HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2019 May 15.

Published in final edited form as:

Anal Chem. 2018 May 15; 90(10): 6028–6034. doi:10.1021/acs.analchem.7b04991.

Mango: A General Tool for CID-cleavable Cross-linked Peptide Identification

Jared P. Mohr, **Poorna Perumalla**, **Juan D. Chavez**, **Jimmy K. Eng**, and **James E. Bruce***

Department of Genome Sciences, University of Washington, Seattle, Washington 98105, United States

Abstract

Chemical cross-linking combined with mass spectrometry provides a method to study protein structures and interactions. The introduction of cleavable bonds in a cross-linker provides an avenue to decouple released peptide masses from their precursor species, greatly simplifying the downstream search, allowing for whole proteome investigations to be performed. Typically, these experiments have been challenging to carry out, often utilizing non-standard methods to fully identify cross-linked peptides. Mango is an open source software tool that extracts precursor masses from chimeric spectra generated using cleavable cross-linkers, greatly simplifying the downstream search. As it is designed to work with chimeric spectra, Mango can be used on traditional high-resolution MS/MS capable mass spectrometers without the need for additional modifications. When paired with a traditional proteomics search engine, Mango can be used to identify several thousand cross-linked peptide pairs searching against the entire *E. coli* proteome. Mango provides an avenue to perform whole proteome cross-linking experiments without specialized instrumentation or access to non-standard methods.

Proteins are specialized molecules designed to carry out innumerable functions in a cell. These functions rely not only on the abundance of individual proteins, but are critically-dependent upon localization, conformation and interactions between assemblies of proteins. Chemical cross-linking combined with mass spectrometry (XL-MS) is emerging to hold great potential for interrogating conformations and interactions that exist within cells[1–4] and elucidating how these networks of interactions can change under different conditions[5,6] XL-MS experiments have been growing in number and appeal over recent years due to the

*Corresponding Author: jimbruce@uw.edu.

CONFLICTS OF INTEREST

The authors declare no competing financial interest.

Supporting Information Available:

Examples of annotated MS2 spectra (Figure S1) from using Mango-Comet. Analysis of searching non-cross-linked samples and using incorrect parameters in mango (Figure S2). Structure and fragments produced by BDP-NHP and how these parameters are used in the Mango-Comet pipeline described (Figure S3). Structure and fragments of DSSO, and how to derive appropriate parameters for Mango's mass relationship and a subsequent search (Figure S4). Effects of varying the mass tolerance parameter on final number of search results (Figure S5). Example of using Msfragger to search Mango's output and how these results compare to Comet (Figure S6). Distribution of e-values for targets and decoys from the Mango-Comet pipeline, and how a false discovery threshold is estimated (Figure S7). Example of searching DSSO, a cross-linker which yields characteristic doublets during fragmentation, cross-linked sample using Mango-Comet (Figure S8). Table of results from the Msfragger data presented in Figure S6 (Table S1). Table of results from the first and second technical replicates of the E. coli dataset present (Table S2 and Table S3). Table of results from the searching a DSSO cross-linked sample (Table S4).

extensive potential applications of cross-linking data[7]. Cross-links within and between proteins provide physical distance constraints that can be used to predict structures. These constraints can be used alone in *ab initio* folding for improved structures[8], or to filter and refine models from homology templates[9,10]. They can also be combined with complementary structural determination tools, such as cryo-EM[11] or x-ray crystallography[12] to potentially produce higher quality structures by restricting the possible solution space. *In vivo* cross-linking experiments uniquely allow for large-scale quantification and monitoring of protein conformational dynamics[6], which can be used to visualize physical effects on conformations and interactions of drug target proteins, as well as off-target effects in cells. Another unique aspect of cross-linking experiments is the ability to identify direct host-pathogen interactions on a structural level[13,14], providing a lens through which to visualize molecular details of pathogenesis of a variety of bacteria and viruses.

Cross-linking experiments in the past have typically utilized stable-backbone cross-linkers, such as BS3 or DSS, which produce a chimeric spectrum of two peptides with known combined mass, but unknown individual peptide masses. With only the precursor mass of the pair to limit the search space, a significant number of peptides or combinations of peptides must be considered and scored with proteome-wide searches. The number of these combinations that will need to be evaluated is quadratic with respect to database size, which makes use of stable-backbone cross-linkers difficult in whole proteome analyses. Despite the complexity of data generated by these cross-linkers, many tools have been developed for evaluating such datasets. These tools implement strategies to reduce the number of peptide pairs that need to be considered for any given spectrum[15–17], which partially mitigate the effect of a large database. A promising development has been the implementation of scoring optimizations that facilitate open window searches to find peptides with unknown modifications[18], which is a prominent issue in searching cross-linking data. Regardless of the methodology employed to search these types of cross-linking experiments, statistical power is lost when each spectrum has a large number of candidates. These problems of quadratic candidate expansion and minimal statistical power are not inherent to all cross-linking experiments, but come about with the use of non-cleavable cross-linkers. Cleavable cross-linkers have provided an alternative to stable backbone cross-linkers by reducing the quadratic search problem to a linear one.

