HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2017 May 06.

Published in final edited form as:

J Proteome Res. 2016 May 6; 15(5): 1725–1731. doi:10.1021/acs.jproteome.6b00014.

Spectral library searching to identify cross-linked peptides

Devin K Schweppe[1], **Juan D Chavez**[1], **Arti T Navare**[1], **Xia Wu**[1], **Bianca Ruiz**[1], **Jimmy K Eng**[2], **Henry Lam**[3], and **James E Bruce**[1,*]

[1]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA

[2]University of Washington Proteomics Resource, University of Washington School of Medicine, Seattle, WA 98195, USA

[3]Department of Chemical and Biomolcular Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Abstract

Methods harnessing protein cross-linking and mass spectrometry (XL-MS) offer high-throughput means to identify protein-protein interactions (PPIs) and structural interfaces of protein complexes. Yet, specialized data dependent methods and search algorithms are often required to confidently assign peptide identifications to spectra. To improve the efficiency of matching high confidence spectra we developed a spectral library based approach to search cross-linked peptide data derived from Protein Interaction Reporter (PIR) methods using the spectral library search algorithm, SpectraST. Spectral library matching of cross-linked peptide data from query spectra increased the absolute number of confident peptide relationships matched to spectra, and thereby number of protein-protein interactions identified. By matching library spectra from bona fide, previously established PIR-cross-linked peptide relationships, spectral library searching reduces the need for continued, complex mass spectrometric methods to identify peptide relationships, increases coverage of relationship identifications and improves the accessibility of XL-MS technologies.

*Contact: James E. Bruce, Ph.D., Department of Genome Sciences, University of Washington School of Medicine, 850 Republican Street, Brotman Building, room 154, Seattle, WA 98109, Tel: (206) 543-0220, Fax: (206) 616-0008, jimbruce@u.washington.edu.

Supporting Information

Supporting Tables

Table S-1. Spectra-spectra match information for the 157 relationships SpectraST identified from the library-query that were not identified by ReACT (see Figure S-3 for reciprocal plots).

Supporting Figures

Figure S-1. Characterization of FDR estimation methods.

Figure S-2. Detailed analysis based on individual cases of false negatives/false positives of those identifications that SpectraST searching was unable to recapitulate and description of potential causes.

Figure S-3. Reciprocal plots for the 157 relationships SpectraST identified from the library-query that were not identified by ReACT (see Table S-1 for relationship information).

Figure S-4. NCE optimization and example spectra for QE+ query searches. Three relationships with their corresponding reciprocal plots and precursor m/z's are shown.

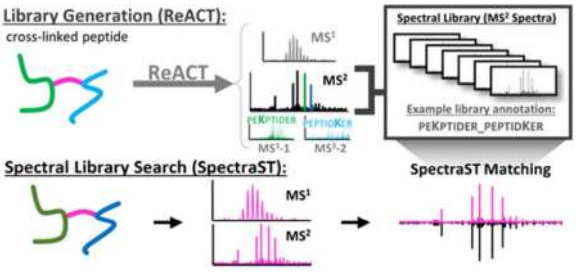
Figure S-5. Example spectra for in vivo *Klebsiella* relationships queried against cryo-crosslinking library. Three relationships with their corresponding reciprocal plots and precursor m/z's are shown.

Figure S-6. SpectraST identifications for non-cleavable cross-linked protein analysis based on public data.

Figure S-7. SpectraST search instructions.

Figure S-8. Comparison of decoy-free binning versus peak-to-peak selection and PeptideProphet distributions.

Graphical Abstract



Keywords

Spectral library; XL-MS; protein cross-linking

Technical Note

Protein interactions underlie nearly all biological processes. Therefore to fully understand these processes requires the large-scale determination of protein interactions. This need has driven the rapid development of techniques to explore protein interactions, including the use of protein cross-linking in conjunction with mass spectrometry (XL-MS). XL-MS combines the ability to determine structural information within individual proteins and to determine interactions between multiple proteins within complex biological samples[1]. Mass spectrometric analysis of cross-linked peptides often require specialized methods to identify cross-linked peptide pairs from complex samples and in many cases the deconvolution of chimeric peptide fragmentation spectra consisting of peptide fragments from both cross-linked peptides[2]. As to the latter challenge, XL-MS pipelines have effectively employed *in silico* methods for spectral deconvolution[2] or exploited the differential physical and chemical properties of cross-linked peptides (e.g. higher mass and charge compared to non-cross-linked peptides) [2–4]. These methods have been highly effective for generating XL-MS data across diverse species and biological mixtures, yet they still require specialized data analysis tools or instrumentation to generate interaction data[2,5]. In an effort to obviate this need, we tested the utility of spectral library searching to identify cross-linked peptide pairs, and thereby protein-protein interactions.

Spectral library searching has the potential to reduce the need for continued complex data analysis or the availability of specialized instrumentation. This is accomplished due to two factors: first, spectral library search algorithms, such as SpectraST[6] and Bibliospec[7], are freely available, highly cited, and frequently implemented platforms for the analysis of large-scale proteomics data beyond XL-MS data analysis. Second, once the spectral library is created from previously established data, the next step of matching newly obtained data is a straightforward process of matching query and library fragmentation patterns without the need for deconvolution of chimeric spectra or the need for custom MS^3-based methods.

