Discovery of Unanticipated Modifications using Protein Prospector

Robert J. Chalkley, Peter R. Baker, Katalin F. Medzihradszky and A. L. Burlingame

NIH NCRR Mass Spectrometry Resource, University of California San Francisco

Introduction

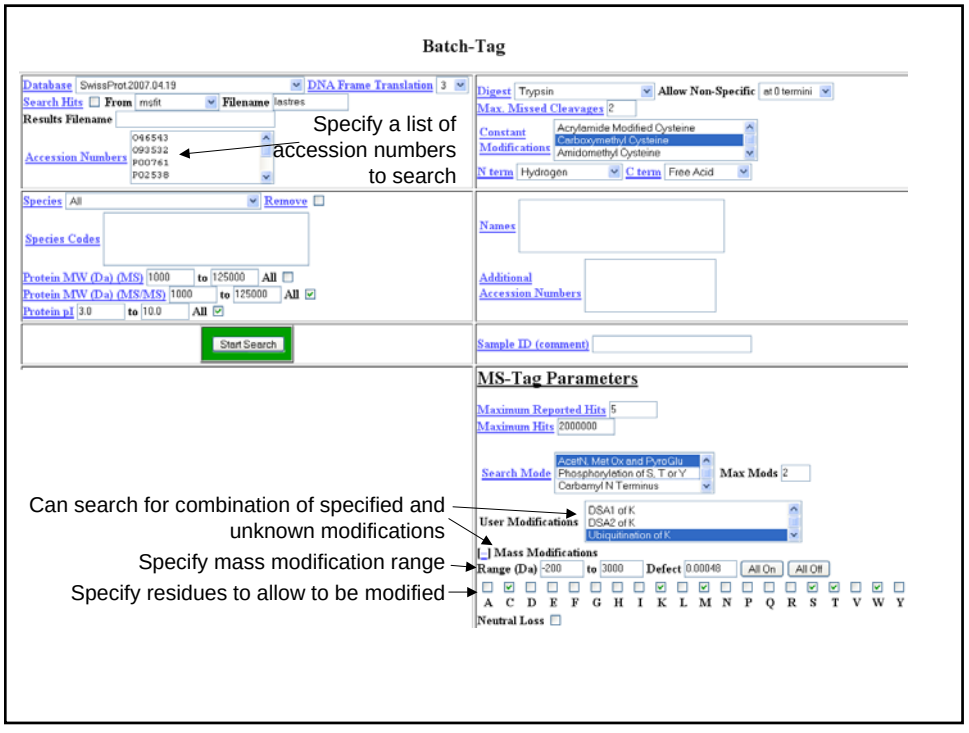
- In LCMSMS analyses large numbers of spectra are not identified by standard database searching strategies.
- Several of these spectra correspond to peptides with unanticipated modifications.
  - This is especially true when some proteins are present at very high abundance relative to others.
- It has been proposed that there are 8-12 modified versions (mainly chemical/artifactual) of each unmodified tryptic peptide.[1]
- If you want to fully characterize your sample you need to be able to explain as much of your data as possible.
  - What is the best strategy to try to identify these peptides?

- Strategies that have been proposed:
  - Identify 'good looking' spectra and search only these for more modification types.[2]
  - Compare unidentified spectra to those identified to find families.[3]
  - Perform database searching allowing for unexpected modifications, searching only against those peptides[4] or proteins[5,6] identified in a standard database search.
  - With very high mass accuracy data *de novo* interpretation.[7]

Poster Overview

•In this poster we present the performance of unexpected modification searching using Protein Prospector:

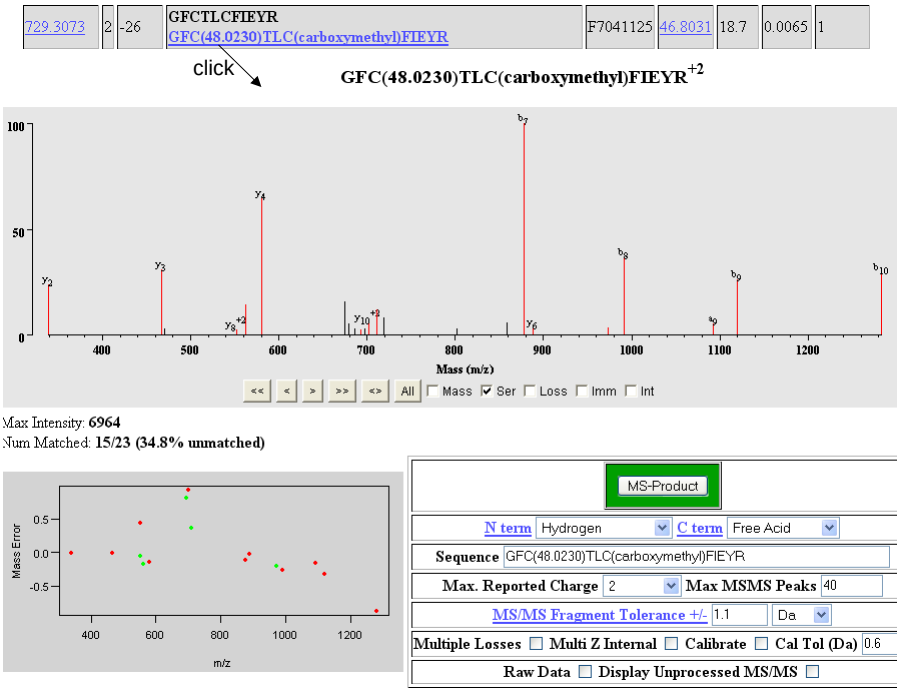
- •How is the search/analysis performed?
- •What does it find?
- •How does it compare to other software?
- •What are the advantages of using Protein Prospector for doing this type of search?