Cleavable cross-linkers such as Protein Interaction Reporter (PIR)[19], DSSO[20], DSBU[21] and others incorporate bonds that can be cleaved predictably in the gas phase. These molecules are engineered such that cross-linker bond cleavage yields two released peptides, which permits the decoupling of the individual cross-linked peptides from their precursor. Once released peptide masses are determined and thus rather than a single quadratic search, two traditional narrow window linear peptide searches can be performed instead. Normal peptide searches are linear with database size, so cleavable cross-linkers perform well in complex biological samples which necessitate a large database. Some commercially available cleavable cross-linkers, such as DSSO[20] and DSBU[21], utilize a characteristic doublet pattern to identify released peptide masses to simplify the search problem and enable the use of a full proteome database. When combined with a tool like XlinkX and an instrument capable of serial fragmentation[22], DSSO can be used to study complex cross-linked lysates. Some

newer instruments now have native support[23] for XlinkX, but high throughput cross-linking experiments remain difficult without access to hardware capable of serial fragmentation.

Hardware requirements and limited instrument control have been a long-standing barrier for optimal use of cleavable cross-linkers on a proteome wide scale. Real-time Analysis of Cross-linked peptide Technology[24] (ReACT) is an XL-MS technique developed for dynamic discovery of protein interaction reporter (PIR) cross-linked peptides[25]. ReACT uses high resolution hardware[26] and software modifications on an LTQ-FT to efficiently identify cross-linked peptides. This method incorporates multiple fragmentation and isolation events to produce spectra of the individual cross-linked peptides at the MS3 level. While this method produces individual peptide spectra compatible with traditional search tools and has produced the majority *in vivo* cross-linked peptides now residing in the database of cross-linked peptides XLinkDB[27,28], it comes at a significant time cost inherent to incorporating multiple scan events. Recently, ReACT has been used to construct libraries that allow for either spectral library searches[29] or PRM-based quantification of previously identified cross-linked peptides, extending some benefits of the methodology to instruments only capable of MS2. A limitation of spectral libraries is that no new cross-linked peptide pairs can be identified during an experiment, which restricts the scope of experiments reliant on these libraries. Extension of PIR identification capabilities to MS2 measurements in the absence of required spectral libraries can significantly extend PIR experiments to many other labs.

To help address the need for improved cross-linked peptide identification capabilities, here we present Mango, an open source search tool for use with CID-cleavable cross-linkers and PIR-technology for the identification of novel cross-links. Mango employs logic similar to ReACT to identify cross-linked peptides, but requires only the capability to produce high resolution MS2 spectra, making the methodology adaptable for use on many commercially available and commonly used mass spectrometers. Unlike other cross-linking search tools, Mango is capable of efficiently analyzing data from cross-linkers with symmetric fragmentation which lack characteristic doublets, even when a full proteome database is used, and it has a standard output format that is compatible with various traditional peptide search tools.

METHODS

E. coli cell culture, cross-linking, and digestion conditions

E. coli (K12) was grown to stationary phase in LB media. *E. coli* cells were pelleted at 1500g for 10 minutes and washed with phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) followed by cross-linking buffer (180 mM Sodium phosphate pH 8.0). The cell pellet was gently resuspended in 500 μL cross-linking buffer, and biotin aspartate proline-N-hydroxyphtalamide (BDP-NHP), synthesized by solid phase synthesis[24], was added to a final concentration of 10 mM. After one hour at room-temperature any remaining reactive cross-linker was quenched with the addition of 1mL of 100 mM ammonium bicarbonate. After quenching, the cells were again pelleted, the cross-linking buffer was removed, and the cells were resuspended in 100 mM ammonium bicarbonate. Urea was added to 8M and then the cells were lysed by sonication using a GE-130 ultrasonic processor. The lysed samples were reduced with 5 mM Tris(2-

carboxyethyl)phosphine for 30 minutes at room temperature, and then alkylated in 10 mM Iodoacetamide for 45 minutes in the dark. The samples were diluted 10-fold with 100 mM ammonium bicarbonate to reduce urea concentration to 0.8M, and the proteins were then digested overnight at 37°C using trypsin (Promega). Digested samples were desalted using C18 sep-pak columns (Waters).