We tested the use of spectral library matching to analyze XL-MS spectra from Protein Interaction Reporter (PIR) data of cross-linked proteins from living bacterial cells (*Acinetobacter baumannii*) and originally analyzed with Real-time Analysis of Cross-linked

peptide Technology (ReACT, Figure 1)[8]. ReACT fragmentation spectra (MS^2) derived from the dissociation of the cleavable PIR cross-linker and the subsequent peptide annotations from SEQUEST searches were used to generate a spectral library for *A. baumannii* (referred to as SLAb, Figure 1A) [9]. The library was generated using SpectraST[6]. SLAb was created from a dataset of 37100 total relationship identifications and consisted of the 1782 unique peptide-peptide relationship identifications (average identifications per relationship = 20) from 51 ReACT analyses. Redundant identifications within the library were filtered based on the relationship mass error (parts per million [ppm]) calculated during ReACT analysis, and this ppm-filtered set of unique peptide-peptide relationships was retained in the final library (n = 1782 spectra).

We used SLAb to test the functionality of spectral library-based searching of XL-MS data by searching query spectra from (1) a mass spectral file used to create the spectral library (library-query) and (2) a new XL-MS dataset from a biological replicate sample not included in the library generation (replicate-query) (Figure 1C, Figure 2). Though SpectraST spectral-spectral matches (SSMs) with a dot product score of 0.5 or greater have previously been considered reliable for peptide SSMs[10], we measured SSM false discovery rates (FDRs) by both decoy-free[11] and precursor swap methods[12] (Figure S-1A). Total correct SSMs from both decoy-free and precursor-swap FDR estimation, 548 and 476 total SSMs, respectively, providing confidence that irrespective of filtering SpectraST can improve SSM sensitivity. The authors note that SSM filtering by either decoy-free or precursor-swap methods was preferred over a simple dot-product threshold, as they offer more conservative estimates of correct hits. To simplify later experiments, we used decoy-free filtering to an FDR of 5% throughout this study (Figure S-1). The original ReACT-analysis resulted in MS^3-based identification of 290 non-redundant cross-linked-peptide pairs from the library-query (Figure 2A). Strikingly, from the same file we were able to identify 419 peptide-peptide relationships through spectral library searching, a 45% increase in sensitivity (Figure S-1B). In total, SLAb searching and ReACT identified 447 non-redundant relationships (Figure 2A). Of the original 290 cross-linked peptide pair identifications, spectral library searching matched 262 (>90% of the relationships). The remaining 28 relationships were a combination of false-negatives and false-positives and are characterized in more detail in Figure S-2. Of the 157 relationships identified by SpectraST that were not identified by ReACT, 55/157 did not trigger MS^3 fragmentation (12% of the SpectraST relationships, Figure S-3, Table S-1). The remaining relationships (102/157 relationships) triggered MS^3 fragmentation but did not pass post-acquisition filtering, most often due to poor MS^3 spectral quality. As expected, when we manually validated individual relationships we found that query and library spectra were highly similar (Figure 2A).

For the replicate-query search, the combination of ReACT and the SLAb search generated 366 unique cross-linked peptide pairs. By comparison, the original ReACT search identified 276 cross-linked peptide pairs (Figure 2B). Again, manual verification of spectral matching showed that the SLAb spectra and query spectra were highly similar (Figure 2B). Of note this biological replicate sample was originally generated ~2 years previous to the spectra used to generate the library, and, while the same cross-linker was used, sample preparation of the replicate query was performed under different conditions (i.e. buffer salt concentrations, incubation time for cross-linking, and gradients for strong cation exchange

[SCX] fractionation). Therefore, when compared to the library-query, the intersection of ReACT and SLAb results for the replicate-query (n=152) was not as comprehensive, yet SLAb was able to identify 55% of the original 276 peptide-relationships, commensurate with previously determined identification overlap between pairs of technical replicates[13]. Moreover, with the continued incorporation of new analyses for samples of the same species, we can improve coverage of spectral identifications and further improve the number of relationships identified per run.

Spectral library searching of cross-linked data offers a new use for the growing number of XL-MS datasets. This new use has the potential for broad utility in the identification of structural features of proteins and protein-protein interactions. First, we tested the utility of searching query data generated on a QE+ Orbitrap instrument against a library of spectra generated on a Velos-FT-ICR instrument. This is particularly important to expand the use of ReACT-based identification of cross-linked interactions to cross-link data obtained from high resolution mass analyzers other than ReACT-compatible platforms (i.e. Velos-FT-ICR) (Figure 3, Figure S-S4). We note two advantages to the identification of cross-linked peptide pairs on the QE+: (1) the scan speed of the QE+ allows for more rapid acquisition of high resolution spectra as compared to the Velos-FT-ICR, and (2) methods harnessing spectral libraries to identify interactions would not require acquisition of MS3 fragmentation spectra, reducing the instrument duty cycle

Cross-linked proteins from *Saccharomyces cerevisiae* (strain BY4742) were analyzed either by ReACT analysis on the Velos-FT-ICR or by top-20 MS/MS data dependent scans on a QE+ mass spectrometer. We observed a 42% overlap between spectral library searches and ReACT searches. As expected for this proof-of-principle study, we observed discrepancies between QE+ and FT peak patterns, such as large precursor ion peaks in the QE+ spectra indicating incomplete dissociation (Figure 3A, pink arrowhead) and ions of different charge states (Figure 3A, green arrowheads). While the QE+ methods employed were optimized for dissociation of the cleavable cross-linker for a standard peptide (Figure S-4A), future work may benefit from cross-linker specific tuning of fragmentation conditions to reduce these disparities with more complete dissociation of precursor ions and release of cross-linked peptides and reporter ions. These improvements would thereby increase dot product scores and the number of confident relationships identified.