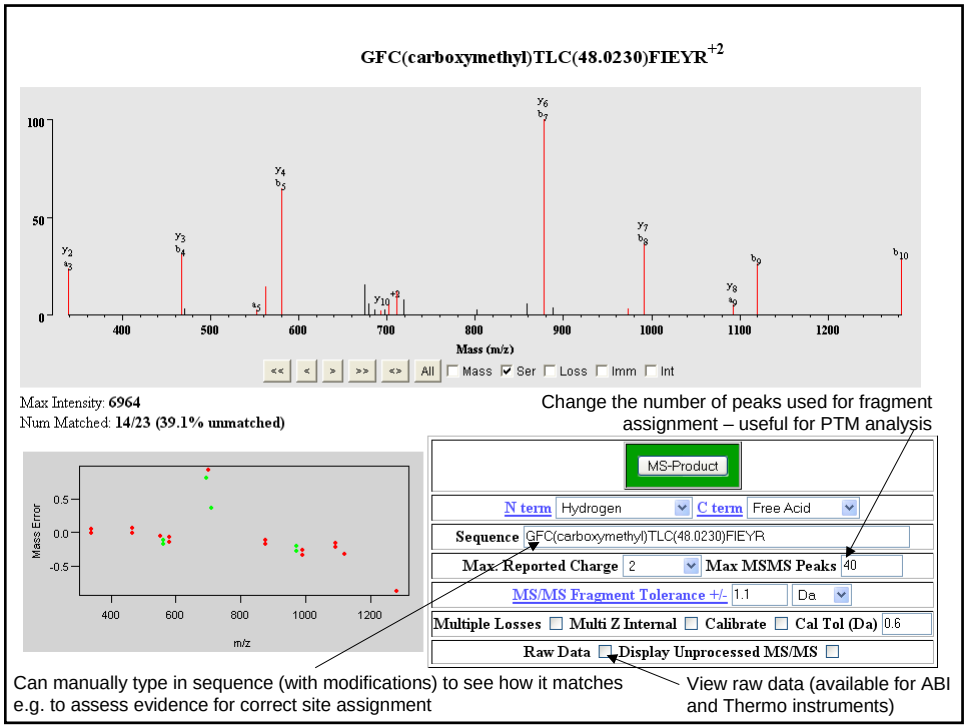


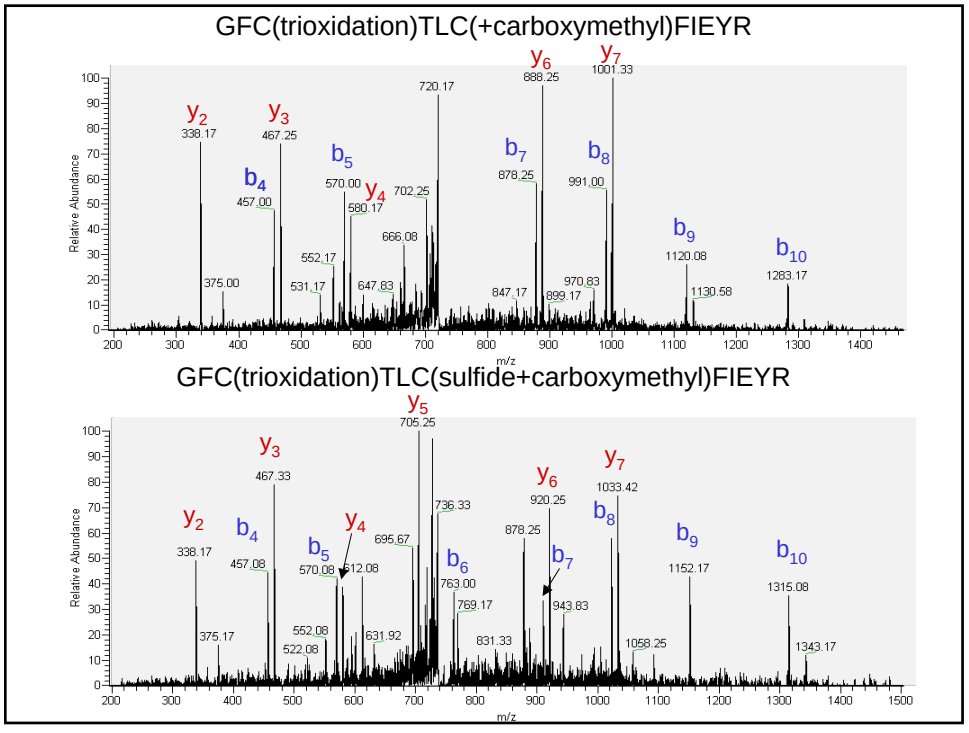
Heterogeneity of Cysteine Modifications

•Cysteines pick up many different modifications.
•Peptides containing two cysteines can be very heterogeneous.
  •Single peptide identified in 11 different forms:

GFC(dehydro)TLC(dehydro)FIEYR (internal disulfide)
GFC(carboxymethyl)TLC(carboxymethyl) FIEYR
GFC(propionamide)TLC(carboxymethyl)FIEYR
GFC(carboxymethyl)TLC(trioxidation)FIEYR
GFC (trioxidation)TLC(carboxymethyl)FIEYR
GFC(trioxidation)TLC(sulfide+carboxymethyl)FIEYR
GFC(oxidation+carboxymethyl)TLC(carboxymethyl)FIEYR
GFC(trioxidation)TLC(Cys->Dha)FIEYR
GFC(propionamide)TLC(sulfide+carboxymethyl)FIEYR
GFC(Cys->Dha)TLC(carboxymethyl)FIEYR
GFC(carboxymethyl+DTT)TLC(carboxymethyl)FIEYR




GFC(carboxymethyl)TLC(48.0230)FIEYR+2
Mass (m/z)
All Mass Ser Loss Imm Int
Max Intensity: 6964
Num Matched: 14/23 (39.1% unmatched)
Mass Error
m/z
Change the number of peaks used for fragment assignment – useful for PTM analysis
MS-Product
N term Hydrogen
C term Free Acid
Sequence GFC(carboxymethyl)TLC(48.0230)FIEYR
Max Reported Charge 2
Max MSMS Peaks 40
MS/MS Fragment Tolerance +/- 1.1 Da
Multiple Losses Multi Z Internal Calibrate Cal Tol (Da) 0.6
Raw Data Display Unprocessed MS/MS
Can manually type in sequence (with modifications) to see how it matches e.g. to assess evidence for correct site assignment
View raw data (available for ABI and Thermo instruments)

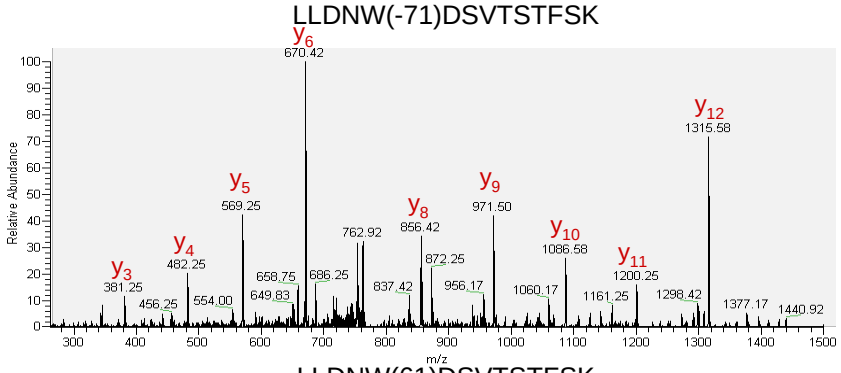

GFC(trioxidation)TLC(+carboxymethyl)FIEYR
GFC(trioxidation)TLC(sulfide+carboxymethyl)FIEYR
Relative Abundance
m/z

