Published in final edited form as:

J Proteome Res. 2013 April 5; 12(4): 1989–1995. doi:10.1021/pr301162j.

XLink-DB: database and software tools for storing and visualizing protein interaction topology data

Chunxiang Zheng¹, Chad R. Weisbrod², Juan D. Chavez², Jimmy K. Eng³, Vagisha Sharma³, Xia Wu², and James E. Bruce^{1,2}

¹Department of Chemistry, University of Washington, Seattle

²Department of Genome Sciences, University of Washington, Seattle

³University of Washington Proteomics Resource

Abstract

As large-scale cross-linking data becomes available, new software tools for data processing and visualization are required to replace manual data analysis. XLink-DB serves as a data storage site and visualization tool for cross-linking results. XLink-DB accepts data generated with any cross-linker and stores them in a relational database. Cross-linked sites are automatically mapped onto PDB structures if available and results are compared to existing protein interaction databases. A protein interaction network is also automatically generated for the entire dataset. The XLink-DB server, including examples, and a help page are available for non-commercial use at URL: <http://brucelab.gs.washington.edu/crosslinkdbv1/>. The source code can be viewed and downloaded at <https://sourceforge.net/projects/crosslinkdb/?source=directory>.

Keywords

cross-linking; database; large scale datasets; protein interaction network; complex system; protein topology; data repository

1. Introduction

Protein interactions support most biological function and are directed by shapes or topologies of the interacting proteins. Improved measurements of protein interaction topologies in cells are needed to increase our understanding of how protein interactions carry out their life supporting functions. Chemical cross-linking with mass spectrometry has been used to study protein structures and complex topologies for several years^{1–21}. Most prior applications have been limited to either purified proteins or complexes due to the complexity and wide dynamic range presented by complex biological samples. Recent technical advancements of the chemical cross-linking methods achieved in a number of labs have allowed this technique to be extended to complex systems^{13, 22–24}. Successful applications of chemical cross-linking to studies of intact virus particles, cell lysates, and even intact bacterial and human cells suggest that in the future, cross-linking methods may provide a majority of structural and topological data on protein complexes as they exist in cells or other complex samples^{22, 24–26}.

As is the case with most large-scale biological data, its usage among investigators in biochemistry, biophysics, cellular and molecular biology, as well as proteomics requires that

new tools be developed to visualize, share and compare these results. This is especially true for large-scale cross-linking data since current growth in data quantity exceeds manual data analysis capabilities. Furthermore cross-linking with mass spectrometry datasets are unique in that they contain multiple tiers of information on protein sequence, interaction, and structural levels for which no single existing data analysis tool can sufficiently support. Often data analysis requires comparison of cross-linking results with existing crystal structure data if available. In addition, cross-linking data are often compared with existing protein interaction data. If previously unknown interactions are discovered, the cross-linked site information can be superimposed by computational docking of interacting structures. These steps can require hours of efforts even with only a few cross-linked peptide pairs in a single experiment and this approach becomes intractable for hundreds of cross-linked peptides.

Here we report development of XLink-DB which was designed to serve both as a storage site and an online data processing and visualization tool to enable analysis of large-scale cross-linked peptide datasets. Importantly, XLink-DB will be useful among biological and proteomics research communities since it provides new analysis capabilities and improved access to complex cross-linking topological data. XLink-DB allows users to upload their cross-linking data and populate a relational database, as well as browse existing datasets. XLink-DB automatically retrieves related protein sequence information from UniProt²⁷ and high resolution structure information from the Protein Data Bank (PDB)²⁸. If relevant structures are available, cross-linked site annotation is automatically performed with XLink-DB and visualized within the Jmol applet (<http://jmol.sourceforge.net/>)²⁹. The cross-linking data is also visualized in a protein interaction network view with an embedded web-based Cytoscape tool³⁰. The data stored in XLink-DB will be compared to existing protein interaction databases such as IntAct³¹ and EciD³². We anticipate that XLink-DB will be a useful tool and benefit the proteomics research community as well as all researchers interested in protein topologies and interactions.