Second, we used spectral libraries to identify the intersection of cross-linked cryo-milled cell lysate interactions and *in vivo* cross linked interactions. Recent studies have taken advantage of cryo-milling cell lysates and protein-cross-linking to determine protein-protein interactions[14–17]. These studies enable the stabilization of protein complexes from the general cellular milieu and can improve both the sensitivity and reproducibility of protein-protein interaction detection[16]. We generated a library from peptide relationships derived from cross-linked cryo-milled *Klebsiella pneumoniae* lysate (cryo-library, n=2046 non-redundant peptide relationships, Figure 4A). We then searched *in vivo* cross-linked dataset of *K. pneumoniae* using against this library (Figure 4A, Figure S-5). Originally, ReACT had identified 599 peptide-peptide relationships from *in vivo* cross-linking of *K. pneumonia.* When combined with the ReACT identifications, spectral matching with the cryo-library at

an FDR of 5% increased the total number of confident identifications to 939 non-redundant peptide relationships, a 57% increase in confident identifications (Figure 4B).

With recent efforts to quantify interactions between proteins, spectral library searching offers a means to expand coverage and thereby increase the number of quantifiable cross-linked peptide pairs and edgotype data[18]. We have shown that spectral library searching on a QE+ using canonical MS/MS analyses can allow confident identification of hundreds of cross-linked relationships. Although spectral libraries do not obviate the need for ReACT analyses using PIR technology, they do allow research groups interested in large-scale protein interactions the ability to perform such studies without special instrumentation or instrument methods. Because spectral library searching is not limited to our platform, it should be adaptable to other high-throughput cross-linking pipelines that take advantage of MS-cleavable cross-linkers, such as DSSO or DSBSO[19,20], and potentially extended to non-cleavable cross-linkers, such as DSS or BS3. We demonstrated increased sensitivity for non-cleavable cross-linkers similar to those demonstrated above for cleavable cross-linkers (>55%) based on publicly available data (Figure S-6)[21]. For those groups conducting studies with highly specialized instrumentation and software, the use of spectral libraries to identify peptide-peptide relationships could improve sensitivity, allowing for more in-depth analyses of protein interactions.

Moving forward, we note two caveats: (1) spectral library searching relies on high-quality input spectra and confidently matched peptide-relationships and (2) library-based searching remains specific to the individual cross-linker used (i.e. query spectra from BDP-NHP cross-linked proteins would need to be searched using library spectra from BDP-NHP cross-linked proteins). Improvements to library generation (such as inclusion of library spectra from multiple precursor charge states) could be used to increase the number of spectral library identifications even further (rescue the examples set forth in Figure S-2B). Furthermore, we believe FDR estimation, e.g. decoy-free or precursor-swap, rather than a dot product cutoff should be used to establish confidence in XL-MS data searched with spectral libraries. Finally, continued generation of cross-linked spectral libraries will aid the accessibility of protein interaction network generation from complex biological samples using protein cross-linking and proteomics.

Experimental Procedures

Cell Culture, Cell Lysis and Sample Preparation for LC-MS/MS

Bacterial and yeast cells were grown to stationary phase at 37°C in either LB broth or YPD media, respectively. Cells were then pelleted and washed twice with PBS, followed by resuspension in cross-linking buffer (0.17M Na_2HPO_4, pH 8.0). Living cells were then cross-linked by incubation with Biotin-Aspartate Proline-PIR *n*-hydroxyphthalimide[8] (BDP-NHP) for 1 hour at room temperature. Excess unreacted cross-linker was quenched with addition of 10mM ammonium bicarbonate, cross-linked cells were then washed with cross-linking buffer, and frozen until use at −80°C.

Cells were lysed by cryo-milling (3x 30Hz, 1 min) with either 8M urea (*S. cerevisiae* and *A. baumannii*) or with 4% SDS buffered with 100mM Tris-Cl pH 8.0, 150mM NaCl with

subsequent detergent removal by buffer exchange. Briefly, lysates were reduced with 10mM dithiothreitol at 55°C for 30 minutes (if urea used) or at 95°C for 10 minutes (if SDS used), then alkylated with 15mM iodoacetamide. SDS lysates were buffer exchanged to 8M urea buffer using the max volume of 30kDa molecular weight cutoff spin filters (15mL capacity, Millipore) until no detectable SDS was present. Protein lysates were resuspended in ~1mL of 8M urea buffer and subsequently diluted to <1M urea with 50mM Tris-Cl pH 8.0 and digested with trypsin (1:200 w/w) overnight at 37°C. After digestion, peptide lysates were acidified by addition of 0.3% trifluoracetic acid and cleared by centrifugation. Peptides were desalted (manufacturers protocol for C18-SepPak, Waters) to be fractionated by SCX chromatography on a Phenomenex Luna SCX column mounted on an Agilent 1200 HPLC from 100% buffer A (30% ACN, 7mM KH_2PO_4, pH 2.5) to 30% buffer B (30% ACN, 7mM KH_2PO_4, 350mM KCl, pH 2.5). SCX fractions were subsequently enriched for biotin containing cross-linked peptide-peptide relationships with monomeric avidin beads (UltraLink, Pierce) and eluted with 70% ACN, 1% trifluoroacetic acid. Eluted cross-linked peptides were then dried by vacuum centrifugation and frozen at −80°C until use.