Tryptophan Modification

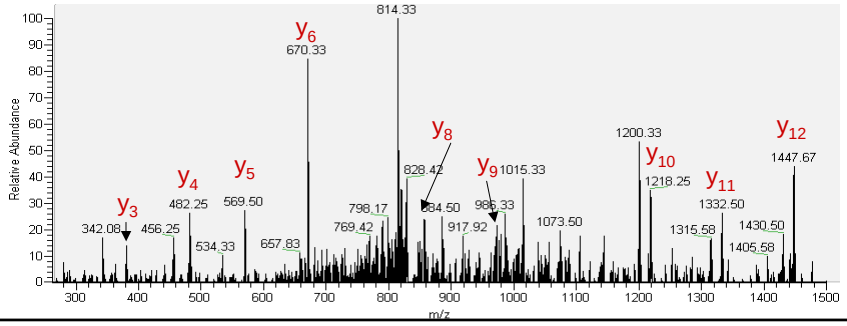
•Tryptophan is also very prone to modification:
e.g.

LLDNWDSVTSTFSK
LLDNW(oxidation)DSVTSTFSK
LLDNW(dioxidation)DSVTSTFSK
LLDNW(Trp->Kynurenin)DSVTSTFSK
LLDNW(-71)DSVTSTFSK
LLDNW(+61)DSVTSTFSK

LLDNW(-71)DSVTSTFSK



LLDNW(61)DSVTSTFSK

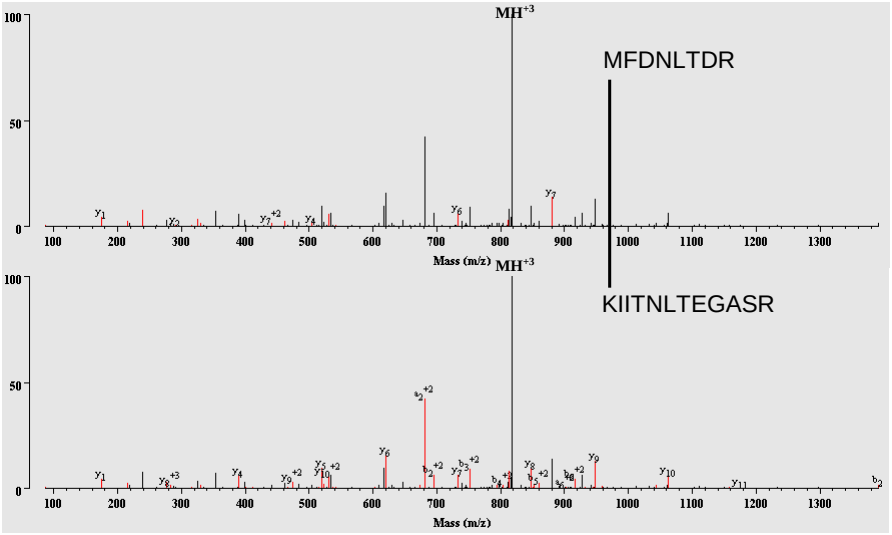


Identification of Cross-linked Peptides

•Data from a study of protein binding interfaces using cross-linking[8].

546.8028	2	0.0024	DAEALYGLLK	33.13	29.3	18.6	7.7e-4	3.39
1029.9161	3	0.036	LDGTAKGGVIFSVADQFGIPIR LDGTAK(826.3965)GGVIFSVADQFGIPIR	40.28	26.6	18.6	1.6e-5	3.39
868.4754	3	0.0062	LDGTAKGGVIFSVADQFGIPIR LDGTAK(342.1642)GGVIFSVADQFGIPIR	47.29	23.9	18.6	5.6e-6	3.39
806.4207	3	-0.037	RSLLKTKENLGSGFISLFR RSLLK(251.1205)TKENLGSGFISLFR	37.51	27.9	18.3	1.5e-5	3.36
817.7567	3	-0.011	KIITNLTEGASR K(1148.5510)IITNLTEGASR	34.25	26.5	17.6	3.5e-4	3.28
832.7576	3	-0.013	IEDLRPFKADDFIEALFAR IEDLRPFK(230.1104)ADDFIEALFAR	45.5	23.4	17.2	4.1e-6	3.24
770.9197	2	-0.0098	KIITNLTEGASR K(238.1142)IITNLTEGASR	18.77	22.8	17.1	3.3e-6	3.23
518.2642	2	0.0092	EEMGEILAK EEmGEILAK	22.54	22.0	16.4	2.0e-4	3.15
424.2013	2	4.7e-4	QFEQQGK qFEQQGK	16.81	25.4	16.1	1.6e-4	3.12

Clicking here will allow you to re-search this one spectrum with different search parameters; e.g. allow for a modification of (817 x 3 =) 2450 Da to try to identify the other peptide.



