups of amino acids within a protein (for

gaining insight into the conformation of a protein) or between different interaction partners (for elucidating interfaces in protein complexes) by a chemical reagent. After the crosslinking reaction, the proteins of interest are usually enzymatically digested, and the resulting peptide mixtures are analyzed by high-resolution mass spectrometry [4].

Analysis of crosslinked peptides by MS makes use of several advantages associated with MS analysis: The mass of the protein or the protein complex under investigation is theoretically unlimited because it is the proteolytic peptides that are analyzed, analysis is fast, and requires only femto- to attomole amounts of protein.

The functional groups of crosslinking reagents that are commonly used for this technique are amine-reactive *N*-hydroxysuccinimide (NHS) esters and photoreactive cross-

Electronic supplementary material The online version of this article (doi:10.1007/s13361-011-0261-2) contains supplementary material, which is available to authorized users.

Correspondence to: Andrea Sinz; e-mail: andrea.sinz@pharmazie.uni-halle.de

Received: 11 July 2011
Revised: 19 September 2011
Accepted: 20 September 2011
Published online: 25 October 2011

linkers (benzophenones, diazirines, and azides) [4, 10]. A number of strategies have been developed either for an enrichment of crosslinker-containing species by affinity chromatography [11], or for a facilitated MS identification of crosslinked products by using isotope-labeled crosslinkers [12–15], or crosslinkers that are MS cleavable and create characteristic neutral losses and fragment ions during tandem mass spectrometry (MS/MS) experiments [16–18].

We have recently shown for laminin N-terminal (LN) domains that even a few distance constraints imposed by chemical crosslinks and disulfide bonds, i.e., “natural” crosslinks, were sufficient for deriving a valid model [19] that closely resembled the structure obtained by X-ray crystallography [20]. So far, the largest protein complex investigated by chemical crosslinking and MS is the 15-subunit 670 kDa complex of RNA polymerase II (Pol II) with the transcription initiation factor TFIIF [21]. Yet, despite its straightforwardness, the greatest challenge of the crosslinking approach is posed by the high complexity of the created peptide mixtures requiring high-resolution MS techniques for analyzing the crosslinked products. Identifying crosslinked peptides poses additional difficulties for data analysis as the number of potential crosslinks increases quadratically with increasing sample complexity. Thus, bioinformatics tools are required in order to handle the large datasets generated during MS and MS/MS analyses of the peptide mixtures. For analyzing these complex datasets, there has been considerable effort to develop specific software tools; nevertheless, a software that allows a fully automated analysis of MS and MS/MS data created from crosslinked product mixtures is still lacking. Therefore, data analysis is still the bottleneck for the chemical crosslinking strategy to evolve into a generally applicable and rapid method for global structural proteomics studies, underlining the need to develop novel and powerful bioinformatics strategies. Among the currently available software for analyzing crosslinked products are General Protein/Mass Analysis for Windows (GPMW) [22], CoolToolBox, a major upgrade of the VIRTUALMSLAB software program [23], xQuest [24], X-Link Identifier [25], xComb [26], and MS-Bridge, which is part of *Protein Prospector* [27]. Summaries of currently available crosslinking software are found in [3] and [28].

In this report, we describe a software, termed StavroX, which is specifically designed for analyzing the highly complex mass spectrometric datasets that are obtained after chemical crosslinking of proteins and a subsequent digestion of the created reaction mixtures. The StavroX software was compared with several existing software programs for cross-linked product identification, with respect to time consumption and manual user input. For evaluating the StavroX software, we chose three diverse biological systems: (1) calmodulin (CaM) crosslinked to a Munc13 derived peptide with a heterobifunctional amine-reactive/photoreactive reagent, (2) disulfide bonds in an N-terminal β -laminin fragment as an example for naturally occurring crosslinks, and (3) the guanylyl cyclase activating protein-2 (GCAP-2) crosslinked to a peptide

derived from the retinal guanylyl cyclase (ROS-GC) with a homobifunctional amine-reactive reagent.

- (1) Munc13 proteins are important presynaptic regulators, which are essential for synaptic vesicle priming and adaptive synaptic mechanisms and are known to interact calcium-dependently with CaM [29]. Here, the amine-reactive photo crosslinker *N*-succinimidyl-*p*-benzoyldihydrocinnamate (SBC) was employed to study the interaction between CaM and a Munc13 peptide comprising the CaM binding region [30].
- (2) Laminins constitute a family of heterotrimeric glycoproteins, which are the main noncollagenous components of the basement membrane [31]. Recently conducted in-depth mass spectrometric analyses of the disulfide patterns in recombinant mouse laminin β 1 N-terminal fragments revealed a novel disulfide pattern for laminin-type epidermal growth factor-like (LE) domains [32].
- (3) The retinal guanylyl cyclase (ROS-GC) is a membrane protein in retina cells, which regulates the adaptation of the retina in response to light [33]. The interaction between ROS-GC and its binding partner GCAP-2 is currently studied in our group using a homobifunctional amine-reactive NHS ester.

Experimental

CaM/Munc13 Peptide Crosslinking

The crosslinking reaction with CaM (bovine brain; Calbiochem) and a 24-amino acid peptide derived from the CaM binding region of a Munc13 protein (synthesized by Dr. Olaf Jahn) was conducted in a two-step fashion. In a first step, the amine-reactive site of the crosslinker SBC [30] was reacted with CaM for 30 min. The reaction mixture contained 10 μ M CaM and the crosslinker SBC (200 or 500 μ M) in a Ca^{2+} /chelator (EGTA) buffer system (free Ca^{2+} concentration 30 nM; 10 mM HEPES buffer, pH 7.2). Excess of crosslinker was quenched with 20 mM NH_4HCO_3 and removed by microfiltration (Microcon YM-10; Millipore). In a second step, Munc13 peptide (10 μ M) was added to SBC-labeled CaM. The crosslinking reaction mixtures were irradiated with UV light in a home-built system (365 nm, irradiation energies 4000 and 8000 mJ/cm^2) to induce photocrosslinking. The reaction mixtures were separated by SDS-PAGE and bands of interest were in-gel digested with trypsin (Promega) according to an existing protocol [14]. Samples were stored at -20°C prior to nano-HPLC/nano-ESI-LTQ-Orbitrap-MS/MS analysis.

Disulfide Pattern in an N-Terminal Laminin β 1 Fragment

For assigning the disulfide pattern in an N-terminal fragment of laminin β 1 (one LN and four LE domains) a complete

alkylation of free cysteines was performed with iodoacetamide to prevent disulfide shuffling. Expression of recombinant mouse laminin β 1 chain fragments in human embryonic kidney 293 cells and purification from serum-free cell culture supernatant was performed as earlier described [32]. Enzymatic digestion of was done with trypsin [enzyme:substrate 1:16 (wt/wt)] overnight at 37 °C at pH 7.5. Enzymatic digestion of laminin β 1 was also performed under acidic conditions (pH 5.5) using LysN (U-ProteinExpress, The Netherlands) at an enzyme:substrate 1:16 (wt/wt) overnight at 50 °C [32]. Reactions were stopped with 10% (vol/vol) TFA solution, and the samples were stored at -80 °C before nano-HPLC/nano-ESI-LTQ-Orbitrap-MS/MS analysis was conducted.