Strong cation exchange fractionation and affinity enrichment

The desalted peptide samples were fractionated by strong cation exchange (SCX) chromatography using an Agilent 1200 HPLC system equipped with a Phenomenex Luna SCX column. A binary linear gradient consisting of buffer A (5 mM KH_2PO_4, pH 2.6, 30% acetonitrile (ACN)) and buffer B (5mM KH_2PO_4, pH 2.6, 30% ACN, 350 mM KCl) was applied at a flow rate of 1.5 mL/min for 97.5 min as follows: 0% B at 0 min, 5% B at 7.5 min, 60% B at 47.5 min, 100% B at 67.5 min, 100% B at 77.5 min, 0% B at 77.51 min to completion. Fractions were taken every 5 min starting at 17.5 min, and fractions were pooled as follows: 1–5, 6–7, 8, 9, 10, 11–14. Fractions 1–5 were not processed any further. The remaining fractions were then reduced to a final volume of 1–2 mL by vacuum centrifugation and pH adjusted to a pH of 8.0 with 1.5M NaOH. After pH adjustment, each sample was incubated for 1 hour with 100 μL of UltraLink monomeric avidin (Thermofisher) with gentle agitation. The avidin matrix was washed 5 times after this incubation period using 3mL aliquots of 100 mM ammonium bicarbonate, and cross-linked peptides were then eluted off the avidin beads by two additions of 500 μL 70% acetonitrile-0.5% formic acid. The resulting eluent was concentrated by vacuum centrifugation.

LC-MS/MS data acquisition description

Peptides recovered from the avidin matrix were analyzed using an EASY-nLC 1000 coupled to a Q Exactive Plus mass spectrometer. Samples were fractionated over a 60 cm x 75 μm inner diameter fused silica analytical column packed with ReproSil-Pur C8 (5 μm diameter, 120 Å pore size particles) by applying a linear gradient from 90% solvent A (0.1% formic acid in water), 10% solvent B (0.1% formic acid in acetonitrile) to 60% solvent A, 40% solvent B over 240 minutes at a flow rate of 300 nL/min.

The mass spectrometer was operated using a data dependent analysis (DDA) method performing one high resolution (70,000 resolving power (RP) at *m/z* 200) MS1 scan from 400–2000 *m/z* followed by MS2 (17,500 RP) on the 20 most abundant ions with a charge between +4 and +8 inclusive detected in the MS1. Parameters for MS2 scans included an automatic gain control target of 50,000 ions, a maximum ion accumulation time of 100 ms, an isolation window of 3.0 m/z, and a normalized collision energy of 30. A dynamic exclusion window of 30 seconds was used to reduce redundant picking of the same parent ion. MS2 spectra were processed using Mango 2017.01 rev. 0 beta 2, whose output was searched using Comet[30] 2017.01 rev. 2.

The same samples were subsequently analyzed using a Water NanoAcquity UPLC coupled to a Thermo Velos Fourier transform ion cyclotron resonance mass spectrometer[26] (Velos-FT). The chromatography gradient employed is identical to the one previously described.

The Velos-FT was operated using the ReACT method[24], where one high resolution (50,000 RP at *m/z* 400) MS1 scan from 400–2000 *m/z* is taken and followed by an MS2 (50,000 RP) on the most abundant ion of at least +4 charge. In real time, if a pair of peaks fulfills the mass relationship (precursor mass = reporter ion mass + peak 1 mass + peak 2 mass) within 20ppm tolerance, then each peak is targeted for 2 additional low resolution MS3 scans in the ion trap. MS3 spectra were searched using Comet 2017.01 rev. 2.

Software description

Mango is an open source tool written in C++ and hosted on GitHub (https://github.com/jpm369/mango) developed for facilitating the identification of cross-linked peptides at the MS2 level. Broadly, Mango is a tool for extracting released peptide masses from MS2 scans of cross-linked peptides that outputs a searchable file with multiple precursors for each scan that correspond to candidate released peptide masses.

Mango takes an mzXML[31] file and a Mango parameters file as its inputs. Mango utilizes Hardklör[32] to preprocess experimental spectra, performing charge deconvolution and de-isotoping to reduce spectral complexity. These reduced spectra are then used to identify pairs of peaks that may have been generated by a cross-linked species. Mango loops through all of the peaks in a deconvoluted MS2 spectrum to identify pairs of peaks that fulfill the mass relationship for a cleavable cross-linker within some user-specified tolerance (Eq. 1). The added requirement of the mass relationship reduces the complexity of the search from quadratic to linear with respect to the number of peptides in a database, analogous to a traditional narrow-window DDA search. This reduction in complexity allows for cross-linking searches to be carried out using an input database containing thousands of proteins without a quadratic increase in search time or loss of statistical power. If a scan contains at least one pair of peaks that fulfills the mass relationship, then each pair of peaks in the spectra that fulfill the mass relationship are written to an ms2 file as a potential precursor mass in place of the MS1 precursor mass isolated. Herein Mango was operated using the following settings: mass_tolerance_relationship = 10.00 ppm, mass_tolerance_peptide = 20.00, reporter_neutral_mass = 751.406080 Da.