Cryo-milled Protein Cross-linking for *Klebsiella*

Bacteria were grown to stationary phase at 37°C in LB broth. Cells were then pelleted and washed twice with PBS, followed by resuspension in cross-linking buffer. Cells were then immediately lysed by cryo-milling (3× 30Hz, 1 min) in cross-linking buffer, the lysates were cleared by centrifugation (15 min, 4°C, 13,200 rpm), and BDP-NHP was added to the supernatant for 2 hours at 4°C to cross-link protein interactions. Protein lysates were then denatured and reduced first with 4% SDS, 10mM dithiothreitol, 0.1M Tris-Cl pH 8.0 followed by reduction and alkylation, and SDS was then replaced by 8M urea via on-filter buffer exchange (30kDa molecular weight cutoff filter, Millipore). Once no SDS was detectable, protein lysate was recovered from the filter with 0.1mM ammonium bicarbonate pH 8.0 followed by trypsin digestion overnight at 37°C (1:200 w/w). All subsequent sample preparation steps were as described above.

LC-MS/MS/MS, LC-MS/MS and Data Analysis

Enriched cross-linked peptides eluted from avidin capture were resuspended (5% ACN/2% formic acid) and injected onto an in-house pulled 45cm C-8 column (Magic, 200A, 5um) run on 4 hour gradients as follows: MS-buffer A (0.1% formic acid in water) and MS-buffer B (0.1% formic acid in acetonitrile [ACN]); 0–1min 2–10% MS-buffer B, 1–241min 10–40% MS-buffer B, 241–261min 40–80% MS-buffer B, 261–281min 80–2% MS-buffer B; eluted peptides were analyzed on an LTQ-Velos-FT-ICR. Spectra were searched using SEQUEST (databases consisted of the full *S. cerevisiae* [Uniprot], *A. baumannii* [Uniprot] and K. *pneumoniae* [PGAT[22]] proteomes) or SpectraST (see below). Cross-linked peptides were fragmented using a ReACT (Real-time Analysis for Cross-linked peptide Technology)[8] data-dependent method on the LTQ-Velos-FT to identify cross-linked peptide-relationships. Briefly, high-charge state precursor ions (MS^1, $z \geq 4$) were isolated. MS^1 data were collected at R = 50000 at 400 m/z, AGC = 500000. Precursor ions were fragmented at low energy (activation Q = 0.25, normalized collision energy = 25.0, isolation width = 3.0 m/z, activation time = 10 ms, excluding precursor charges 1,2,3), releasing individual cross-linked peptides and a reporter ion (m/z = 752.41) observed in MS^2 spectra (R = 50000, AGC

= 200000). On-the-fly, PIR relationships that summed to the MS^1 precursor mass (precursor mass = reporter mass + $peptide^1$ mass + $peptide^2$ mass), within 25ppm, were selected for MS^3 fragmentation (activation Q = 0.25, normalized collision energy = 35.0, isolation width = 3.0 m/z, activation time = 10 ms, AGC = 50000) to generate spectra for SEQUEST spectral matching. For data comparison of SpectraST libraries used to search QE+ data, the same sample originally analyzed on the LTQ-Velos-FT were injected onto an in-house pulled 45cm C-18 column (Magic, 200A, 5um) run on 4 hour gradients as follows: MS-buffer A (0.1% formic acid in water) and MS-buffer B (0.1% formic acid in acetonitrile [ACN]); 0–1min 2–10% MS-buffer B, 1–241min 10–40% MS-buffer B, 241–261min 40–80% MS-buffer B, 261–281min 80–2% MS-buffer B; eluted peptides were analyzed on the QE+. Top-20 QE+ methods settings were: Full-MS Resolution = 70000, Full-MS AGC target = 1e6, dd-MS^2 resolution = 35000, dd-MS^2 AGQ target = 5e4, dd-MS^2 NCE = 25, dd-MS^2 isolation width = 1.6 m/z. QE+ collision energy was optimized based on fragmentation of cross-linked bradykinin[23]. Briefly, a 10uM of cross-linked bradykinin was resuspended in a 1:1 (v/v) solution of methanol/water and directly infused into the QE+ using the electrospray manifold. HCD NCE optimizations (NCE = 5 to NCE = 80) were acquired for the 817.4059 m/z (z=4) precursor ion, the NCE that generated the maximal base peak was ~25. QE+ data were then searched using SpectraST as described below.

SpectraST Library Generation, Spectral Matching and False Discovery Rate Estimation

Spectra from mzXML files were assembled into a spectral library based on peptide relationships that passed ReACT filtering, as described previously. ReACT, written in the ion trap control language (native language used for Thermo mass spectrometers) was run using default parameters as described above and previously[4,8,24]. SEQUEST parameters: peptide mass tolerance = 20 ppm; isotope error = 1; maximum internal cleavage sites = 3; variable modifications = 15.9949 (M, oxidation), 197.032422 (K, stump mass), required modifications = 57.021464 (C, carbamidomethylation). For searching, precursor mass tolerance was set to 20ppm and fragment ion tolerance was set to 0.36 Da. Spectra were searched against a target-decoy database, and resultant peptide-spectral matches were filtered based on expect score to an FDR < 5%. Of note, while the FDR filter for peptide identifications was set to a maximum of 5%, the final relationship FDR has often been found to be less than 1% for individual experiments[4,24].

SpectraST library generation was run with default values to generate a full spectral library for each of the above species. A probability of 1 was used for each peptide relationship assuming each ReACT relationship identified was fully correct. To input peptide-peptide relationships as opposed to single peptides, an underscore "_" was appended to the beginning of each relationship, which allowed for input of a relationship rather than just a peptide (e.g. "_PEPTIK@ER_PK@EPTIDER"). For spectral matching, mzXML files were matched using default SpectraST values with the exception of: (1) '-sz' was toggled on, (2) the no binning option was turned on ('-s_NOB'), and (3) the p-value calculation was turned on ('-s_PVL') (Figure S-7). FDRs for the runs was assessed by decoy-free estimation of incorrect hits as described previously[11]. When using the decoy-free approach, we found that using peak-to-peak, rather than Th bins, was the most conservative means to determine

FDRs by the decoy-free approach for high-resolution spectra (Figure S-8A). Spectral libraries will be available at brucelab.gs.washington.edu/.