GCAP-2/ROS-GC Peptide Crosslinking

Crosslinking between a peptide derived from the retinal guanylyl cyclase (ROS-GC, amino acids 965–981, YRIHVNRSTVQILSALN) and its binding partner GCAP-2 was performed with the amine-reactive homobifunctional NHS ester BS²G (*bis*[sulfosuccinimidyl]glutarate, Thermo Fisher Scientific). GCAP-2 was expressed in *E. coli* and purified according to an existing protocol (manuscript in preparation). Equimolar amounts (10 μ M) of GCAP-2 and ROS-GC peptide in 20 mM HEPES, pH 7.5, were equilibrated for 10 min with either 1 mM Ca²⁺ or 10 mM EGTA. The crosslinking reaction was started by adding the crosslinker BS²G at 100-fold excess (1 mM) and the reaction was quenched with 20 mM NH₄HCO₃. Aliquots were taken after 30 and 60 min. The crosslinking reaction mixtures were separated by SDS-PAGE, bands of interest were excised, and subjected to in-gel proteolysis with trypsin and GluC using an existing protocol [14]. The peptide mixtures were stored at -80 °C before nano-HPLC/nano-ESI-LTQ-Orbitrap-MS/MS analysis was performed.

Nano-HPLC/Nano-ESI-LTQ-Orbitrap Mass Spectrometry

Fractionation of proteolytic peptide mixtures was carried out on an Ultimate nano-HPLC system (Dionex Corporation, Idstein, Germany) using reversed phase C18 columns (precolumn: Acclaim PepMap, 300 μ m • 5 mm, 5 μ m, 100 Å, separation column: Acclaim PepMap, 75 μ m • 250 mm, 3 μ m, 100 Å, Dionex Corporation). After washing the peptides on the precolumn for 15 min with water containing 0.1% TFA, peptides were eluted and separated using gradients from 0% to 50% B (varying between 30 to 90 min), 50% to 100% B (1 min), and 100% B (5 min), with solvent A being 5% acetonitrile (ACN) containing 0.1% FA and solvent B being 80% ACN containing 0.08% FA. The nano-HPLC system was directly coupled to the nano-ESI source (Proxeon, Odense, Denmark) of an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). MS data were acquired in

data-dependent MS/MS mode: Each high-resolution full scan (m/z 300 to 2000, R=60,000) in the Orbitrap was followed by three or five product ion scans in the LTQ and/or the orbitrap (R=7500) on the three or five most intense signals in the full-scan mass spectrum (isolation window 2.5 u). Dynamic exclusion (exclusion duration 180 s, exclusion window -1 to 2 Th) was enabled to allow detection of less abundant ions. Data acquisition was controlled via XCalibur 2.0.7 (Thermo Fisher Scientific) in combination with DCMS link 2.0 (Dionex).

Analysis of Crosslinked Products

For comparison with the StavroX software, crosslinked peptides were analyzed using General Protein Mass Analysis for Windows (GPMW) [22] ver. 8.10 (Lighthouse Data, Odense, Denmark, <http://www.gpmaw.com>), CoolToolBox (CTB), which is a major upgrade of VIRTUALMSLAB [23], xQuest [24] (http://prottools.ethz.ch/orinner/public/htdocs/xquest/index_review.html), X-Link-Identifier [25] (<http://du-lab.org/XlinkIdentifier>), and MS-Bridge [27] (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msbridgestandard>).

Results and discussion

Program Workflow

The general workflow of StavroX is presented in Figure 1. StavroX uses the amino acid sequences of the proteins to be crosslinked, with potential amino acid modifications, using the protease as specified by the user (In1). From this, a peptide map is calculated. A list of all calculated peptides is displayed (Out1). Using the amino acid sequences and considering the properties of the crosslinker and the specified mass tolerances (In2), all possible crosslinks are calculated (D1).

Mass spectrometric data are loaded as mgf (Mascot generic file) files (In3; Figure 1). The precursor ion masses (MS), which are extracted from the mgf file, are compared with the masses of potential crosslinked products with user-defined mass accuracy. If no match is found for the extracted precursor ion mass, the next precursor ion mass from the mgf file is compared with masses of potential crosslinked products (negative). All identified matches are crosslinked product candidates, which are further analyzed (positive). The software calculates b- and y-type ions for all crosslinks between two peptides of a potential candidate and compares them to MS/MS data of the precursor ion (P5). The theoretical ion masses are calculated by adding the masses of the amino acids of the respective peptide. Conclusively, the crosslinked amino acid carries the additional mass of the crosslinker and the second peptide. The noise of a fragment ion mass spectrum is calculated iteratively. Ions exceeding the given signal-to-noise ratio are not