The modified ms2 files as output by Mango can be directly searched by Comet (Version 2017.01). Comet searches were performed with the default settings and the following changes: mango_search = 1; variable modifications: 15.9949 M, required modifications: 197.032422 at an internal K. Comet loops through the list of released peptide masses in the precursor header of the file and uses each mass identified by Mango as a precursor mass to perform a narrow window search for that spectrum. Comet scores all fragments that could be generated by each linear precursor mass but does not score fragments containing the second peptide (Figure S1). This results in computing a cross-correlation score, E-value, and all other standard Comet metrics for each precursor mass queried, facilitating established downstream analysis. A custom parameter option in Comet 2017.01, mango_search, directs Comet to also provide a unique identifier for each pair of released peptide masses in a scan to facilitate reassembling the individual linear identifications into a cross-linked identification. This identifier is appended to the spectrum title and contains a pair index and the letter A or B to indicate which identifications should be paired (e.g. 001_A and 001_B).

Corresponding linear peptide identifications are paired together according to their unique identifier, and then the results are filtered to a 1% FDR using a target-decoy based filter at the peptide-spectrum match (PSM) level. A PSM for a pair of cross-linked peptides refers to the pair of peptide assignments to a single spectrum. FDR filtering was performed by first assigning a cross-linked pair of peptides an E-value equal to the worse[33] of the two E-values assigned by Comet. Each spectrum is then assigned its best scoring pair, and then selecting an E-value cut-off to limit final results to contain no more than 1% of the pairs that contained one or two decoy hits. This FDR metric was validated searching cross-linked and non-cross-linked samples using a variety of reporter masses in Mango (Figure S1).

RESULTS

Identifying released peptides masses using Mango

Cleavable cross-linkers provide an avenue by which the individual masses of a pair of cross-link peptides can be determined. In recent years, a number of search tools have emerged that take advantage of the predictable fragmentation products of cleavable cross-linkers to reduce the number of pairs of peptides that must be scored to identify a cross-linked species[22,24,34]. However, some of these methods require instrumentation capable of MS3, or equipped with other non-standard capabilities, to generate suitable data. Mango provides a tool to facilitate whole-proteome, *in vivo* cross-linking experiments using a standard Orbitrap-based or other high resolution mass spectrometer.

Mango makes use of the same mass relationship, originally implemented in ReACT[19], to target candidate cross-linked peptide masses on-the-fly. Instead of utilizing this information in real-time which requires specialized methodology, Mango utilizes this relationship only in post-processing of spectra, allowing a number of CID/HCD-enabled mass spectrometers to be used in conjunction with Mango. This mass relationship is simply a formulation of the conservation of mass, namely that a crosslinked species will fragment to produce two species whose neutral masses sum together with the reporter mass to match selected precursor ion neutral mass with tight tolerance. For a bi-functional cross-linker, this equation can be generalized to:

$$MS1.Precursor = Xlink\,.\,reporter + MS2.mass1 + MS2.mass2 \quad \text{(Eq.1)}$$

Where the reporter is a constant that corresponds to the cross-linker specific mass offset determined by its specific structure and fragmentation products. For example, in BDP-NHP, a biotinylated peptidic cross-linker, this reporter mass stems from the biotinylated region of the molecule with a mass of 751.4106 (Figure S3). In DSSO this reporter mass does not correspond to a physical species, but rather stems from the mass defect resulting from pairs of the light or heavy stump modifications compared to the mass modification associated with the intact cross-linked pair[22] (Figure S4).

Since the reporter mass of a cross-linker can be determined from its structure or measured experimentally and provided as an input to equation 1, the problem of identifying candidate released peptide masses is quadratic with respect to the number of peaks in a spectrum.

Mango loops through all of the peaks in a tandem mass spectrum, generating all combinations of peaks and evaluates their sum, reporting any that fulfill equation 1 within a user-specified mass tolerance. Empirically, approximately 5 times the expected mass accuracy of the instrument appears to be the optimal tolerance to limit the number of candidates considered without losing many true or forward hits (Figure S5). Increasing the mass tolerance to higher values results in a large number of spectra having many candidate pairs (Figure 1). The increased number of candidates to score increases downstream search time and decreases statistical power, which can lead to an overall lower number of identifications after FDR filtering (Figure S5).

A novel feature of Mango among cross-linking analysis tools is that it does not contain a native peptide search engine, and instead outputs an .ms2 file encoding all pairs of masses that fulfill equation 1 as a precursor mass for the scan from which they were extracted from. A .mgf output file is also available that lists each precursor mass query as a separate scan for search tools that will not handle multiple precursor masses per spectrum, allowing alternative search engines to be used (Figure S6, Table S1). This will facilitate the integration of Mango into existing pipelines without the need for significant change to existing post-processing and analysis methods. Mango can be used as a pre-processing step before peptide scoring, and the appropriate peptides can be paired together at the end of all post-processing steps to identify their progenitor cross-linked species.