Overview

The XLink-DB website was developed with PHP 5.5 and JavaScript, data analysis tools were programmed with Java 1.6 and data were stored in a MySQL database. The functionality of the website also depends on both Java applets and flash plug-in. As shown in Figure 1, the website contains two major modules: 1) Data upload, process and storage and 2) Data visualization. Five different views (interaction network, protein structure, search, site and table views) are available for cross-linked peptide data analysis. Interaction network view shows the protein interaction network generated from the dataset. Protein structure view shows the cross-linking peptide pairs on the existing PDB structure. A key feature of XLink-DB is the ability to map cross-linked sites on protein complexes for which individual protein crystal structures exist, but no co-crystal have been reported. Site view is designed to display the sites when the co-crystal structure does not exist. Search view is a sub-network of the dataset. The table view is a summary of the dataset in a table. To help users get familiar with the features of the database, we have created a video tutorial which can be found in the help page. In addition, we have also put tooltips on some parameters to guide the users. Details on each module are discussed below.

1) Data upload, process and storage

The users can choose if they want their data to be publically available. If they choose not to release their data to the public, they will get a table name after the data upload is finished and their data will not appear in the drop-down list to choose. Instead, the users can use the table name to access their non-public data. Their data will be stored in the database for 90 days. If the user chooses to make their data public available, the data will be permanently

stored in the database and will appear in a dropdown list in the selection box under “Choose a dataset”. The users can access their published and previously uploaded data from the dropdown list. Data are uploaded in XLink-DB in a tab-delimited file format with column arrangements as indicated on XLink-DB help page (<http://brucelab.gs.washington.edu/crosslinkdbv1/help.php>). XLink-DB parses the input file to extract the UniProt identifiers for each cross-linked protein contained within the dataset. The UniProt files (.txt files) containing protein annotation is then automatically downloaded from the UniProt database. The sequence information and identifiers for each labeled protein are parsed from the UniProt file and stored within the database in XLink-DB. If available, the PDB code associated with each protein is also retrieved from the UniProt annotation. For cases where more than one PDB code is associated with one protein, XLink-DB will select and retrieve the PDB structure based on the following rules: First XLink-DB will find all the PDB files which contain structural information covering the cross-linked site. If two cross-linked peptides originate from different protein sequences, which identifies a hetero interaction, all the co-crystal structures containing the two labeled proteins will be put in the candidate pool for later selection. Next, if the cross-linked peptide pair contains identical or overlapping peptide sequences that originate within a single protein sequence, all oligomer structure files containing both sites will be put in the candidate pool. If the cross-linked peptide pair does not fall into either of the two categories above, individual structure files containing both sites will be put into the candidate pool. Finally, the software will choose the structure with highest sequence coverage from the candidate pool to use for visualization of the cross-linked peptide pair. The structure with highest sequence coverage is chosen because they allow the best representation of the entire protein and greatest chance to cover cross-linked sites. If no structural file can be found which contains both labeled sites, the software will choose the best individual structures for each labeled site.

After the PDB codes are assigned to each protein, the PDB files for these proteins are automatically downloaded. XLink-DB then computes atom numbers for all cross-linked peptide sites with the following steps: First, the peptide sequence is mapped to the protein sequence in the PDB file. Next, the atom numbers and coordinates of every copy of the cross-linked peptide in the PDB file are identified. The chosen atoms are the alpha carbon of the cross-linked lysine residues. The shortest distance between the two cross-linked sites contained in each cross-linked peptide pair is then calculated from the atomic coordinates of the alpha carbon atoms. Finally, the associated atom numbers of the cross-linked sites are stored within the database embedded in XLink-DB.

