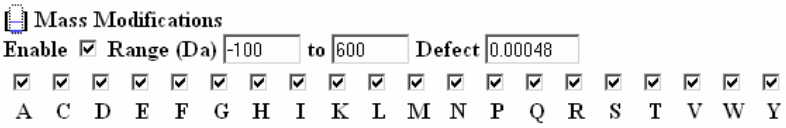
Improved Methods for Comprehensive Sample Analysis Using Protein Prospector

Peter R. Baker[1], Katalin F. Medzihradszky[1] and Alma L. Burlingame[1]

[1] Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of California, San Francisco, USA

Algorithm for Matching Additional Spectra

An algorithm has been written that can consider a range of integer mass shifts on any selected amino acid. For example for the 3000 spectra in the test dataset it can consider any mass shift between -100 Da and +600 Da on 2 proteins in under 1 minute on a standard dual processor computer. This is equivalent to looking for 14000 different modifications at the same time. The mass defect can also be changed.



Experimental

KatG Overexpression and Purification – His-tagged Mycobacterium tuberculosis catalase/peroxidase mutants were expressed in *Escherichia coli.* Cells were lysed by sonication in the presence of protease inhibitors, at 4°C. Cellular debris was pelleted at 30,000 *g* for 30 min, resulting in a viscous, red-brown crude extract, which was first purified by affinity chromatography employing Ni-NTA-agarose. Fractions containing KatG, identified by their red/brown color and the presence of an 85-kDa band on a 10-12% SDS-polyacrylamide gel, were pooled and concentrated using an Amicon Diaflow concentrator equipped with a 50,000–molecular weight cutoff membrane (BioMax-50). The concentrate was further purified by size-exclusion chromatography. Fractions containing KatG exhibiting no impurities by SDS-PAGE analysis were pooled and concentrated as above. [Ghiladi *et al*., *J. Am. Chem. Soc.* **127,** 13428-13442 (2005).]

Tryptic digestion and mass spectrometry – the mutant recombinant proteins were in-gel digested with side-chain protected porcine trypsin without reduction and alkylation. The digests were subjected to nano-LC/MS/MS experiments. Peptides were fractionated in 0.1% formic acid-containing mobile phase, on a 75 μm ID C18 column utilizing an Eksigent pump coupled with a Spark autosampler. Data were acquired on a QSTAR XL mass spectrometer in a data-dependent mode: 1 sec MS surveys were followed by 3 sec CID analyses on computer-selected multiply charged ions.

Introduced Protein Modifications

Mutations in the *KatG* enzyme
Sample 1: Tyr-229 → “acetyl”-Tyr
Sample 2: Tyr-229 → “amino”-Tyr
Sample 3: Tyr-229 → “azido”-Tyr
Sample 4: Tyr-229 → “iodo”-Tyr

In each of these unusual amino acids the hydroxyl group of the tyrosine was replaced with the substituent indicated

Sample 5: Met-255 → Cys

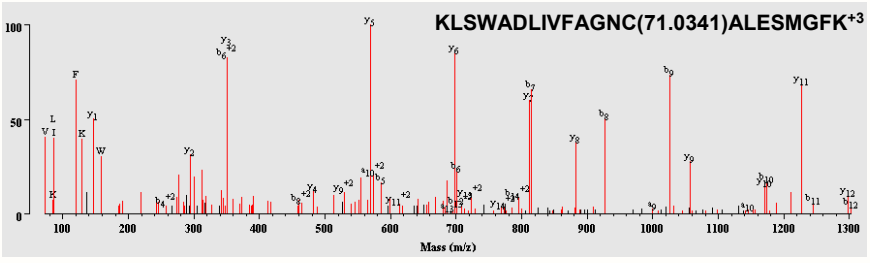
Also in all the above proteins a Gly-234 → Ala mutation occurred.

Test Data Set

The test data set is an LC-MS analysis of 5 *Mycobacterium tuberculosis* catalase/peroxidase mutants. Around 3000 Q-STAR spectra were collected in 5 LC/MS analyses.
An initial database search identified the KatG protein as well as a contaminating *E coli* protein. A second search was then done restricting the search to these 2 proteins and looking for a standard set of modifications (acetylated N-terminus, pyroglutamic acid and oxidized methionine).
The second search resulted in 184/606 (30.4%) of the spectra being matched in fraction 1 with expectation values less than 0.0001.
A No Enzyme search was able to match a further 19 spectra (3.1%).
Use of the new algorithm enabled an extra 292 (48.2%) of the spectra to be matched. 111 (18.3%) of the spectra remain uninterpreted.