Whole proteome cross-linking at the MS2 level

Mango and Comet were used to identify cross-links generated by cross-linking *E. coli in vivo* with BDP-NHP. BDP-NHP is a biotinylated-peptidic cross-linker capable of penetrating cell membranes[35], allowing it to preserve protein-protein interactions in their native context during *in vivo* cross-linking experiments. For this strategy, it is necessary to produce fragment ions for both the intact released peptides, as well backbone fragment ions corresponding to the amino acid sequences of the released peptides (Figure 2). The intact fragments are necessary for Mango to be able to identify candidate masses that can effectively constrain the search, while peptide backbone fragment ions are necessary to assign primary sequences for the peptides.

Comet was used to search the output of Mango from an *in vivo* cross-linking experiment. After Mango has extracted candidate released peptide masses, the subsequent database search is identical to a normal peptide search, which is linear with respect to database size. Consequently this pipeline can be used to search for cross-links from the whole *E. coli* proteome (4309 target and 4309 decoy proteins).

Across single injections of five SCX fractions, Mango and Comet were able to assign 4170 PSMs (Table S2), mapping to 2334 non-redundant peptide pairs at less than or equal to 1% FDR (Figure S7). These identifications are not evenly distributed across all fractions, but rather concentrated in the later fractions (8–14) which are enriched for highly charged species by SCX (Figure 3A). Beyond being able to identify thousands of cross-linked peptide pairs in a single sample, these can also be found with reproducibility similar to traditional DDA runs with duplicate analyses (Table S2, Table S3) yielding roughly 75% overlap due to the stochastic nature of DDA sampling (Figure 3B).

While Mango as presented is applied to a PIR cross-linked sample, it is agnostic to the cross-linker and can work with any cross-linker for which a mass relationship (Equation 1) can be formulated. To test this, Mango was used to identify cross-links from published DSSO cross-linked HeLa lysate data[22] (Figure S8, Table S4).

Comparison with ReACT

Mango's performance was compared directly to ReACT, a dynamic MS3-based method for identifying cross-links. In contrast to Mango, ReACT is able to dynamically target candidate released peptides masses for MS3[24] increasing the certainty in fragment assignment by avoiding chimeric spectra. However, ReACT requires 2 high resolution scans and 4 low resolution scans totaling to approximately 3 seconds to fully query a cross-linked peptide pair, which limits the depth of coverage and reproducibility in very complex samples. While MS2 chimeric spectra containing fragments from both peptide sequences are more difficult to score than independent MS3 spectra of peptides generated by ReACT, each cross-linked pair can be investigated more quickly by eliminating the MS3 requirement entirely and acquiring lower resolution MS2 spectra. The combination of both these measures produces nearly an order of magnitude more investigations per analytical run, at the cost of increased ambiguity in fragment assignment associated with chimeric spectra.

The same fractions of the *E. coli* sample analyzed in the previous section were analyzed using ReACT on the Velos-FT and compared to the results achieved by Mango. This data set consisted of 273886 HCD-MS2 spectra from the Q-Exactive Plus for Mango, and 30916 CID-MS3 spectra from the Velos-FT generated by ReACT. In each fraction, Mango was able to identify at least 50% more non-redundant peptide pairs as compared to ReACT (Figure 4A). Overall, Mango identifies 2334 non-redundant peptide pairs compared to ReACT's 1135, which are mapped to 1974 and 1002 unique lysine-lysine site interactions respectively. However, a greater fraction of paired ReACT spectra are successfully assigned a high-confidence PSM, indicating that the identification improvements presented by Mango can be largely attributed to having a larger number of queries compared to ReACT.

Comparing the overlap of non-redundant PSMs from the two datasets reveals that each method identifies exclusive sets of cross-linked peptide pairs (Figure 4B), despite the overall increase in results achieved with Mango. The lysine-lysine site level shows a similar trend as the peptide level, with the two methods sharing 675 unique interactions. While some portion of these missed identifications can be explained by the stochastic sampling employed, differences in the fragmentation strategies between the two methods may further explain some missed identifications. ReACT utilizes multiple isolation and fragmentation events to produce fragments from one released peptide at a time with a traceable relationship to the cross-linked peptide pair first isolated from the MS1. Mango instead relies on a single isolation and fragmentation event to produce fragments that correspond to both species. Intuitively, one would expect that the characteristics of cross-linked species that provide good fragmentation for Mango may not be identical to those that produce good spectra in ReACT. It is likely that some pairs of peptides are difficult to identify from chimeric spectra due to one peptide fragmenting significantly better than its partner[33], a problem which is alleviated in ReACT by isolating and fragmenting each peptide individually. Ultimately this

suggests complementarity between Mango and MS3-based methods, where there is a trade-off between the increased sampling rate enabled by producing chimeric spectra versus the time-intensive independent isolation and fragmentation of individual peptides offered by a method like ReACT.