FDRs were estimated by both the precursor swap and decoy-free methods from within SpectraST 5.0 (Figure S-1A). In both cases PeptideProphet was used to model the distributions of 'correct' and 'incorrect' SSMs to estimate the final FDRs, which were then filtered to an FDR < 5% (Figure S-8B).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants: U19-AI107775-02 (JEB), R01-AI101307-03 (JEB), R01-GM086688-06 (JEB), and R01-HL110879-04 (JEB). This work was supported in part by the University of Washington's Proteomics Resource (UWPR95794). The authors would like to thank the Chad Weisbrod for technical assistance.

References

1. Navare AT, et al. Probing the protein interaction network of Pseudomonas aeruginosa cells by chemical cross-linking mass spectrometry. Structure. 2015; 23:762–773. [PubMed: 25800553]
2. Rinner O, et al. Identification of cross-linked peptides from large sequence databases. Nature methods. 2008; 5:315–318. [PubMed: 18327264]
3. Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. Journal of mass spectrometry : JMS. 2003; 38:1225–1237. [PubMed: 14696200]
4. Schweppe DK, et al. Host-Microbe Protein Interactions during Bacterial Infection. Chem Biol. 2015; 22:1521–1530. [PubMed: 26548613]
5. Yang B, et al. Identification of cross-linked peptides from complex samples. Nature methods. 2012; 9:904–906. [PubMed: 22772728]
6. Lam H, et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 2007; 7:655–667. [PubMed: 17295354]
7. Frewen B, MacCoss MJ. Using BiblioSpec for creating and searching tandem MS peptide libraries. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis … [et al.] Chapter 13, Unit 13 17. 2007
8. Weisbrod CR, et al. In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. Journal of proteome research. 2013; 12:1569–1579. [PubMed: 23413883]
9. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry. 1994; 5:976–989. [PubMed: 24226387]
10. Depke M, et al. A peptide resource for the analysis of Staphylococcus aureus in host-pathogen interaction studies. Proteomics. 2015
11. Shao W, Zhu K, Lam H. Refining similarity scoring to enable decoy-free validation in spectral library searching. Proteomics. 2013; 13:3273–3283. [PubMed: 24115759]
12. Cheng CY, Tsai CF, Chen YJ, Sung TY, Hsu WL. Spectrum-based method to generate good decoy libraries for spectral library searching in peptide identifications. Journal of proteome research. 2013; 12:2305–2310. [PubMed: 23560440]
13. Tabb DL, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. Journal of proteome research. 2010; 9:761–776. [PubMed: 19921851]

14. Shi Y, et al. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. Molecular & cellular proteomics : MCP. 2014; 13:2927–2943. [PubMed: 25161197]
15. Hakhverdyan Z, et al. Rapid, optimized interactomic screening. Nature methods. 2015; 12:553–560. [PubMed: 25938370]
16. Subbotin RI, Chait BT. A pipeline for determining protein-protein interactions and proximities in the cellular milieu. Molecular & cellular proteomics : MCP. 2014; 13:2824–2835. [PubMed: 25172955]
17. Byrum S, Smart SK, Larson S, Tackett AJ. Analysis of stable and transient protein-protein interactions. Methods Mol Biol. 2012; 833:143–152. [PubMed: 22183593]
18. Chavez JD, et al. Quantitative interactome analysis reveals a chemoresistant edgotype. Nat Commun. 2015; 6
19. Kaake RM, et al. A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. Molecular & cellular proteomics : MCP. 2014; 13:3533–3543. [PubMed: 25253489]
20. Liu F, Rijkers DT, Post H, Heck AJ. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nature methods. 2015
21. Trowitzsch S, et al. Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. Nat Commun. 2015; 6:6011. [PubMed: 25586196]
22. Brittnacher MJ, et al. PGAT: a multistrain analysis resource for microbial genomes. Bioinformatics. 2011; 27:2429–2430. [PubMed: 21765097]
23. Park SG, Anderson GA, Navare AT, Bruce JE. Parallel Spectral Acquisition with an Ion Cyclotron Resonance Cell Array. Anal Chem. 2016; 88:1162–1168. [PubMed: 26669509]
24. Chavez JD, et al. Quantitative interactome analysis reveals a chemoresistant edgotype. Nat Commun. 2015; 6:7928. [PubMed: 26235782]

Significance

Protein cross-linking and mass spectrometry (XL-MS) techniques have emerged as useful tools for the large scale determination of interactions and structural interfaces between proteins. Searching against a spectral library generated from previous XL-MS studies improved the sensitivity and throughput of cross-linked peptide relationship identification. The presented approach, including the establishment of publicly-available spectral libraries, provides a platform to increase the accessibility of XL-MS studies to the broader community and provides the first application of spectral library identification of cross-linked interactions from complex samples.

