Identifying Cross-linked Peptides using Protein Prospector

Robert J. Chalkley[1], Michael J. Trnka[1], Nicholas Michael[2], Peter R. Baker[1]
[1]NIGMS Mass Spectrometry Resource, University of California San Francisco, USA
[2]Reading Scientific Services Ltd, Reading, UK

Introduction

Cross-linking through disulfide bridges is used by nature for enforcing protein structure. Chemical cross-linkers can be exploited to provide distance restraints for three-dimensional structural analysis of proteins and protein machines. In both cases, after enzymatic digestion the product is two (or more) peptides covalently linked together. Identification of these products using tandem mass spectrometry and database searching is challenging. Recently, a couple of bioinformatic software tools and approaches have been presented to try to identify these products. In this poster, case studies are presented to demonstrate newly developed cross-linking analysis features in Protein Prospector.

Methods

Protein Prospector identifies each peptide in the cross-linked product as single peptides with large mass modifications on the site of cross-linking. Hence, regular protein databases can be queried, rather than the alternative approach of creating artificial peptide databases containing all potential combinations of cross-linked peptides. This makes the analysis of complexes containing many protein subunits tractable. Also, by identifying the two peptides separately before combining the identifications for one result, one can use the evidence for the less confident peptide identification as the reliability measure for the cross-linked peptide complex identification, producing results with a high reliability while still maintaining sensitivity. A further improvement in performance is achieved through employing separate thresholds for intra-protein and inter-protein identifications.

Chemical Cross-linking

Used for detecting interaction interfaces within (intra-protein) and between proteins (inter-protein). Most reagents target primary amino groups found in lysine side chains and protein N-termini

Case Study 1

UTP-B is a six member complex that is part of the small ribosome subunit in yeast. For details of its purification see [1]. Crosslinking was performed using a 50:50 mix of [d0]- and [d4]-bis(sulfosuccinimidyl) suberate. LCMS-MS data was acquired byHCD fragmentation using a LTQ-Orbitrap. Data was analyzed using Protein Prospector, then compared to results previously published from this dataset using alternative cross-linking, pLink [1]

Cross-linked residues identified between members of the UTP-B complex

UTP1 to UTP1	**UTP6 to UTP6**	**UTP12 to UTP21**
6:46	321:361	752:929
27:46	389:397	890:906
27:85	389:439	
46:85		**UTP13 to UTP13**
56:674		51:91
96:129	**UTP6 to UTP13**	86:94
96:180	72:751	179:181
88:129		181:228
98:166	**UTP12 to UTP12**	533:555
98:674	163:187	699:751
129:180	230:318	741:780
166:264	237:318	
211:264	253:337	
536:557	279:337	**UTP18 to UTP18**
536:572	279:381	134:154
557:572	404:420	154:170
572:674	486:503	538:585
733:753	595:635	
	774:780	**UTP18 to UTP21**
UTP1 to UTP6	774:787	245:341
65:572	877:884	288:341
		288:538
UTP1 to UTP12	**UTP12 to UTP13**	408:538
111:262	1:741	
800:884	111:741	**UTP21 to UTP21**
	381:533	6:9
UTP1 to UTP18	381:555	7:9
418:572	404:569	9:9
	515:546	9:19
UTP1 to UTP21	533:555	9:661
6:661	699:843	408:435
27:730	855:751	502:539
85:661	866:815	794:804
96:382	877:815	794:819
9:129	884:815	806:819
102:129	909:780	
129:382		
245:761		

At estimated 5% FDR (for each software):

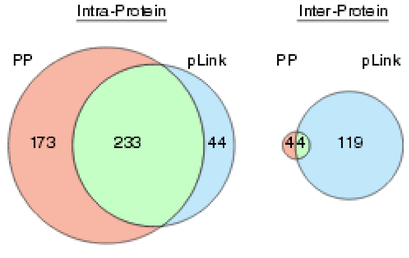
Identified by both software (67)
Unique to Protein Prospector (17)
Unique to pLink (11)

At 5% FDR, estimated 3 incorrect results. These are likely to be results unique to one software. False positives are more likely for inter-protein results than intra-protein.

- 3 of the unique pLink IDs report new protein-protein interactions.
  - Protein Prospector reports 11 protein:protein interactions
  - pLink reports 14 protein:protein interactions
- All protein:protein interactions reported by Protein Prospector are supported by at least two cross-links.
- False discovery rate among protein:protein interactions by pLink = 3/14 = 21%?