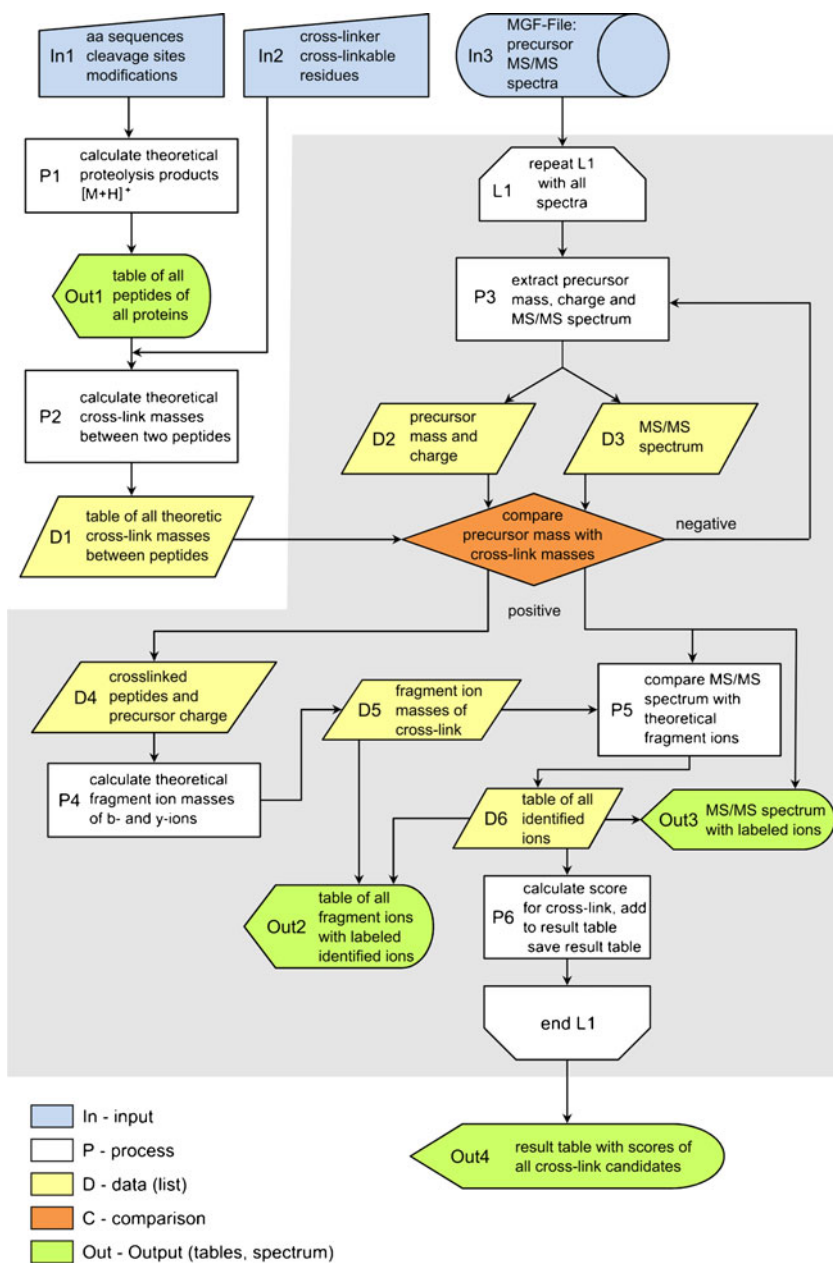


Figure 1. Workflow of the StavroX software

considered in the calculation, unless the noise remains identical within two consecutive iterations. In a first step, b- and y-type ions as well as neutral losses of the precursor ion (water, ammonia) are compared with the spectrum. Neutral losses (ammonia, water) of b- or y-type ions are only taken into account for previously identified b- or y-type ions. The identified ions of crosslinked residues are saved in csv file format for each peptide pair. Based on identified hits and ion series, a score is calculated (P6) for each crosslink-candidate and summarized in a results table. The graphical user interface includes the results table (Out4), a summary of all crosslink candidates, potential fragment ion masses of one pair of reactive sites with labeled identified ions

(Out2) as well as the respective fragment ion mass spectrum with ions labeled (Out3).

Software Description

As analyzing crosslinked products presents a complex task, human intervention is still required, but is greatly simplified by the StavroX software. The major strengths of StavroX comprise its easy-to-use graphical user interface as well as the highly automated analysis of MS and MS/MS data. StavroX runs without installation on Microsoft Windows platforms (98, NT, XP, Vista, 7). StavroX is a single executable file that was programmed and compiled using

Borland Delphi 4. The software can be obtained by sending an e-mail to michael.goetze@biochemtech.uni-halle.de.

StavroX is divided into three main parts: (1) In-silico proteolysis and calculation of proteolytic peptides and crosslinks, (2) comparison of mass spectrometric data (MS and MS/MS data) to calculated masses of crosslinked products, and (3) presentation of results. In-silico proteolysis is required to calculate potential crosslinked products for the proteins under investigation. Therefore, the user has to provide the respective amino acid sequences as well as the proteases used for enzymatic cleavage of crosslinked proteins. Amino acid sequences are imported in FASTA format. Enzymatic cleavage sites are defined by entering the specific amino acid next to the cleavage site with a question mark specifying N- or C-terminal cleavage, e.g., K? – cleavage occurs C-terminal of lysine, ?N – cleavage occurs N-terminal of asparagine. Amino acid sequences for highly sequence specific proteolysis can also be specified by the user, i.e., ENLYFQG? defining a TEV protease cleavage site. For each cleavage, the number of missed cleavages needs to be defined (“*missed cleavage factor*”). A missed cleavage factor of 1 implies that all resulting peptides contain up to one potential cleavage site that has not been subjected to proteolysis. It is also possible to enter an amino acid that prevents proteolysis at a specific site, i.e., trypsin will cleave with low frequency in case an Arg or Lys is followed by Pro. Variable modifications might be defined, such as carbamidomethylation of cysteines or oxidation of methionine, by entering the respective modified amino acid as well as the maximum number of modifications per peptide. Methionine oxidation is defined as follows: “M”

(methionine) is changed to “m” (oxidized methionine). The single letter code is used for entering the amino acids. Modified amino acids are simply added to the amino acid code by their elemental composition with the letters defined by the user. Conclusively, the resulting mass list of proteolytic peptides is extended by the number of modified peptides. For calculating the masses of crosslinked products, the mass of the crosslinker is added to the masses of two proteolytic peptides. The crosslinker is defined by its elemental composition and its reactivities at both reactive sites separately, e.g., amine- and photo-reactive. Depending on the MS method used, the accuracies and mass limits for precursor ion mass measurements and fragment ion mass measurements are defined. These settings can be saved individually for different analyses.

MS data are loaded as standard Mascot generic file (mgf) containing all MS/MS data for each precursor ion that was fragmented. As soon as data analysis is started by the user data analysis will proceed without any further input by the user. Analysis of each dataset presented herein required calculation times between 30 to 60 min. The results are automatically saved for a subsequent analysis and are summarized in a table presenting the identified crosslink candidate with its corresponding peptides, proteins, masses, mass deviation, scan number, and score (Figure 2). By double-clicking on one crosslink candidate, a summary is shown, including all combinations of crosslinked amino acids, the number of identified fragment ions, and fragment ion series (b- and y-type ions). Double-clicking on one pair of crosslinked residues will show the fragment ion comparison sheet with all identified ions and the fragment ion mass

Score	Measured	Calculated	ppm	Protein	From	To	Peptide 1	Protein	From	To	Peptide 2	Scan
176	4673.1393	4673.1368	0.5	>spIP11012ILAM_N	150	166	[AYDCESSFPGISTGPMK]	>spIP11012ILAM_N	167	191	[KVDDIICDSRYSDIEPSTEG	4044
156	4673.1393	4673.1368	0.5	>spIP11012ILAM_N	150	165	[AYDCESSFPGISTGPM]	>spIP11012ILAM_N	166	191	[KKVDDIICDSRYSDIEPSTE	4044
149	4673.1393	4673.1368	0.5	>spIP11012ILAM_N	150	160	[AYDCESSFPGI]	>spIP11012ILAM_N	161	191	[STGPMKKVDDIICDSRYSDI	4044
143	2800.3671	2800.3644	1.0	>spIP11012ILAM_N	32	42	[KLSVTSTCGLH]	>spIP11012ILAM_N	43	56	[KPEPYCIVSHLQED]	2855
142	4673.1393	4673.1368	0.5	>spIP11012ILAM_N	150	155	[AYDCES]	>spIP11012ILAM_N	156	191	[SFPGISTGPMKKVDDIICDS	4044
140	2782.3559	2782.3539	0.7	>spIP11012ILAM_N	32	56	[KLSVTSTCGLHKPEPYCIVS	>intrapeptidal	0	1	(i)	3575
136	4386.9905	4386.9967	-1.4	>spIP11012ILAM_N	37	57	[STCGLHKPEPYCIVSHLQEC	>spIP11012ILAM_N	406	423	[AGQCRCRKLHVEGERCDVC]	2502
128	2229.9932	2229.9903	1.3	>spIP11012ILAM_N	10	31	[SYGCAEGSCYPATGDLLIGF	>intrapeptidal	0	1	(i)	3886
125	2229.9920	2229.9903	0.7	>spIP11012ILAM_N	10	31	[SYGCAEGSCYPATGDLLIGF	>intrapeptidal	0	1	(i)	4134
124	2229.9930	2229.9903	1.2	>spIP11012ILAM_N	10	31	[SYGCAEGSCYPATGDLLIGF	>intrapeptidal	0	1	(i)	4597
120	4386.9905	4386.9967	-1.4	>spIP11012ILAM_N	316	324	[KKCNCNEHS]	>spIP11012ILAM_N	401	430	[SVGLIAGQCRCRKLHVEGER	2502
119	3240.5057	3240.4976	2.5	>spIP11012ILAM_N	1	31	[APLVQPEPEFSYGCAEGSCY	>intrapeptidal	0	1	(i)	5008
119	3240.5057	3240.4976	2.5	>spIP11012ILAM_N	0	31	[APLVQPEPEFSYGCAEGSCY	>intrapeptidal	0	1	(i)	5008
119	2782.3599	2782.3539	2.2	>spIP11012ILAM_N	32	56	[KLSVTSTCGLHKPEPYCIVS	>intrapeptidal	0	1	(i)	3427
117	4386.9905	4386.9839	1.5	>spIP11012ILAM_N	34	56	[SVTSTCGLHKPEPYCIVSHL	>spIP11012ILAM_N	388	405	[SENGGICDGYTDFSVGLI]	2502
116	4386.9905	4386.9967	-1.4	>spIP11012ILAM_N	316	323	[KKCNCNEH]	>spIP11012ILAM_N	401	431	[SVGLIAGQCRCRKLHVEGER	2502