The final data processing step is to compare the uploaded data with an existing protein interaction database. For this case we used the databases IntAct³¹ and EciD³². We chose these two databases based on the coverage of protein interaction data. IntAct is used for human data. For *E.coli* data, EciD is used instead because it has a better coverage on the *E.coli* protein interaction data. The computed distances between two cross-linked proteins serve as measurement from the reference protein interaction network composed from existing protein interaction database information. For example, if two cross-linked proteins were previously known to interact, the computed distance within the reference protein interaction network is 0, otherwise the computed distance is the smallest number of nodes or proteins that exist in the reference network linking the two cross-linked proteins. If the cross-linked proteins cannot be connected in the reference network, “N/A” will be returned for this computed distance.

2) Data visualization

2.1 Network view—In Network View, a protein interaction network of the cross-linked peptide dataset will be generated with Cytoscape plugin, and be displayed on the left side of

the page. A complete set of features available in the Cytoscape plugin are described by Lopez, *et al.*³⁰ Each node represents a protein and each edge represents all the cross-linked peptide pairs linking the two proteins. The users can open files, save files and change the layout and style options from the menu on the top. The toolbox at the right bottom corner of the network graph enables panning and zooming in the graph. Every node and edge in the graph can be selected, dragged and edited. The right-hand side of the page contains three tabs: Visual Style, Filter and Properties. The Visual Style tab allows users to change the color of the nodes, edges and background. The Filter tab allows users to filter the nodes based on the value of attributes. The properties tab is automatically activated when nodes or edges are selected. When one or more nodes are selected, the interacting partners of the selected nodes will be listed in a table. The name of each interacting partner is converted into a button which will lead to the Protein View of this protein complex. When one or more edges are selected, the interactions which are represented by the selected edges will be listed in a table. Each interaction is converted to a button which will lead to the Protein View of the pair. In addition, the protein interaction network developed with cross-linking data is compared with previous known protein structural and interaction information. For instance, the size of the node indicates a crystal structure for the protein exists in PDB. The thickness of the edges is related to the number of cross-linked peptide pairs that have been identified in the dataset, with thicker lines indicative of 2 or more cross-links. The color of the edge indicates the distance of connection of the two proteins in reference protein interaction database. Red edges indicate direct interactions between linked proteins are found in IntAct or EciD. Green edges indicate linked proteins have been found to share a common interactor in the reference database and are therefore one node away. Black edges indicate linked proteins are more than one node away or were not found in the reference databases. It should be noted that, for linkages that contain two peptides from the same protein, the edge color will appear red unless one or more cross-linked pairs are comprised of two peptides with overlapping sequences indicating unambiguous linkage of a homodimer. In these unambiguous homodimer cases, proteins previously known to form homomultimers will appear with red edges, while those not yet known to form homomultimers will appear with green edges.

2.2 Protein View—Protein View page contains a Jmol applet²⁹ on the top if the structure is available, and a result table on the bottom. The user can change basic display options with right-click menu in the Jmol layer. Two buttons are available to change the display of cross-linked peptide pairs. “Display all” button illustrates all cross-linked sites associated with the two proteins displayed in the Jmol layer. “Reset complex” button will remove all the cross-linking pairs labeled on the structure. The bottom part of the page contains a result table with all the pairs associated with the two proteins. This table contains peptide sequence, gene name, PDB code, number of cross-linked pairs that involve the peptide and display option button. The number of cross-linking pairs involving the peptide is a measurement of reactivity and spatial proximity of the labeled site. A larger number indicates the labeled site is close to many other sites and the labeled site is highly reactive. The “display single pair” button will display the selected pair on the structure. The users can also use their own favorite structure if they do not appreciate the pre-assigned structures. They need to input the PDB code and the chain IDs for the respective proteins.

2.3 Table View—The Table View page can be accessed from the Network View by clicking on the “Generate table view” button. The result table page contains two parts; the top part shows the link to the network view and the title. The bottom part is the result table with peptide sequence, protein accession, PDB code, distance of connection and links to protein view. This table can be sorted by entries within each column by clicking on the column heading. Each entry in Peptide A/B columns is hyperlinked to the Site view page

which will be described later. Protein names shown in columns Protein A/B within the table are hyperlinked to relevant UniProt pages for each protein to facilitate further investigation. Similarly, “PDB code for peptide A/B” names are hyperlinked to the relevant PDB page for additional structure information if needed. The “Show structure” button produces a protein-level view of the cross-linked pair.