CONCLUSION

XL-MS provides a unique way of interrogating protein-protein interactions in their native cellular context. Knowledge of the identities of these interacting proteins and the structures of their assemblies can provide insight into a variety of biological systems. Here we describe Mango as a high throughput solution for *in vivo* or *in vitro* cross-linking experiments that requires accurate mass MS/MS capabilities that are widely available in many labs.

Mango has been employed successfully in conjunction with Comet for identifying a large number of cross-linked peptides in a complex *in vivo* cross-linked *E. coli* sample using a Q-Exactive Plus for data acquisition. These results compare favorably to those acquired using an MS3-based cross-linked identification strategy, but the two data sets also show complementarity. The depth of sampling achievable by Mango could likely be improved with dynamic fragmentation whereby each species is fragmented at multiple energies, to expand the cross-linked peptide space Mango is able to sample. While Mango is presented here using BDP-NHP and a Q-Exactive Plus, the software is agnostic to the cross-linker and instrument used, requiring only that a CID-cleavable cross-linker be used on an instrument able to acquire high-resolution MS2 spectra.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge all members of the Bruce lab and UWPR for productive discussions during development. We would also like to acknowledge the following funding sources: NIH 5R01GM08668 and 2R01GM097112.

References

1. Zheng C, Yang L, Hoopmann MR, Eng JK, Tang X, Weisbrod CR, Bruce JE. Mol Cell Proteomics. 2011; 10:M110.006841.
2. Navare AT, Chavez JD, Zheng C, Weisbrod CR, Eng JK, Siehnel R, Singh PK, Manoil C, Bruce JE. Structure. 2015; 23:762–773. [PubMed: 25800553]
3. Wu X, Chavez JD, Schweppe DK, Zheng C, Weisbrod CR, Eng JK, Murali A, Lee SA, Ramage E, Gallagher LA, et al. Nat Commun. 2016; 7:1–14.
4. Guerrero C, Tagwerker C, Kaiser P, Huang L. Mol Cell Proteomics. 2006; 5:366–378. [PubMed: 16284124]
5. Chavez JD, Schweppe DK, Eng JK, Zheng C, Taipale A, Zhang Y, Takara K, Bruce JE. Nat Commun. 2015; 6:1–12.
6. Chavez JD, Schweppe DK, Eng JK, Bruce JE. Cell Chem Biol. 2016; 23:716–726. [PubMed: 27341434]
7. Leitner A, Faini M, Stengel F, Aebersold R. Trends Biochem Sci. 2016; 41:20–32. [PubMed: 26654279]
8. Yang J, Zhang Y. Nucleic Acids Res. 2015; 43:W174–W181. [PubMed: 25883148]

9. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. Proc Natl Acad Sci. 2000; 97:5802–5806. [PubMed: 10811876]
10. Rozbesky D, Sovova Z, Marcoux J, Man P, Ettrich R, Robinson CV, Novak P. Anal Chem. 2013; 85:1597–1604. [PubMed: 23249299]
11. Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, Chait BT. Mol Cell Proteomics. 2014; 13:2927–2943. [PubMed: 25161197]
12. Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimermančič P, Boehringer D, Sali A, Aebersold R, et al. Cell. 2014; 158:1123–1135. [PubMed: 25171412]
13. Deblasio SL, Chavez JD, Alexander MM, Ramsey J, Eng JK, Mahoney J, Gray SM, Bruce JE, Cilia M. J Virol. 2015; 90:JVI.01706-15.
14. Schweppe DK, Harding C, Chavez JD, Wu X, Ramage E, Singh PK, Manoil C, Bruce JE. Chem Biol. 2015; 22:1521–1530. [PubMed: 26548613]
15. Hoopmann MR, Zelter A, Johnson RS, Riffle M, Maccoss MJ, Davis TN, Moritz RL. J Proteome Res J Proteome Res. May.2015 1:2190–2198.
16. McIlwain S, Draghicescu P, Singh P, Goodlett DR, Noble WS. J Proteome Res. 2010; 9:2488–2495. [PubMed: 20349954]
17. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, et al. Nat Methods. 2012; 9:904–906. [PubMed: 22772728]
18. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. Nat Methods. 2017; 14:513–520. [PubMed: 28394336]
19. Tang X, Munske GR, Siems WF, Bruce JE. Anal Chem. 2005; 77:311–318. [PubMed: 15623310]
20. Kao A, Chiu C, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L. Mol Cell Proteomics. 2011; 10:M110.002212.
21. Müller MQ, Dreiocker F, Ihling CH, Schäfer M, Sinz A. Anal Chem. 2010; 82:6958–6968. [PubMed: 20704385]
22. Liu F, Rijkers DTS, Post H, Heck AJR. Nat Methods. 2015; 12:1179–1184. [PubMed: 26414014]
23. Liu F, Lössl P, Scheltema R, Viner R, Heck AJR. Nat Commun. 2017; 8:15473. [PubMed: 28524877]
24. Weisbrod CR, Chavez JD, Eng JK, Yang L, Zheng C, Bruce JE. J Proteome Res. 2013; 12:1569–1579. [PubMed: 23413883]
25. Tang X, Bruce JE. Mol Biosyst. 2010; 6:939–947. [PubMed: 20485738]
26. Weisbrod CR, Hoopmann MR, Senko MW, Bruce JE. J Proteomics. 2013; 88:109–119. [PubMed: 23590889]
27. Zheng C, Weisbrod CR, Chavez JD, Eng JK, Sharma V, Wu X, Bruce JE. J Proteome Res. 2013; 12:1989–1995. [PubMed: 23413830]
28. Schweppe DK, Zheng C, Chavez JD, Navare AT, Wu X, Eng JK, Bruce JE. Bioinformatics. 2016; 32:2716–2718. [PubMed: 27153666]
29. Schweppe DK, Chavez JD, Navare AT, Wu X, Ruiz B, Eng JK, Lam H, Bruce JE. J Proteome Res. 2016; 15:1725–1731. [PubMed: 27089058]
30. Eng JK, Jahan TA, Hoopmann MR. Proteomics. 2013; 13:22–24. [PubMed: 23148064]
31. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, et al. Nat Biotechnol. 2004; 22:1459–1466. [PubMed: 15529173]
32. Hoopmann MR, Finney GL, Maccoss MJ. 2008; 79:5620–5632.
33. Trnka MJ, Baker PR, Robinson PJJ, Burlingame AL, Chalkley RJ. Mol Cell Proteomics. 2014; 13:420–434. [PubMed: 24335475]
34. Götze M, Pettelkau J, Fritzsche R, Ihling CH, Schäfer M, Sinz A. J Am Soc Mass Spectrom. 2014; 26:83–97. [PubMed: 25261217]
35. Chavez JD, Weisbrod CR, Zheng C, Eng JK, Bruce JE. Mol Cell Proteomics. 2013; 12:1451–1467. [PubMed: 23354917]

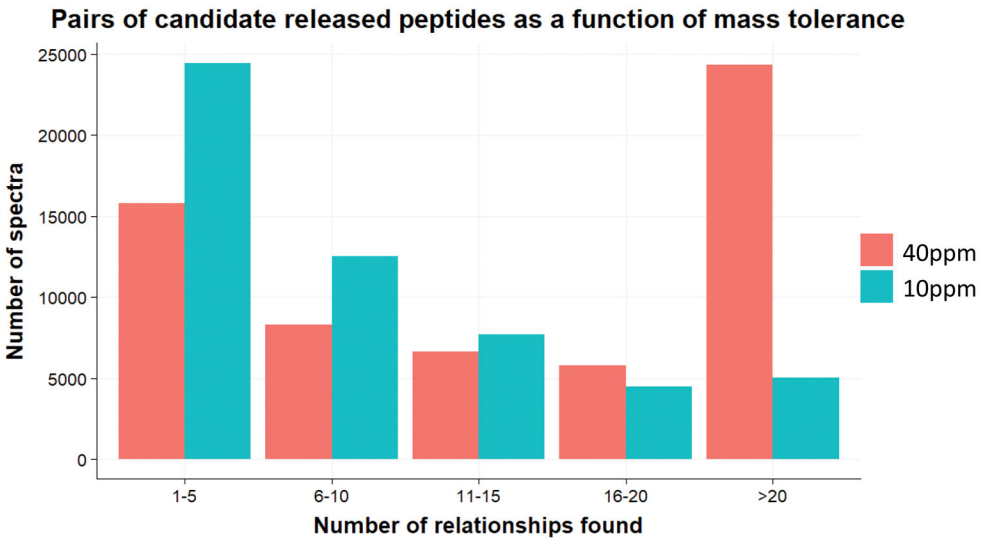

Figure 1.

The effect of tolerance on the number of putative released peptide pairs found using Mango across an entire dataset. High mass accuracy improves the results obtained with Mango, as it allows for a reduction in the tolerance of the mass relationship utilized for identifying candidate released peptides. A tighter mass tolerance reduces the number of candidates that must be scored for any given spectra in a downstream search, which improves both search time and statistical power.