Figure 2. Screenshot of a result list showing the identified crosslink candidates with scores, masses, mass deviations, amino acid sequences, and the scan number (MS), exemplified for laminin β 1 disulfide analysis. The table can be sorted by clicking the header. By double clicking on one candidate in the list, the respective candidate can be investigated in more detail

spectrum (MS/MS) with all fragment ions assigned (Figure 3). Thus, ions resulting from fragmentation of a potential crosslink candidate are readily visible. So far, StavroX calculates b- and y-type ions, which mainly occur during collision-induced dissociation (CID), in addition to constant neutral losses (water and/or ammonia) from previously identified fragment ions as well as from the precursor ion.

Scoring

The scoring algorithm reflects the quality of the respective fragment ion mass spectrum, which is calculated from the number of signals above a specified signal-to-noise ratio. The score is based on the number of identified b- and y-type ions as well as on the number and length of the ion series.

The score is calculated as follows:

$$\text{Score} = -50 \cdot \log \left[\prod_n e^{-\frac{s_n}{p1+p2}} \cdot \left(0.2 \cdot \left(1 - e^{-\frac{|d-300|^5}{10^{12}}} \right) + 0.2 \cdot e^{-\frac{l}{6}} + 0.4 \cdot e^{-7 \cdot \frac{k}{l}} + 0.2 \cdot e^{-20 \cdot \frac{l}{d}} \right) \right] \quad (1)$$

with: s_n : Length of series n (b- or y-type ions); $p1$, $p2$: Length of crosslinked peptides 1 and 2; d : Number of fragment ions in the observed spectrum above threshold

($140 < d < 460$); i : Number of signals above 10% relative intensity; k : Number of identified fragment ions; h : Number of all identified ions ($h > d/10$)

StavroX uses non-probabilistic parameters to determine the score for a crosslink candidate. To estimate the quality of a fragment ion spectrum the total number of fragment ions above the threshold as well as the number of signals with relative intensities above 10% are taken into account. The length of the respective b- or y-type ion series also influences the score: Each y- and b-type ions series of every crosslinked peptide is divided by the total length of the peptide. This prohibits a potential under-representation of short crosslinked peptides by taking into account that short peptides do not produce long series of fragment ions. In case the respective crosslink candidate is a true match, the calculated exponential term is very small. The calculated probability is reflected by the fact that the observed match between the experimental data and the predicted mass of a crosslink candidate is a random event. A logarithmic conversion of this probability yields the score that is displayed by StavroX. All factors of eq 1 were adapted to obtain score values larger than 100 for highly probable crosslinked products.

False-Positive Rate

For estimating the false-positive rate, six datasets were taken into account comprising a total of 3539 crosslink

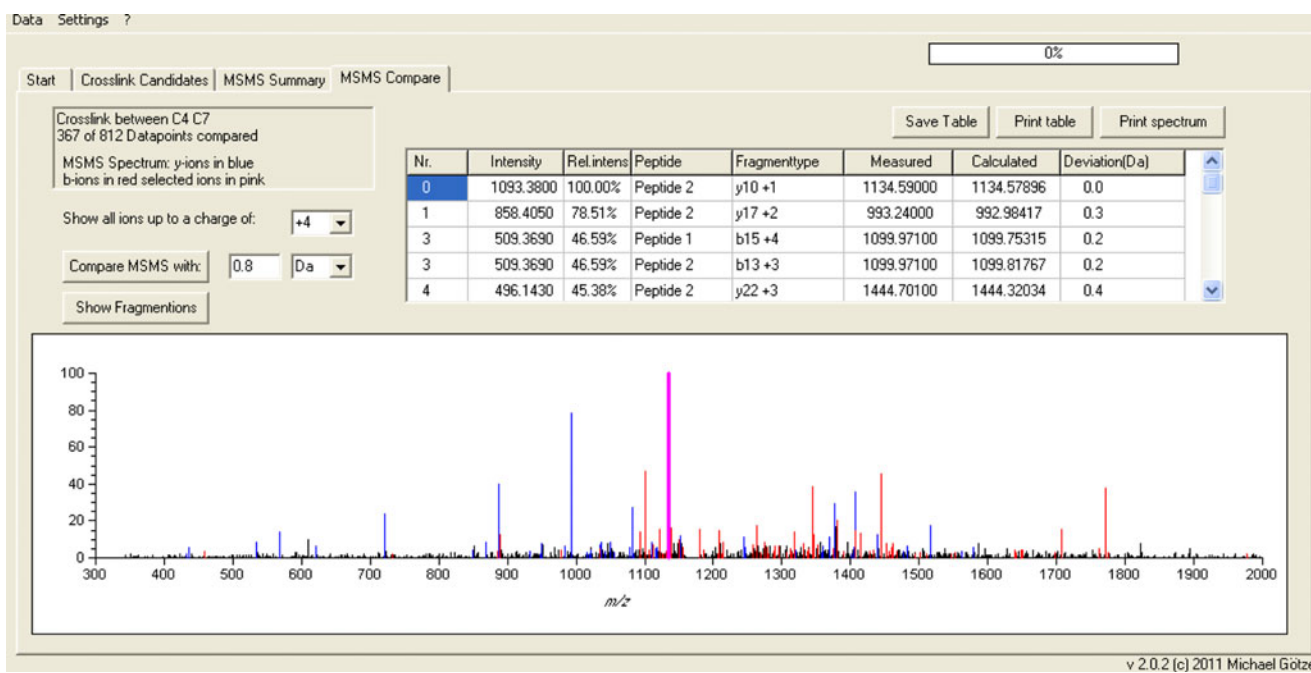


Figure 3. Screenshot of the fragment ion mass spectra (MS/MS) comparison sheet. The identified ions are presented as a table and are labeled in the spectrum; signals of y-type ions are shown in blue, while signals of b-type ions are shown in red. The fragment ion selected from the table is shown in pink. Each fragment ion mass spectrum can be reanalyzed using different user-defined parameters. As default, fragment ions with a charge up to the charge of the precursor are listed and labeled

candidates. Searches were performed with correct parameters as well as with a decoy search using reversed protein sequences. Candidate lists from both searches were united and only those candidates that were found to be true crosslinks during manual inspection were assigned as positives. The number of false-positives above a certain score value were added and divided by

the total number of crosslink candidates. For scores larger than 100, the false-positive rate is ca. 2%.