Case Study 2

E. Coli whole cell lysate was cross-linked using a 50:50 mix of [d0]- and [d4]-bis(sulfosuccinimidyl) suberate. then data acquired by HCD fragmentation using a LTQ-Orbitrap. Data was analyzed using Protein Prospector, then compared to results previously published from this dataset using alternative cross-linking, pLink [1]



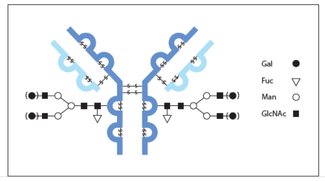
At estimated 5% FDR
Protein Prospector: 414 IDs
pLink: 390 IDs

Significantly different results were reported by the two software , partly due to Protein Prospector evaluating the reliability of inter-protein and intra-protein results separately. Protein Prospector reported only 8 reliable inter-protein cross-links, whereas a third of all results reported by pLink were inter-protein. Conversely, Protein Prospector reported many more intra-protein cross-links.

Disulfide-Bridge Mapping

Case Study 3

Intact mAb Mass Check Standard sold by Waters as a LC/MS standard for confirming LC/MS system operation



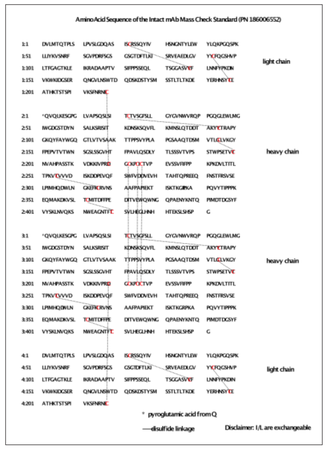


Figure 2. Protein Sequence Information

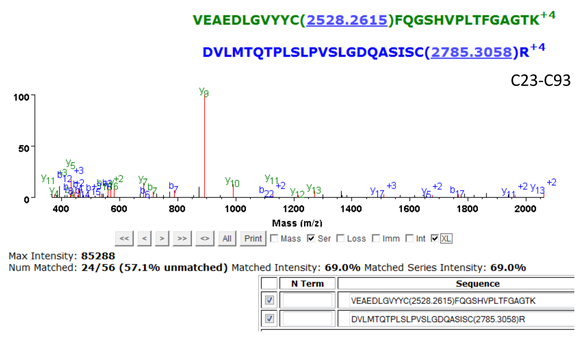
(from literature supplied with sample).

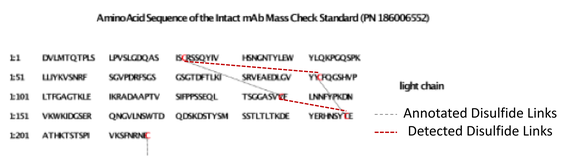
Analysis:

1. Digest with Trypsin.
2. Acquire LC-MSMS data using Q-Exactive.
3. Analyze for disulfide links using Protein Prospector.