2.4 Site view—As mentioned above, the Site View shows the two labeled sites in two parallel windows. This enables users to visualize the location of the labeled peptide in the protein. When the crystal structure is available for the either protein but not the complex, the site will be highlighted magenta on the structure; otherwise the entire cross-linked peptide will be highlighted red in the protein sequence.

2.5 Search view—Search view can be accessed from the home page. The user can choose UniProt ID, UniProt accession or gene name to search for any protein of interest. The user can either search one protein or give a list of protein IDs to search. The search will be performed against all the datasets for the selected organism.

Results

Two datasets are used to demonstrate the features of XLink-DB. One is a large scale cross-linking experiment performed in our laboratory on intact *E. coli* cells (See companion manuscript by Weisbrod et al.) “Weisbrod *et al.*” dataset is used here to denote this data from *E. coli* cells. The other dataset was extracted from a recent publication by Yang *et al.* in which the researchers performed cross-linking on *E. coli* cell lysate.²⁴ “Yang *et al.*” is used here to refer to this dataset. Both datasets comprise the largest reported cross-linking datasets and contain several hundred unique cross-linked sites. There are a few differences in the two experiments. Weisbrod *et al.* used customized cross-linker which is mass spectrometry cleavable and has biotin affinity tag for purification. Yang *et al.* used commercially available DSS which is non-cleavable. Both dataset used strong cation exchange to enrich high charge peptides. Weisbrod *et al.* performed avidin capture to enrich biotin-tagged peptides prior to mass spectrometry analysis. Using XLink-DB to analyze these datasets provides unique insight into datasets which would have been difficult and time consuming to get manually. Figure 2 illustrates the distribution of cross-linked distances mapped by XLink-DB. These distances are extracted from XLink-DB and plotted in Excel. Both datasets show broad distributions of observed cross-linked distances. Disuccinimidyl suberate (DSS) a cross-linker with a relatively short spacer arm length (11.4Å) was applied in the “Yang *et al.*” dataset. The cross-linker used in the “Weisbrod *et al.*” dataset has a spacer arm longer than 30Å, the fact that both datasets show similar cross-linked distance distributions suggests that cross-linker size is less important than protein flexibility in determination of which protein sites are cross-linked in complex mixtures.

Using XLink-DB both datasets were compared to the *E. coli* protein interaction database (EciD, only considering interactions from experimentally derived data). Figure 3 shows the distribution of the node distances of both datasets and a Monte-Carlo simulation of the expected distance for randomly selecting two proteins. Both cross-linking datasets consist of approximately 130 inter-protein interactions. For the Monte-Carlo simulation, 130 randomly selected protein pairs were chosen to represent the sample size of the cross-linking experiment. The experiment was repeated 100 times and the average percentage of each distance is plotted in Figure 3. Based on the Monte-Carlo simulation, the most probable expected distance of two randomly chosen proteins is 2 nodes. The majority of the distances for the two cross-linking datasets is below or equal to one node, suggesting that both “Weisbrod *et al.*” dataset and “Yang *et al.*” cross-linking experiments show good correlation with other experimental techniques. Furthermore, the “Weisbrod *et al.*” dataset

contains the highest percentage (25%) of known direct interactors (0 nodes), whereas random simulation predicts about 4%. This suggests that data from either the “Weisbrod *et al.*” or “Yang *et al.*” cross-linking experiments is significantly different from random data based on existing known interactions from EciD.

Discussion

Several protein interaction databases have been established and embraced by the scientific community, such as PDB, EciD and IntAct. But none of them provide the features that XLink-DB offers. While PDB represents a significant resource in terms of available protein crystal structures and databases like EciD and IntAct contain significant wealth of data on protein interactions, there currently is a void of databases that contain protein interaction topological data. This likely stems from the lack of technological capabilities to produce data of this kind, but new technologies and advancements are rapidly changing the situation^{13, 16, 22, 24, 26}. XLink-DB was developed to help fill this void in database availability and maximize the access and utility of protein interaction topological data that is now available and will come from these technological advancements.