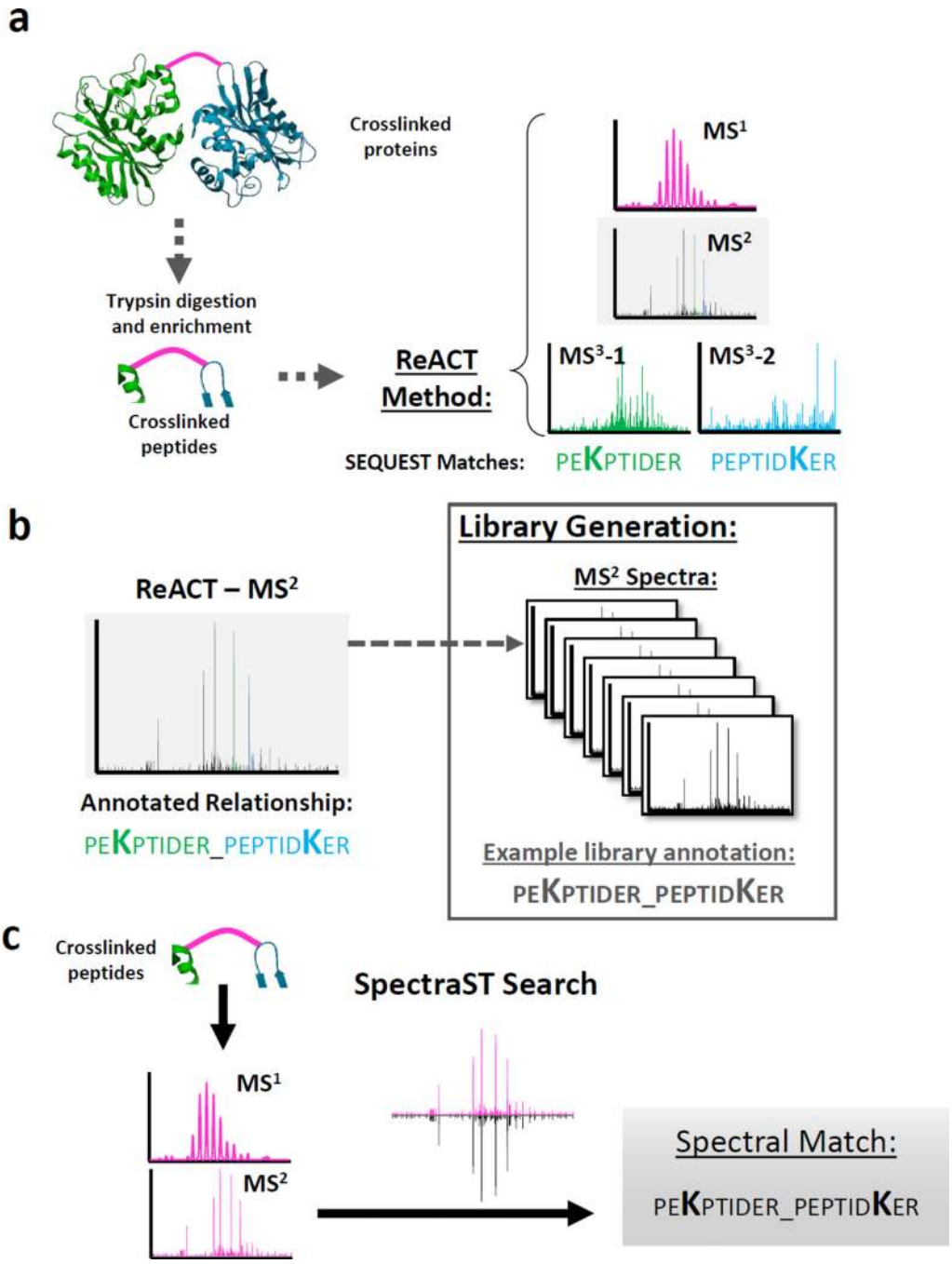


Figure 1.

Generation of cross-linked relationship spectral libraries. (a) ReACT identifies peptide matches from a cross-linked proteome. Unique, high-resolution MS^2 spectra from ReACT are generated that correspond to each relationship identification. MS^2 spectra include precursor masses for species generated from the low-energy dissociation of the CID-cleavable cross-linker. (b) MS^2 spectra are compiled into a spectral library. (c) This spectral library is then used to identify cross-linked peptide relationships.