[+] Display and Parameter Settings

1	2	3	4	8	9	16	18	19	20	21	22	23	24
m/z	z	ppm	Crosslinked Peptide	Score	Expect	L Sc	Start 1	Start 2	End 1	End 2	Acc #	XLink AA 1	Intra 1
946.3059	6	-1.5	SRVEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK DVLMTQTPLSLPVSLGDQASISC(3141.5230)R	42.5	4.0e-14	8.8	80	1	108	24	1	93	1
946.3075	6	0.22	SRVEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK DVLMTQTPLSLPVSLGDQASISC(3141.5230)R	36.4	8.3e-14	7.3	80	1	108	24	1	93	1
949.3100	6	2.6	SRVEAEDLGVYYC(2544.2564)FQGSHVPLTFGAGTK+Incorrect Mono 2Da DVLMtOxidation(T)QTPLSLPVSLGDQASISC(3141.5230)R+Incorrect Mono 2Da	27.6	3.3e-10	6.0	80	1	108	24	1	93	1
1329.9019	4	-0.82	VEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK DVLMTQTPLSLPVSLGDQASISC(2785.3058)R	40.6	1.8e-15	12.2	83	1	108	24	1	93	1
760.6600	7	-3.9	VEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK+Incorrect Mono 2Da DVLMTQTPLSLPVSLGDQASISC(2785.3058)R+Incorrect Mono 2Da	37.0	2.1e-14	6.0	83	1	108	24	1	93	1
1064.1227	5	-1.1	VEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK DVLMTQTPLSLPVSLGDQASISC(2785.3058)R	33.4	4.2e-13	9.8	83	1	108	24	1	93	1
1064.1250	5	1.1	VEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK DVLMTQTPLSLPVSLGDQASISC(2785.3058)R	32.5	5.3e-13	7.0	83	1	108	24	1	93	1
1067.9200	5	-4.5	VEAEDLGVYYC(2544.2564)FQGSHVPLTFGAGTK+Incorrect Mono 3Da DVLMtOxidation(T)QTPLSLPVSLGDQASISC(2785.3058)R+Incorrect Mono 3Da	32.8	1.0e-12	7.9	83	1	108	24	1	93	1
1067.3236	5	0.72	VEAEDLGVYYC(2544.2564)FQGSHVPLTFGAGTK DVLMtOxidation(T)QTPLSLPVSLGDQASISC(2785.3058)R	29.9	1.1e-11	10.8	83	1	108	24	1	93	1
1773.8700	3	-0.90	VEAEDLGVYYC(2528.2613)FQGSHVPLTFGAGTK+Incorrect Mono 3Da DVLMTQTPLSLPVSLGDQASISC(2785.3058)R+Incorrect Mono 3Da	21.6	2.1e-10	6.3	83	1	108	24	1	93	1
1075.7236	5	-1.3	VEAEDLGVYYC(2586.2669)FQGSHVPLTFGAGTK Acetyl-DVLMtOxidation(T)QTPLSLPVSLGDQASISC(2785.3058)R	37.4	1.5e-9	10.3	83	1	108	24	1	93	1
827.2302	6	-1.3	RADAAPTVSIFPPSSEQLTSGGASVVCC(1287.5302)FLNNFYPK HNSYTC(3667.7981)EATHK	30.1	2.8e-11	6.2	113	194	147	204	1	139	1
992.4752	5	-0.88	RADAAPTVSIFPPSSEQLTSGGASVVCC(1287.5302)FLNNFYPK HNSYTC(3667.7981)EATHK	31.6	5.4e-11	6.2	113	194	147	204	1	139	1
827.2307	6	-0.69	RADAAPTVSIFPPSSEQLTSGGASVVCC(1287.5302)FLNNFYPK HNSYTC(3667.7981)EATHK	36.4	8.3e-11	10.0	113	194	147	204	1	139	1
992.4745	5	-1.6	RADAAPTVSIFPPSSEQLTSGGASVVCC(1287.5302)FLNNFYPK HNSYTC(3667.7981)EATHK+Incorrect Mono 3Da	27.6	2.3e-9	6.2	113	194	147	204	1	139	1
827.7300	6	-3.6	RADAAPTVSIFPPSSEQLTSGGASVVCC(1287.5302)FLNNFYPK+Incorrect Mono 3Da HNSYTC(3667.7981)EATHK+Incorrect Mono 3Da	23.6	4.1e-8	6.2	113	194	147	204	1	139	1





Access to Annotated spectra of Cross-linking Study Results

Sharing and dissemination of cross-linked peptide identifications is currently difficult, as community standard formats (e.g. mzIdentML) do not support reporting of cross-linked results and there is no proteomic repository that can display these type of results.

MS-Viewer is spectral viewing software that is part of Protein Prospector.[3] It provides a web portal where anyone can upload, share and view proteomic results. It supports annotation of crosslinked peptide identifications.

Annotated spectra from the three datasets discussed on this poster can be browsed and viewed:

http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer

Dataset Search Key:

UTP-B: eg9vql948
E. coli: 9w2x3pqwmi
mAb: 3z1vkx1pgt

There are video tutorials that describe how to use Protein Prospector, including how to perform cross-linking analysis and use of MS-Viewer:

https://vimeo.com/channels/194363/

Conclusions

- Protein Prospector supports free online cross-linked peptide searching analysis.
- By evaluating results reliability based on less confident of two peptide identifications it is better able to control the FDR rate while maintaining high sensitivity.
- By reporting Inter- and intra-protein crosslinks separately one is able to draw different score thresholds for each:
  - Intra-protein matches are much less likely to be random, so a more liberal threshold can be drawn compared to inter-protein matches.
- Protein Prospector was able to show that some of the disulfide linkages reported in the literature accompanying Waters Intact mAb LC/MS standard are incorrect.

If you have any questions, or want help trying the software please contact us:

ppadmin@cgl.ucsf.edu

Acknowledgements

We thank Drs He and Dong for access to their raw data.
his work was supported by NIH NIGMS grant 8P41GM103481.

References

1. Yang *et al.* Nat Methods (2012) **9** 9 904-6
2. Baker *et al.* Mol Cell Proteomics (2014) **13** 2 420-34
3. Baker and Chalkley Mol Cell Proteomics (2014)**13** 5 1392-6