MS-Bridge Assignment of Cross-linked Peptides (from different proteins)

1	1	8	0	(-)MFDNLTDR(L)
2	247	258	1	(R)**K**IITNLTEGASR(K)

Comparison of Protein Prospector Mass Modification Searching with Alternative Software

•InsPecT is freely available software designed for finding modified peptides in iontrap MSMS data.[9]
  - •It works by finding sequence tags and has no bias towards particular modifications.

•We compared mass modification searching in Protein Prospector to InsPecT results of the same dataset.

Software Results Comparison

•The dataset: One LCMSMS analysis of a gel-purified modified protein.
•The sample was initially searched without looking for modifications and 17 database entries were subsequently used for the modification searches using Protein Prospector and InsPecT.
•Number of spectra acquired: 521.
•Number of unmodified spectra identified (PP E-value <0.01): 73

•After modification searching:
•Number of assignments Prospector and InsPecT agree completely upon: 116
•Number agree on modification but not site: 27
•Number agree on peptide but not modification: 37

•Both search engines returned 163 matches with E-value/P-value <0.3, of which the majority of spectra uniquely identified by one or other were due to slightly different search parameters.

•Conclusion: the two different searching and scoring strategies produce essentially the same assignments, with neither being significantly more sensitive/reliable than the other.

Where Protein Prospector is better

- Protein Prospector allows the user to search for a combination of specified and unexpected modifications: e.g. search allowing for methionine oxidation and one unknown modification:
  - InsPecT only allows searching for either specified or unknown
- Hence, Protein Prospector is able to more reliably identify multiply modified spectra and more consistently assign the correct site where a common modification is present.
- In Protein Prospector you can specify to allow modification to only certain amino acids*:
  - Reduction in 'search space' can give more confident and reliable answers whilst still identifying unexpected modifications.
- Protein Prospector user-friendly interface allows easier assessment and verification of results.

*not utilized in this software comparison

Conclusions

- Protein Prospector is able to identify many MSMS spectra to peptides with modifications:
  - It does find many biological modifications as well as the chemical modifications highlighted in this poster; e.g. ubiquitination, tyrosine nitration…
  - In some cases the 'modification' is an unexpected cleavage site.
  - Modification can be large (a cross-linked peptide).
  - Site assignment is much less reliable than peptide and modification assignment.
- Certain residues account for a large percentage of all the modifications, most notably cysteines and tryptophans.
- Protein Prospector allow easy assessment and manual verification of assignments.

- Performance of software is at least as good as InsPecT (most other alternative software is not freely available).

- A new version of Protein Prospector (including mass modification searching) will be available to the public in the next month or two!

References

1. Nielsen, M. L., Savitski, M. M., and Zubarev, R. A., Mol Cell Proteomics, 2006. **5**(12): p. 2384-91.
2. Nesvizhskii, A. I., Roos, F. F., Grossmann, J., Vogelzang, M., Eddes, J. S., Gruissem, W., Baginsky, S., and Aebersold, R., Mol Cell Proteomics, 2006. **5**(4): p. 652-70.
3. Tsur, D., Tanner, S., Zandi, E., Bafna, V., and Pevzner, P. A., Nat Biotechnol, 2005. **23**(12): p. 1562-7.
4. Savitski, M. M., Nielsen, M. L., and Zubarev, R. A., Mol Cell Proteomics, 2006. **5**(5): p. 935-48.
5. Searle, B. C., Dasari, S., Turner, M., Reddy, A. P., Choi, D., Wilmarth, P. A., McCormack, A. L., David, L. L., and Nagalla, S. R., Anal Chem, 2004. **76**(8): p. 2220-30.
6. Baker, P. R., Medzhiradszky, K. F., and Burlingame, A. L. in *Proceedings of the 54th ASMS Conference*. 2006. Seattle, WA.
7. Frank, A. M., Savitski, M. M., Nielsen, M. L., Zubarev, R. A., and Pevzner, P. A., J Proteome Res, 2007. **6**(1): p. 114-23.
8. Chu, F., Shan, S. O., Moustakas, D. T., Alber, F., Egea, P. F., Stroud, R. M., Walter, P., and Burlingame, A. L., Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16454-9.
9. Tanner, S., Shu, H., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., and Bafna, V., Anal Chem, 2005. **77**(14): p. 4626-39.

Acknowledgements

This work was supported by NIH NCRR grants 001614, 0012961, 0019934 and the Vincent J. Coates Foundation.