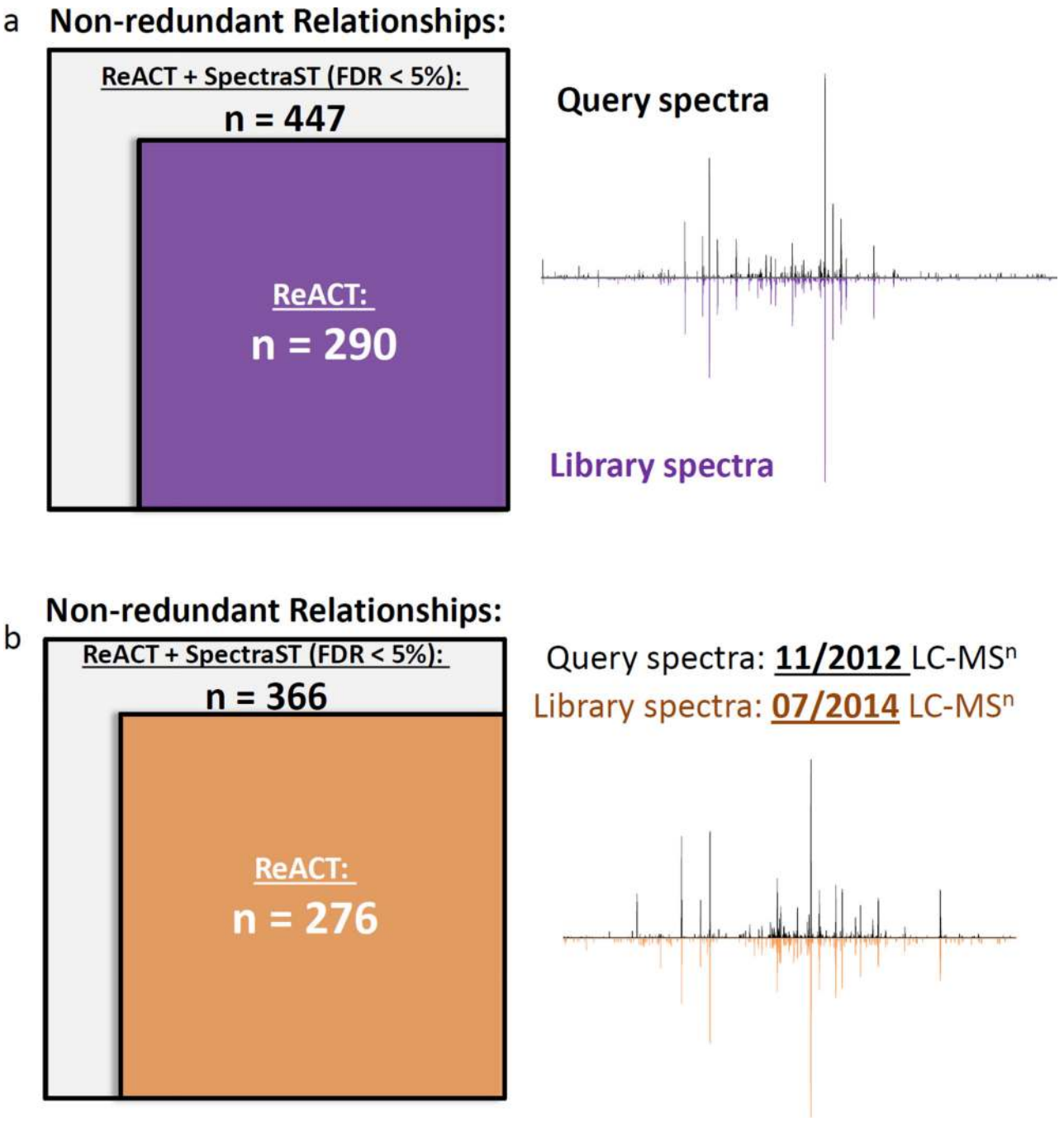


Figure 2.

Identification of cross-linked-peptide relationships by spectral matching. (a) The library-query sample was compared to the original ReACT identifications (n = 290). Spectral matching increased the number of confidently identified relationships by 157 at an FDR <5%. Inset depicts a spectral match between query and library spectra. (b) The replicate-query sample similarly showed a large increase in the number of confident identifications upon implementation of spectral library searching with ReACT and SpectraST (n = 366)

compared to ReACT alone (n = 276). Inset depicts a spectral match between query and library spectra.