Evaluation of StavroX

The StavroX software was tested on three diverse biological systems: (1) The complex between CaM and a peptide

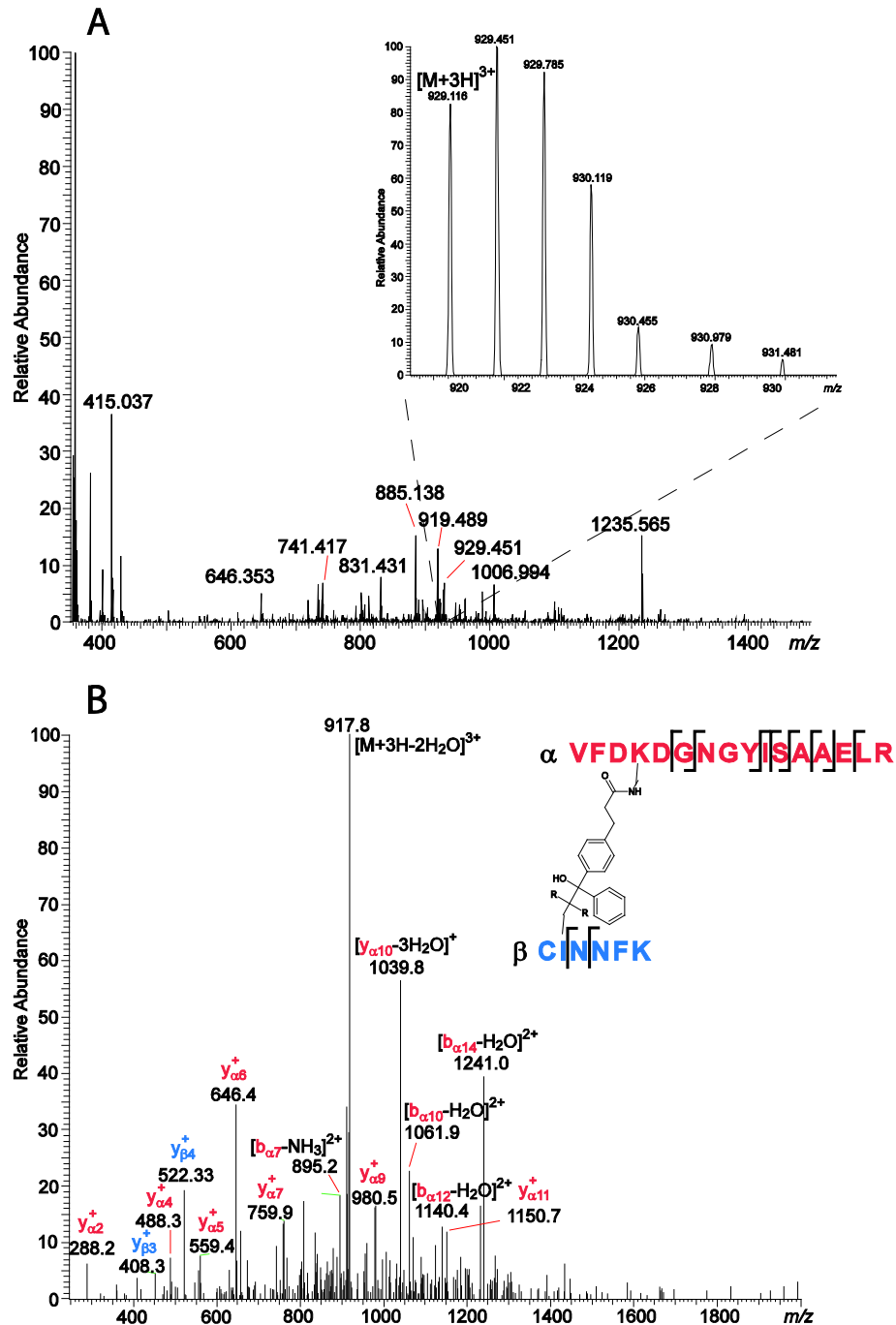


Figure 4. **(A)** Mass spectrum with the enlarged signal of the triply charged precursor ion at m/z 929.116 that was selected for fragmentation. **(B)** Fragment ion mass spectrum of the crosslink between calmodulin (CaM) (amino acids 91–106) and a Munc13 peptide (amino acids 1–6); Lys-94 of CaM was found to be crosslinked with Ile-2 of the Munc13 peptide. Ions of the crosslinked α -peptide are represented in red, while ions of the β -peptide are represented in blue. The nomenclature of the crosslinked product is according to [34]

derived from Munc13, (2) an N-terminal β 1 laminin fragment, and (3) the complex between guanylyl cyclase activating protein-2 (GCAP-2) and a peptide derived from the retinal guanylyl cyclase (ROS-GC).

Calmodulin/Munc13 Peptide Interaction

Using CoolToolBox (CTB), we had previously identified several crosslinks between CaM and a Munc13 peptide [29]. When we reanalyzed the data with StavroX, those crosslinks were confirmed—and more importantly—some additional ones were discovered (Figure 4; Table 1). It should be mentioned that some of the crosslinks that were identified by CTB gained only low scores with StavroX, but these ambiguous crosslinks were readily confirmed by a quick manual inspection of fragment ion mass spectra.

The crosslinks identified between CaM and the Munc13 peptide revealed that the amine-reactive site of the crosslinker SBC had mainly reacted with lysines 21 and 94 of CaM. Merely one crosslinked product was identified with Lys-13 of CaM (Table 1). The photophore of SBC was found to have reacted with the hydrophobic amino acids Leu-9 and Leu-14, Val-8, and Ile-2 (Figure 4) of the Munc13 peptide. Additional crosslinks were identified with Lys-12 and Lys-17 of the Munc13 peptide.

Disulfide Analysis for Laminin β 1 N-Terminal Fragment

Previously conducted in-depth mass spectrometric analyses of the disulfide patterns in recombinant mouse laminin β 1 N-terminal fragments comprised of one LN and four LE domains had revealed a novel disulfide pattern for LE domains, in which the last cysteine of one LE domain is connected to the first cysteine in the following domain [32]. Reanalyzing the data with StavroX not only confirmed the disulfide pattern that had already been predicted (Figure 5; Table 2), but exceeded those previous results: a number of disulfides were identified in mass spectra of additional

precursor ions resulting in an overall higher number of identified disulfide bonds. Therefore, StavroX possesses advantages for analyzing disulfide patterns in proteins underlining its versatile application.

GCAP-2/ROS-GC Peptide Interaction

Analyses of the interaction between GCAP-2 and potential binding peptides derived from ROS-GC by chemical crosslinking and MS revealed the presence of several crosslinks when investigating the data with GPMW, xQuest, and StavroX (Figure 6; Table 3). In all crosslinks with the amine-reactive crosslinker BS²G, the N-terminal tyrosine of the ROS-GC peptide was involved. Although the number of identified crosslinks is identical for StavroX and GPMW, employing StavroX greatly facilitated data analysis and reduced the time to comprehensively screen the datasets. In contrast to GPMW, StavroX calculates all masses of crosslinks between different amino acids with a number of variable modifications simultaneously, while GPMW allows merely searching for crosslinks between two defined amino acids, with fixed modifications of the respective crosslinked peptides. Conclusively, the crosslink between Ser-37 of GCAP-2 and the N-terminal tyrosine of the ROS-GC peptide was only identified with GPMW after conducting a number of additional time-consuming analyses requiring a manual variation of the fixed modifications. With StavroX, only a single analysis cycle was required—without the need for further manual input—in order to gain the same amount of information.