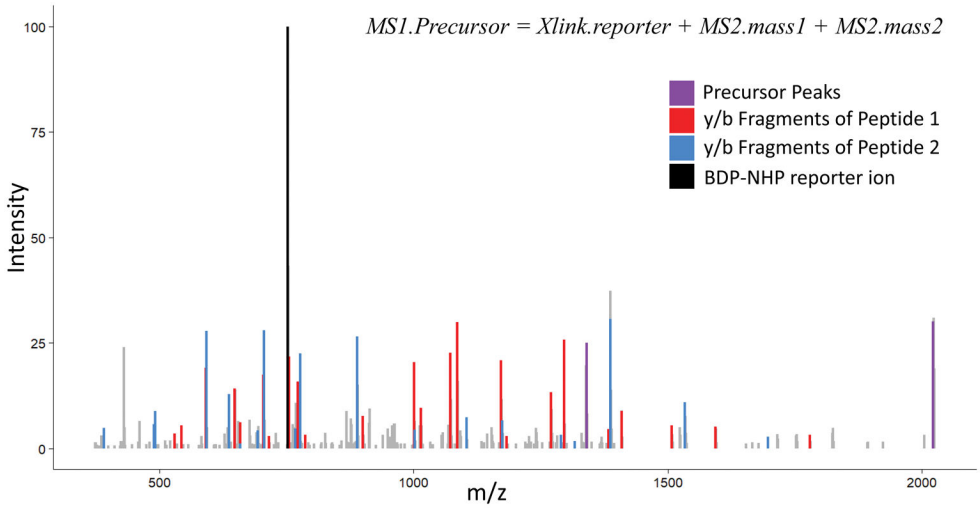

Figure 2.

Mango identifies pairs of peaks (purple) in an MS2 spectrum that fulfill a mass relationship within some tolerance. These masses can then be used by a peptide search engine, such as Comet, to perform multiple narrow window searches on the same spectrum, twice for each pair of masses output by Mango. These sequential narrow window searches yield a peptide identification for each mass in the pair based off observed fragments for each peptide (red and blue). These identifications can then be paired using a unique identifier assigned by Mango to identify the initial cross-linked species isolated.

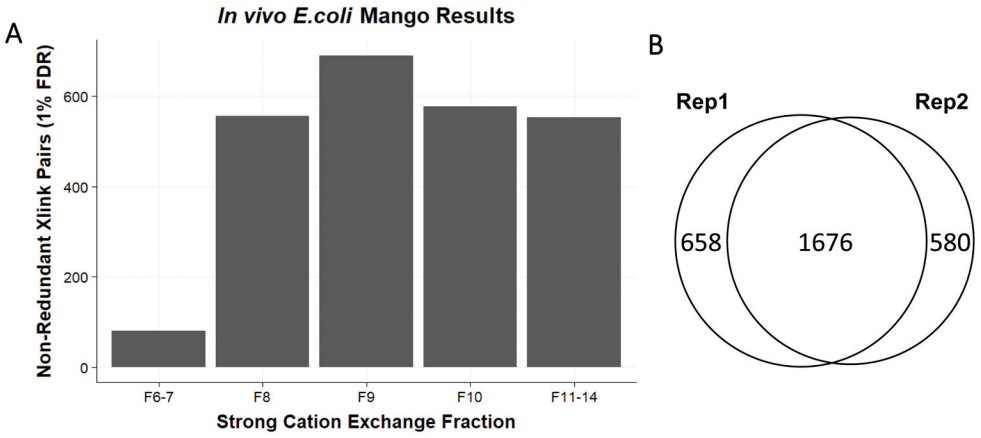

Figure 3.

Summary of Mango analysis of a cross-linked E. coli sample. (A) Number of non-redundant IDs found in each SCX fraction analyzed after FDR filtering. (B) Overlap of non-redundant peptide pairs identified post-FDR filtering between technical replicates of the same 5 SCX fractions.

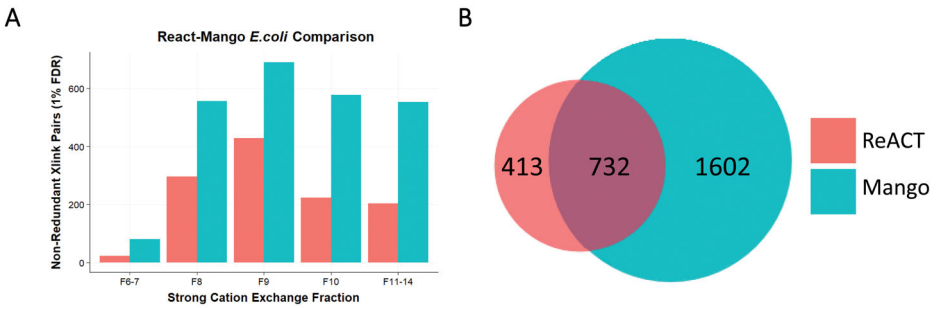

Figure 4.

Summary of comparison between Mango and ReACT for a whole proteome in-vivo cross-linking experiment. (A) The per fraction comparison between Mango and ReACT. (B) The overall overlap of non-redundant peptides post-FDR filtering identified by Mango and ReACT.