In conclusion, XLink-DB presents a new way to organize and demonstrate protein interaction data with topological information. Conventional databases either lack the interaction information or lack the topological information for the protein complexes. With the advancement of new cross-linking technologies, large scale protein interaction studies are now becoming reality. XLink-DB is the first database to allow compilation and analysis of large-scale cross-linking data. It will not only help the cross-linking community to store, share and process their data, but also share the data with other scientists with interests in protein interactions and topologies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by National Institutes of Health [R01RR023334, R01GM086688, R01GM097112, S10RR02510].

References

1. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nature reviews Molecular cell biology*. 2007; 8 (8):645–54.
2. Muller DR, Schindler P, Towbin H, Wirth U, Voshol H, Hoving S, Steinmetz MO. Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Analytical chemistry*. 2001; 73 (9):1927–34. [PubMed: 11354472]
3. Rappsilber J, Siniosoglou S, Hurt EC, Mann M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Analytical chemistry*. 2000; 72 (2):267–75. [PubMed: 10658319]
4. Huang BX, Kim HY, Dass C. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *Journal of the American Society for Mass Spectrometry*. 2004; 15 (8):1237–47. [PubMed: 15276171]
5. Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *Journal of molecular biology*. 2003; 331 (2):303–13. [PubMed: 12888339]
6. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from

- intramolecular cross-links and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97 (11):5802–6. [PubMed: 10811876]
7. Chen T, Jaffe JD, Church GM. Algorithms for identifying protein cross-links via tandem mass spectrometry. *J Comput Biol*. 2001; 8 (6):571–83. [PubMed: 11747613]
 8. Chu F, Mahrus S, Craik CS, Burlingame AL. Isotope-coded and affinity-tagged cross-linking (ICATXL): an efficient strategy to probe protein interaction surfaces. *J Am Chem Soc*. 2006; 128 (32):10362–3. [PubMed: 16895390]
 9. Gomes AF, Gozzo FC. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J Mass Spectrom*. 2010; 45 (8):892–9. [PubMed: 20635431]
 10. Kalkhof S, Ihling C, Mechtler K, Sinz A. Chemical cross-linking and high-performance Fourier transform ion cyclotron resonance mass spectrometry for protein interaction analysis: application to a calmodulin/target peptide complex. *Anal Chem*. 2005; 77 (2):495–503. [PubMed: 15649045]
 11. Muller MQ, Zeiser JJ, Dreier F, Pich A, Schafer M, Sinz A. A universal matrix-assisted laser desorption/ionization cleavable cross-linker for protein structure analysis. *Rapid Commun Mass Spectrom*. 2011; 25 (1):155–61. [PubMed: 21157862]
 12. Petrotchenko EV, Xiao K, Cable J, Chen Y, Dokholyan NV, Borchers CH. BiPS, a photocleavable, isotopically coded, fluorescent cross-linker for structural proteomics. *Mol Cell Proteomics*. 2009; 8 (2):273–86. [PubMed: 18838738]
 13. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. *Nat Methods*. 2008; 5 (4): 315–8. [PubMed: 18327264]
 14. Silva RA, Hilliard GM, Fang J, Macha S, Davidson WS. A three-dimensional molecular model of lipid-free apolipoprotein A-I determined by cross-linking/mass spectrometry and sequence threading. *Biochemistry*. 2005; 44 (8):2759–69. [PubMed: 15723520]
 15. Sinz A, Wang K. Mapping spatial proximities of sulfhydryl groups in proteins using a fluorogenic cross-linker and mass spectrometry. *Anal Biochem*. 2004; 331 (1):27–32. [PubMed: 15245993]
 16. Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem*. 2005; 77 (1):311–8. [PubMed: 15623310]
 17. Yang T, Horejsh DR, Mahan KJ, Zaluzec EJ, Watson TJ, Gage DA. Mapping cross-linking sites in modified proteins with mass spectrometry: an application to cross-linked hemoglobins. *Anal Biochem*. 1996; 242 (1):55–63. [PubMed: 8923964]
 18. Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kaminski T, Rasmussen M, Lariviere L, Bukowski-Wills JC, Nilges M, Cramer P, Rappsilber J. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *Embo J*. 29(4):717–26. [PubMed: 20094031]
 19. Herzog F, Kahraman A, Boehringer D, Mak R, Bracher A, Walzthoeni T, Leitner A, Beck M, Hartl FU, Ban N, Malmstrom L, Aebersold R. Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science*. 337(6100):1348–52. [PubMed: 22984071]
 20. Maiolica A, Cittaro D, Borsotti D, Sennels L, Ciferri C, Tarricone C, Musacchio A, Rappsilber J. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol Cell Proteomics*. 2007; 6 (12):2200–11. [PubMed: 17921176]
 21. Karadzic I, Maupin-Furlow J, Humbard M, Prunetti L, Singh P, Goodlett DR. Chemical cross-linking, mass spectrometry, and in silico modeling of proteasomal 20S core particles of the haloarchaeon *Haloferax volcanii*. *Proteomics*. 12(11):1806–14. [PubMed: 22623373]
 22. Zheng C, Yang L, Hoopmann MR, Eng JK, Tang X, Weisbrod CR, Bruce JE. Cross-linking measurements of in vivo protein complex topologies. *Mol Cell Proteomics*. 2011; 10 (10):M110 006841.
 23. Walzthoeni T, Claassen M, Leitner A, Herzog F, Bohn S, Forster F, Beck M, Aebersold R. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Methods*. 2012

24. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ. Identification of cross-linked peptides from complex samples. *Nat Methods*. 2012
25. Chavez JD, Cilia M, Weisbrod CR, Ju HJ, Eng JK, Gray SM, Bruce JE. Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus-plant interactions. *J Proteome Res*. 2012; 11 (5):2968–81. [PubMed: 22390342]
26. Zhang H, Tang X, Munske GR, Tolic N, Anderson GA, Bruce JE. Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol Cell Proteomics*. 2009; 8 (3):409–20. [PubMed: 18936057]
27. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004; 32(Database issue):D115–9. [PubMed: 14681372]
28. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*. 1977; 112 (3):535–42. [PubMed: 875032]
29. Herraiz A. Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Educ*. 2006; 34 (4):255–61. [PubMed: 21638687]
30. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. Cytoscape Web: an interactive web-based network browser. *Bioinformatics*. 2010; 26 (18):2347–8. [PubMed: 20656902]
31. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeifferberger E, Porras P, Raghunath A, Roehert B, Orchard S, Hermjakob H. The IntAct molecular interaction database in 2012. *Nucleic acids research*. 40(Database issue):D841–6. [PubMed: 22121220]
32. Andres Leon E, Ezkurdia I, Garcia B, Valencia A, Juan D. EcID. A database for the inference of functional interactions in *E. coli*. *Nucleic Acids Res*. 2009; 37(Database issue):D629–35. [PubMed: 19004873]

Figure 1.
Internal structure and algorithms in XLink-DB
A) Web structure of XLink-DB
B) The data process scheme for uploaded data
C) The algorithms of choosing the best PDB structure

Figure 2. Distribution of interlinked distances of large-scale cross-linked peptide datasets from cells and cell lysates. Distances are computed in XLink-DB from all cross-linked sites that appear within crystal structures available from the PDB. Cell lysate data (Yang *et al.*, 2012) is shown in blue and Weisbrod *et al.* data shown in red.

Figure 3.

Distribution of the node distances observed in cross-linked peptide datasets from cell lysates (Yang et al., 2012) shown in blue and intact cells shown in red as determined from the *E. coli* protein interaction database EciD. Also shown in green is the expected nodal distance distribution for random selection of 2 proteins shown in green.