Comparison of StavroX to Existing Crosslinking Software

To appreciate the efficiency of StavroX, we compared our software with existing programs for analyzing crosslinked products, namely GPMW [22], CTB [23], xQuest [24], Xlink-Identifier [25], and MS-Bridge (ProteinProspector) [27]. In conclusion, all crosslinks identified by CTB or GPMW were also identified by StavroX, but StavroX identified

Table 1. Analysis of crosslinks between CaM and a Munc 13 peptide with the heterobifunctional crosslinker SBC. The crosslinked product shown in Figure 4 is highlighted

<i>m/z</i> measured	Calculated molecular weight	Charge state	Mass deviation (ppm)	Crosslinked aa (CaM)	Crosslinked aa (Munc13)	Identified with StavroX	Identified with CTB	Identified with Xlink-Identifier	Identified with xQuest
1082.047	4325.161	4	1.6	K13	K17	√	–	–	–
794.178	3173.685	4	2.0	K94	V8/L9	√	√	–	√
929.116	2785.329	3	0.9	K94	I2	√	√	–	√
696.109	2781.409	4	1.1	K21	K12	√	√	√	√
927.483	2780.425	3	2.4	K21	L14	√	–	√	√
916.808	2748.399	3	–2.3	K94	L9	–	√	–	–
898.140	2691.389	3	1.3	K94	K12	√	√	√	√
673.358	2690.405	4	2.1	K94	L14	√	–	√	√
875.449	2624.324	3	2.9	K21	L14	√	–	–	√
845.442	2534.304	3	2.8	K94	L14	√	–	–	√
842.097	2524.272	3	2.4	K21	L14	√	–	√	√
831.094	2491.262	3	2.2	K94	V8/L9	√	–	–	√
812.091	2434.251	3	2.7	K94	L14	√	–	–	√

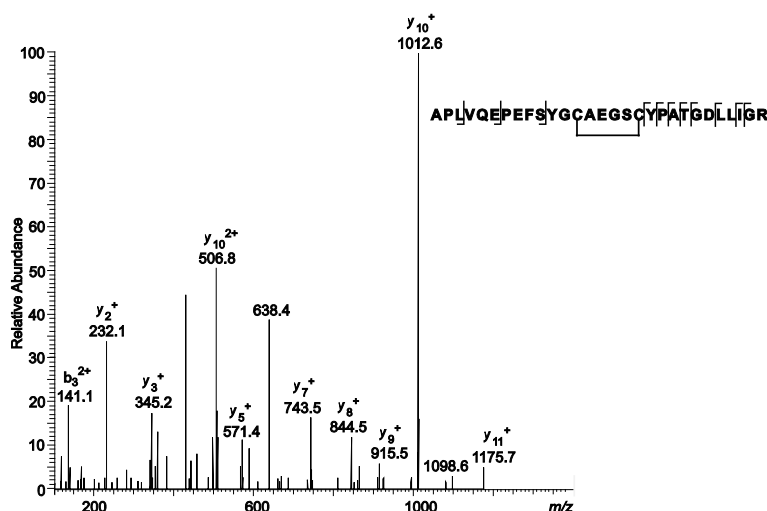


Figure 5. Fragment ion mass spectrum of an intramolecular disulfide bond (amino acids 1–29, cysteines 13 and 18 are connected) in the LN domain of laminin β 1. Fragmentation sites as well as disulfide bridged cysteine residues are indicated in the amino acid sequence

additional crosslinks without requiring any further input by the user. One major advantage of StavroX consists in the possibility to simultaneously analyze crosslinked products for peptides with different variable modifications. Additionally, crosslinks are calculated for all amino acid combinations that are specified by the user, and not only for selected ones (i.e., lysines in case amine-reactive crosslinkers are employed). In GPMAW, separate searches have to be performed for each peptide combination with different modifications. Moreover, mass lists have to be copied in groups of 500 entries into the program and each crosslink analysis is performed separately. Considering the high versatility of StavroX that allows searching for all combinations of crosslinked peptides, time

consumption for analyzing a whole crosslinking dataset is much lower compared with a similar analysis with GPMAW.

A further strength of StavroX over GPMAW, CTB, and MS-Bridge is the direct inclusion of MS/MS data and the visualization of labeled fragment ion mass spectra, which allows the user to directly decide about the quality of a crosslink assignment. In order to validate a crosslink candidate with CTB or GPMAW, time-consuming manual intervention is required to compare MS/MS data with theoretical fragmentation patterns, as neither of those programs can automatically handle MS/MS data. Xlink-Identifier and xQuest allow examining fragment ion mass spectra online, but only with Xlink-Identifier it is possible to

Table 2. Analysis of the disulfide pattern of a laminin β 1 N-terminal fragment. The crosslinked product shown in Figure 5 is highlighted; n. a. denotes these disulfide bonds that were found after proteolysis with LysN, which could not be analyzed either by Xlink-Identifier or by xQuest

<i>m/z</i> measured	Calculated molecular weight	Charge state	Mass deviation (ppm)	Cys	Cys	Identified with StavroX	Identified with GPMAW	Identified with MS-Bridge	Identified with Xlink-Identifier	Identified with xQuest
1149.939	5745.663	5	-2.0	153	173	√	√	–	n. a.	n. a.
1322.618	5287.443	4	1.5	153	173	√	√	–	n. a.	n. a.
1169.040	4673.137	4	0.5	153	173	√	√	–	n. a.	n. a.
860.886	3440.524	4	-0.6	59	62	√	–	–	√	–
819.869	3276.453	4	0.3	153	173	√	–	√	√	√
815.871	3260.458	4	1.0	153	173	√	–	–	√	√
1087.492	3260.458	3	1.3	153	173	√	√	–	√	√
1079.821	3237.446	3	0.6	153	173	√	√	–	n. a.	n. a.
1050.125	3148.358	3	0.6	153	173	√	–	–	√	√
787.846	3148.358	4	1.0	153	173	√	–	–	√	√
1044.794	3132.363	3	0.9	153	173	√	√	–	√	√
1566.689	3132.363	2	2.1	153	173	√	√	–	√	√
783.847	3132.363	3	1.2	153	173	√	√	–	√	√
1014.474	3041.402	3	1.2	13	18	√	√	√	√	–
1521.206	3041.402	2	0.5	13	18	√	√	√	√	–
761.107	3041.402	4	1.0	13	18	√	√	√	√	–
728.368	2910.449	4	0.2	39	48	√	√	√	√	–
696.345	2782.354	4	1.8	39	48	√	√	√	√	–
738.976	2214.912	3	1.0	473	476	√	√	√	–	–
641.947	1923.827	3	0.1	357	360	√	√	√	–	–
421.171	841.334	2	0.6	59	62	√	√	√	–	–