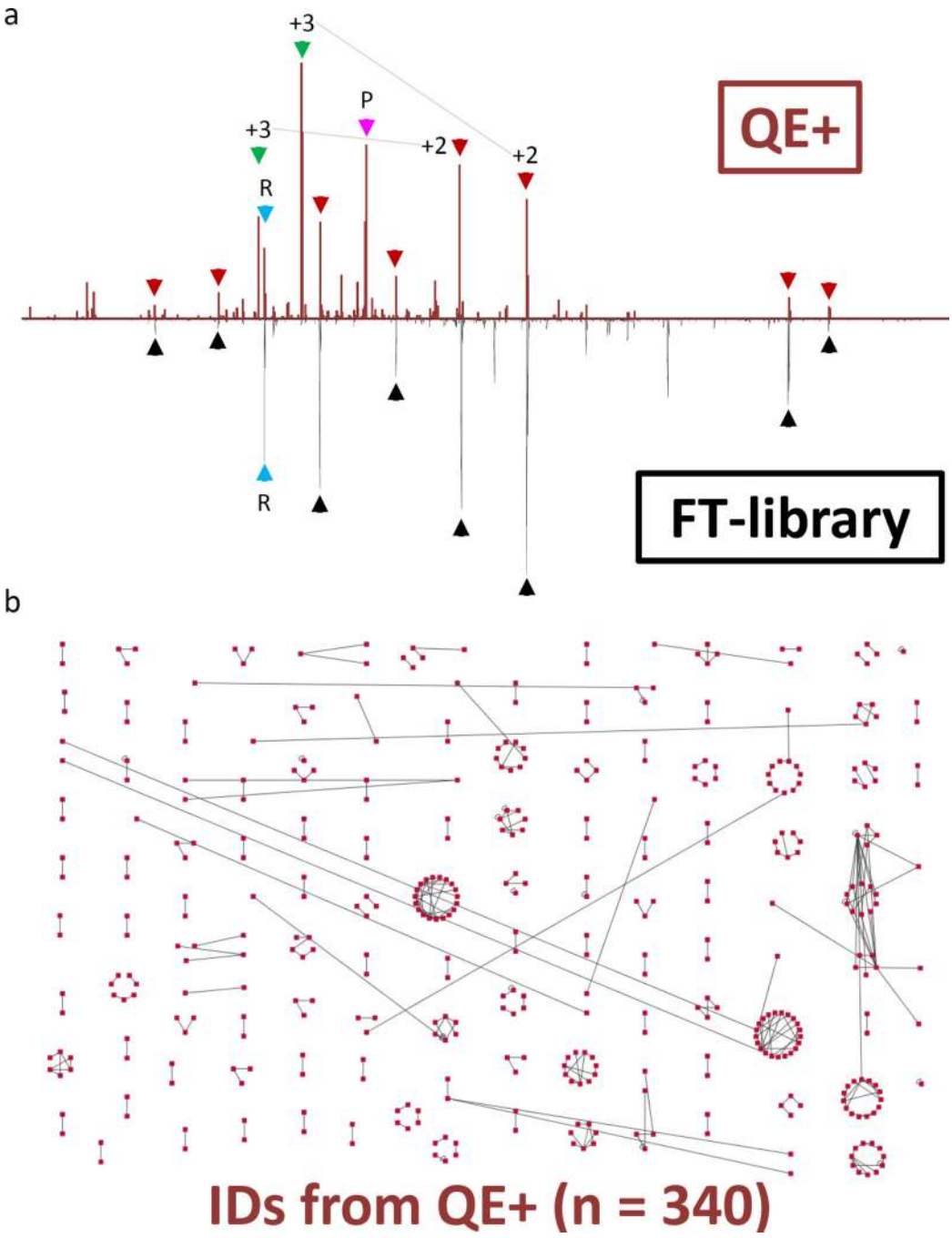


Figure 3.

Cross-linked peptide identification across platforms. (a) Using a ReACT-based library, samples were identified from canonical LC-MS/MS runs on a QE+. The high similarity of MS^2 spectra post-low-energy fragmentation is demonstrated by the high overlap of peaks across platforms. (b) Interaction network for the QE+ dataset demonstrates the ability of spectral-library searching to determine interactions via non-specialized MS-methods on a widely used high-resolution platform.

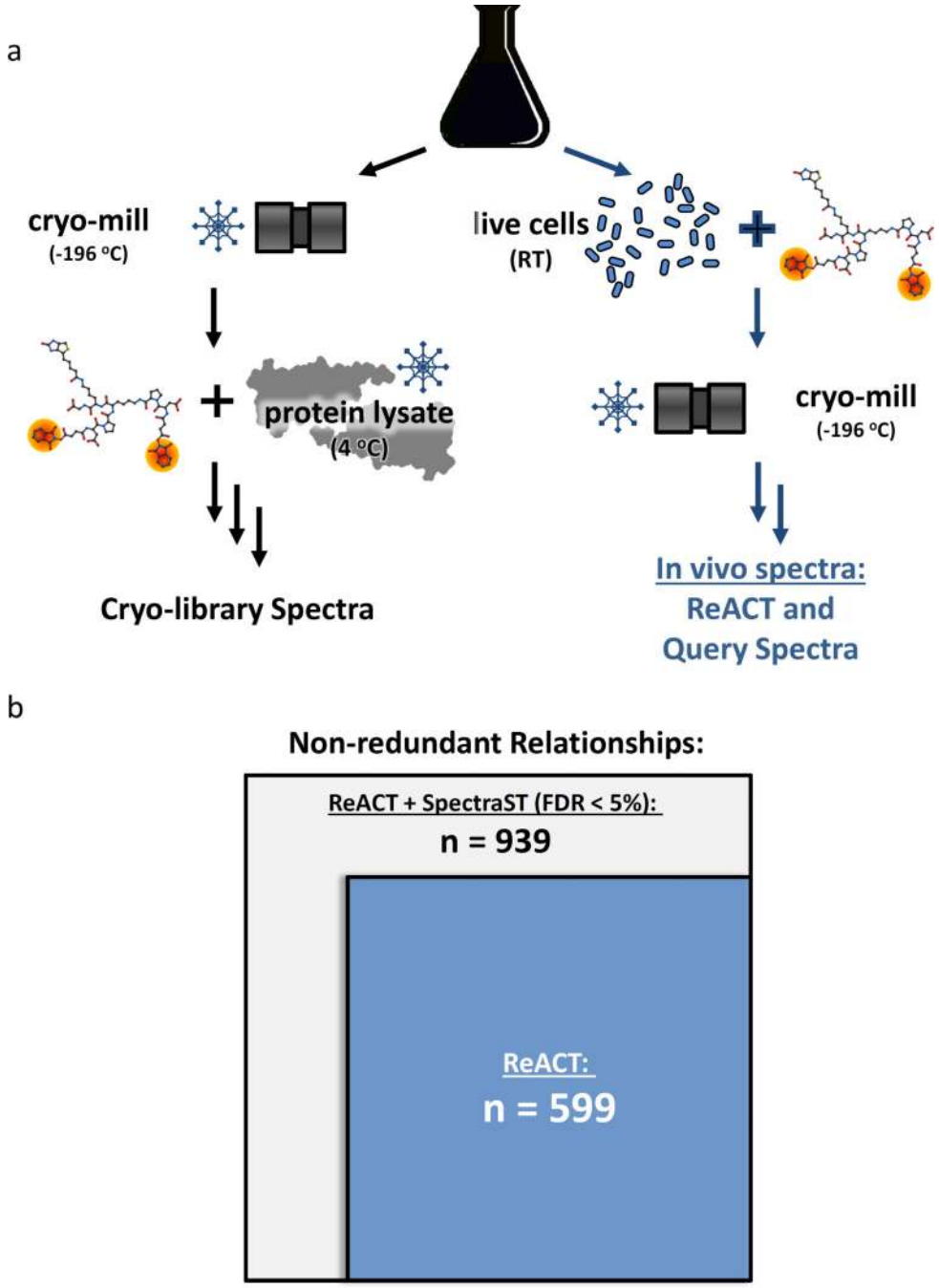


Figure 4.

Identification of *in vivo* cross-linked peptide relationships from a library generated with cryo-milled samples. (a) Library spectra were generated via ReACT from cryo-milled lysates that were cross-linked after lysis. The query spectra were generated from samples cross-linked in vivo and searched both by ReACT and using spectral library searching. (b) Consistent with our bacterial results, spectral-library searching improved the number of confidently identified relationships at an FDR <5%.