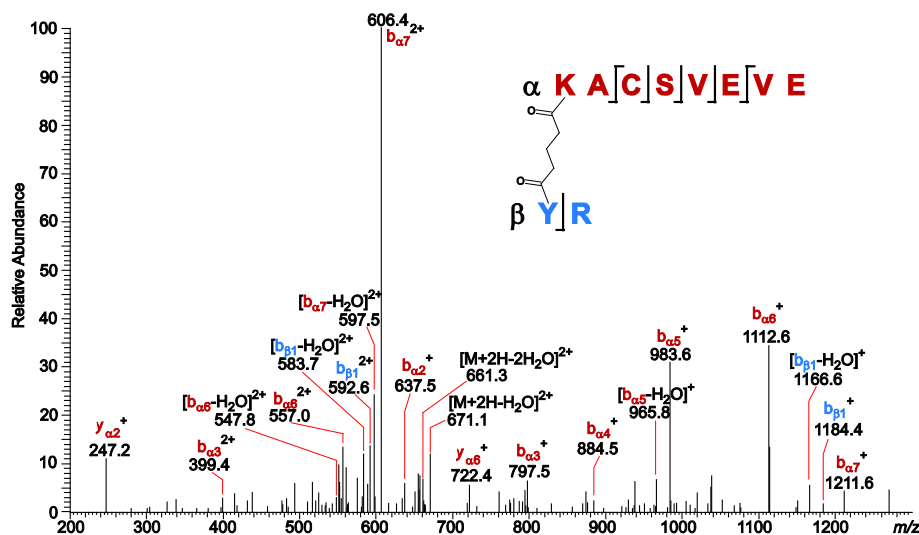


Figure 6. Fragment ion mass spectrum of the crosslink between GCAP-2 (amino acids 129-136) and a ROS-GC peptide (amino acids 1-2). Lys-129 of GCAP-2 and the N-terminal Tyr of the GC peptide were found to be crosslinked. The nomenclature of the crosslinked product is according to [34]

download these spectra and store them. Xlink-Identifier and MS-Bridge exhibit a number of drawbacks: Xlink-Identifier does not allow analyzing crosslinked peptides that have been generated by a protease other than trypsin, while MS-Bridge only accepts crosslinkers with reactivities towards amine or sulfhydryl groups. For some proteins studied by chemical crosslinking it is relevant to specify additional fixed amino acid modifications, e.g., an acetylated N-terminus or a methylated lysine, which is not implemented in xQuest. Also, the number of fixed and variable modifications in xQuest is restricted to a maximum of two or three, which

might be problematic for highly oxidized, methylated, and acetylated proteins.

Analyzing three datasets with different crosslinking software revealed the advantages of our StavroX software, namely, the options to define crosslinker reactivities, specific modifications of single amino acids, and a high number of fixed and variable modifications. Visualization of fragment ion mass spectra greatly shortens and simplifies data analysis. StavroX combines the advantages of existing software programs for analyzing crosslinked products and allows screening of crosslinking data in a highly versatile and efficient manner.

Table 3. Interaction analysis between GCAP-2 and a ROS-GC peptide with and without calcium using the homobifunctional amine-reactive crosslinker BS²G. For the ROS-GC peptide, the N-terminal Tyr was involved in all crosslinks. The crosslinked product shown in Figure 6 is highlighted

<i>m/z</i> measured	Calculated molecular weight	Charge state	Mass deviation (ppm)	aa of GCAP-2 crosslinked with ROS-GC peptide	With Ca ²⁺	Without Ca ²⁺	Identified with StavroX	Identified with GPMaw	Identified with xQuest
531.291	1061.572	2	3.3	K-29	√	–	√	√	–
588.812	1176.613	2	2.5	K-200	√	–	√	√	√
594.850	1188.692	2	0.0	K-126	–	√	√	√	–
594.851	1188.692	2	2.8	K-126	√	–	√	√	–
622.327	1243.644	2	2.6	K-178	√	–	√	√	√
623.822	1246.636	2	0.0	K-106	–	√	√	√	√
630.358	1259.708	2	0.1	K-98	–	√	√	√	√
630.360	1259.708	2	3.1	K-98	√	–	√	√	√
679.832	1358.656	2	0.3	K-129	–	√	√	√	–
679.834	1358.656	2	2.9	K-129	√	–	√	√	–
467.255	1399.745	3	3.3	K-178	√	–	√	√	√
781.378	1561.747	2	0.4	K-50	–	√	√	√	–
781.380	1561.747	2	3.1	K-50	√	–	√	√	–
603.649	1808.927	3	3.3	K-102	√	–	√	√	√
711.991	2133.956	3	0.6	S-37	–	√	√	√	–
1125.090	2249.171	2	0.2	K-142	–	√	√	√	–
1125.092	2249.171	2	2.9	K-142	√	–	√	√	–
750.396	2249.171	3	0.5	K-142	–	√	√	√	–
750.398	2249.171	3	3.3	K-142	√	–	√	√	–
1127.037	2253.065	2	0.5	K-50	–	√	√	√	–
754.689	2262.051	3	0.9	K-30	–	√	√	√	–

Conclusions

The combination of chemical crosslinking of proteins and MS has matured into an alternative technique for gaining structural information on proteins. Yet, the greatest deficit of this approach is still presented by the lack of efficient bioinformatics tools that allow analyzing data in a fully automated fashion. Therefore, the development of novel software programs for a facilitated analysis of crosslinked products is of utmost importance. The StavroX software presented herein is highly advantageous for analyzing data of crosslinked products in respect to its easy-to-use graphical user interface and its highly automated analysis of MS and MS/MS data resulting in short analysis times.

Acknowledgments

A.S. acknowledges financial support from the DFG (German Research Foundation, projects Si 867/13-1 and Si 867/15-1) and the BMBF (ProNet-T3, Project To-06). M.G. is funded by the Graduiertenkolleg 1591 “*Posttranscriptional Control of Gene Expression: Mechanisms and Role in Pathogenesis*”, S.S. is funded by the Graduiertenkolleg 1026 “*Conformational Transitions in Macromolecular Interactions*” at the Martin-Luther University Halle-Wittenberg. K.B. is funded by the EU (LUNGTARGET project). Dr. Olaf Jahn is acknowledged for synthesizing the Munc13 peptide. The authors are indebted to Professor Elmar Wahle for supporting this work.

References

- 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 Crosslinks and Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5802–5804 (2000)
- Back, J.W., de Jong, L., Muijsers, A.O., de Koster, C.G.: Chemical Crosslinking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **331**, 303–313 (2003)
- Fabris, D., Yu, E.T.: Elucidating the Higher-Order Structure of Biopolymers by Structural Probing and Mass Spectrometry: MS3D. *J. Mass Spectrom.* **45**, 841–860 (2010)
- Sinz, A.: Chemical Crosslinking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–Protein Interactions. *Mass Spectrom. Rev.* **25**, 663–682 (2006)
- Sinz, A.: Investigation of Protein–Protein Interactions in Living Cells by Chemical Crosslinking and Mass Spectrometry. *Anal. Bioanal. Chem.* **397**, 3433–3440 (2010)
- Lee, J.Y.: Mass Spectrometric Analysis of Crosslinking Sites for the Structure of Proteins and Protein Complexes. *Mol. Biosyst.* **4**, 816–823 (2008)
- 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 Crosslinking and Mass Spectrometry. *Mol. Cell. Proteomic.* **8**, 409–420 (2009)
- Rappsilber, J.: The beginning of a beautiful friendship: crosslinking/mass spectrometry and modeling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540 (2011)
- Petrotchenko, E.V., Borchers, C.H.: Crosslinking Combined with Mass Spectrometry for Structural Proteomics. *Mass Spectrom. Rev.* **29**, 862–876 (2010)
- Weber, P.J.A., Beck-Sickinger, A.G.: Comparison of the Photochemical Behavior of Four Different Photoactivatable Probes. *J. Peptide Res.* **49**, 375–383 (1997)
- Hurst, G.B., Lankford, T.K., Kennel, S.J.: Mass Spectrometric Detection of Affinity Purified Crosslinked Peptides. *J. Am. Soc. Mass Spectrom.* **15**, 832–839 (2004)
- Müller, D.R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., Steinmetz, M.O.: Isotope-Tagged Crosslinking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **73**, 1927–1934 (2001)
- Schmidt, A., Kalkhof, S., Ihling, C., Cooper, D.M.F., Sinz, A.: Mapping Protein Interfaces by Chemical Crosslinking and FTICR Mass Spectrometry: Application to a Calmodulin/Adenylyl Cyclase 8 Peptide Complex. *Eur. J. Mass Spectrom.* **11**, 525–534 (2005)
- Schulz, D.M., Kalkhof, S., Schmidt, A., Ihling, C., Stingl, C., Mechtler, K., Zschörnig, O., Sinz, A.: Annexin A2/P11 Interaction: New Insights into Annexin A2 Tetramer Structure by Chemical Crosslinking, High-Resolution Mass Spectrometry, and Computational Modeling. *Proteins* **69**, 254–269 (2007)
- Petrotchenko, E.V., Xiao, K., Cable, J., Chen, Y., Dokholyan, N.V., Borchers, C.H.: BiPS, a Photocleavable, Isotopically Coded, Fluorescent Crosslinker for Structural Proteomics. *Mol. Cell. Proteom.* **8**, 273–286 (2009)
- Soderblom, E.J., Goshe, M.B.: Collision-Induced Dissociative Chemical Crosslinking Reagents and Methodology: Applications to Protein Structural Characterization Using Tandem Mass Spectrometry Analysis. *Anal. Chem.* **78**, 8059–8068 (2006)
- Müller, M.Q., Dreiocker, F., Ihling, C., Schäfer, M., Sinz, A.: Cleavable Crosslinker for Protein Structure Analysis: Reliable Identification of Crosslinked Products by Tandem MS. *Anal. Chem.* **82**, 6958–6968 (2010)
- Dreiocker, F., Müller, M.Q., Sinz, A., Schäfer, M.: Collision-Induced Dissociative Chemical Crosslinking Reagent for Protein Structure Characterization: Applied Edman Chemistry in the Gas Phase. *J. Mass Spectrom.* **45**, 178–189 (2010)
- Kalkhof, S., Haehn, S., Paulsson, M., Smyth, N., Meiler, J., Sinz, A.: Computational Modeling of Laminin N-Terminal Domains Using Sparse Distance Constraints from Disulfide Bonds and Chemical Crosslinking. *Proteins* **78**, 3409–3427 (2010)
- Hussain, S.A., Carafoli, F., Hohenester, E.: Determinants of Laminin Polymerization Revealed by the Structure of the $\alpha 5$ Chain Amino-Terminal Region. *EMBO Rep.* **12**, 276–282 (2011)
- Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J. C., Nilges, M.: Architecture of the RNA Polymerase II-TFIIF Complex Revealed by Crosslinking and Mass Spectrometry. *EMBO J.* **29**, 717–726 (2010)
- Peri, S., Steen, H., Pandey, A.: GPMAW—A Software Tool for Analyzing Proteins and Peptides. *Trends Biochem. Sci.* **26**, 687–689 (2001)
- Koning, L.J., Kasper, P.T., Back, J.W., Nessen, M.A., Vanrobaeys, F., Beeumen, J., Gherardi, E., Koster, C.G., Jong, L.: Computer-Assisted Mass Spectrometric Analysis of Naturally Occurring and Artificially Introduced Crosslinks in Proteins and Protein Complexes. *FEBS J.* **273**, 281–291 (2006)
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., Aebersold, R.: Identification of Crosslinked Peptides from Large Sequence Databases. *Nat. Methods* **5**, 315–318 (2008)
- Du, X.X., Chowdury, 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 Crosslinked Peptides Using Tandem Mass Spectrometry. *J. Protein Res.* **3**, 923–931 (2011)
- Panchaud, A., Singh, P., Shaffer, S.A., Goodlett, D.R.: xComb—A Crosslinked Peptide Database Approach to Protein–Protein Interaction Analysis. *J. Prot. Res.* **5**, 250–2515 (2010)
- Chu, F.X., Baker, P.R., Burlingame, A.L., Chalkley, R.J.: Finding Chimeras: A Bioinformatics Strategy for Identification of Crosslinked Peptides. *Mol. Cell. Proteom.* **9**, 25–31 (2010)
- Mayne, S. L. N.; Patterson, H. G. Bioinformatics Tools for the Structural Elucidation of Multi-Subunit Protein Complexes by Mass Spectrometric Analysis of Protein–Protein Crosslinks. *Brief Bioinform.* **2011**, published online, doi:10.1093/bib/bbq087
- Dimova, K., Kalkhof, S., Pottratz, I., Ihling, C., Rodriguez-Castaneda, F., Liepold, T., Griesinger, C., Brose, N., Sinz, A., Jahn, O.: Structural Insights into the Calmodulin-Munc13 Interaction Obtained by Crosslinking and Mass Spectrometry. *Biochemistry* **48**, 5908–5921 (2009)
- Krauth, F., Ihling, C.H., Rüttinger, H.H., Sinz, A.: Heterobifunctional Isotope-Labeled Amine-Reactive Photo-Crosslinker for Structural

- Investigation of Proteins by Matrix-Assisted Laser Desorption/Ionization Tandem Time-of-Flight and Electrospray Ionization LTQ-Orbitrap Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 2811–2818 (2009)
31. Odenthal, U., Haehn, S., Tunggal, P., Merkl, B., Schomburg, D., Frie, C., Paulsson, M., Smyth, N.: Molecular Analysis of Laminin N-Terminal Domains Mediating Self-Interactions. *J. Biol. Chem.* **279**, 44504–44512 (2004)
32. Kalkhof, S., Witte, K., Ihling, C.H., Müller, M.Q., Keller, M.V., Haehn, S., Smyth, N.R., Paulsson, M., Sinz, A.: A Novel Disulfide Pattern in Laminin-Type Epidermal Growth Factor-Like (LE) Modules of Laminin 1 and 1 Chain. *Biochemistry* **49**, 8359–8366 (2010)
33. Lange, C., Duda, T., Beyermann, M., Sharma, R.K., Koch, K.W.: Regions in Vertebrate Photoreceptor Guanylyl Cyclase ROS-GC1 Involved in Ca(2+)-Dependent Regulation by Guanylyl Cyclase-Activating Protein GCAP-1. *FEBS Lett.* **460**, 27–31 (1999)
34. Schilling, B., Row, R.H., Gibson, B.W., Guo, X., Young, M.M.: MS2 Assign, Automated Assignment and Nomenclature of Tandem Mass Spectra of Chemically Crosslinked Peptides. *J. Am. Soc. Mass Spectrom.* **14**, 834–850 (2003)