Some Modifications Found

- Oxidation and dioxidation of Trp (+16, +32)
- Trp oxidation to kynurenin (+4)
- Acrylamide modified Cys (+71)
- Deamidation of Asp (+1)



Acrylamide Modified Cysteine

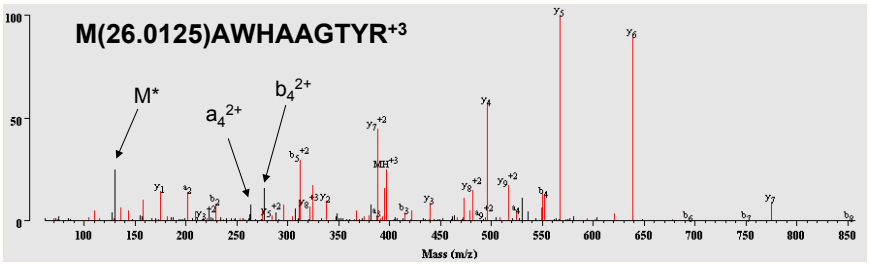
Mysterious N-terminal Modification

+26 - + C_2H_2? = 26.0157
CH_2=CH-NH~ or CH_3-CH=N~

Other two options:
→ O=C=N~ that would add 25.9793
→ N=N=N~ that would add 25.9906
– however these are more reactive species, plus mass wise do not fit.

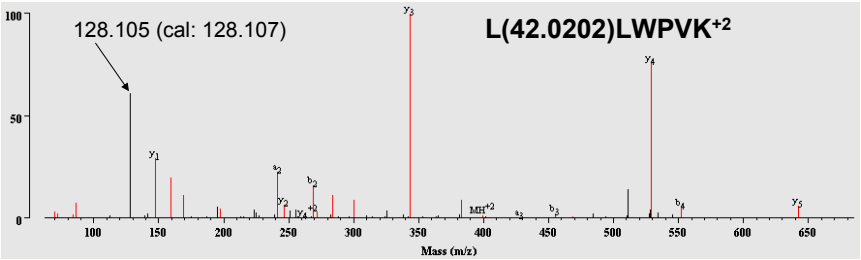
The modified peptides elute later than their unmodified counterparts
~125 examples in these 5 runs!

Example of +26 Modification



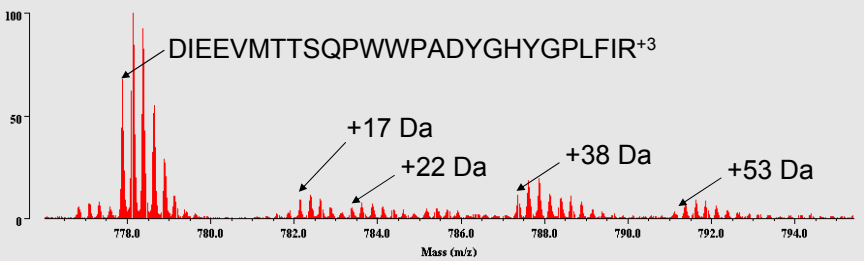
Variation On The Same Theme

NO b_1 ion, BUT an abundant modified immonium ion → NOT acetyl
- maybe HO-CH_2-CH=N~ ?

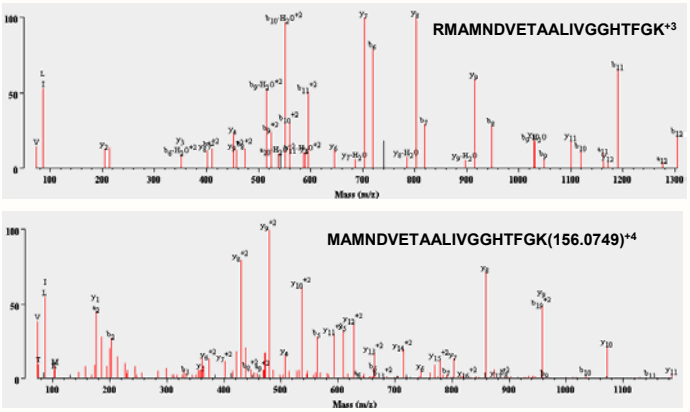


Adducts

Adducts found by the algorithm include Ammonium (+17 Da), Na (+22 Da), K (+38 Da) and Fe (+53 Da).



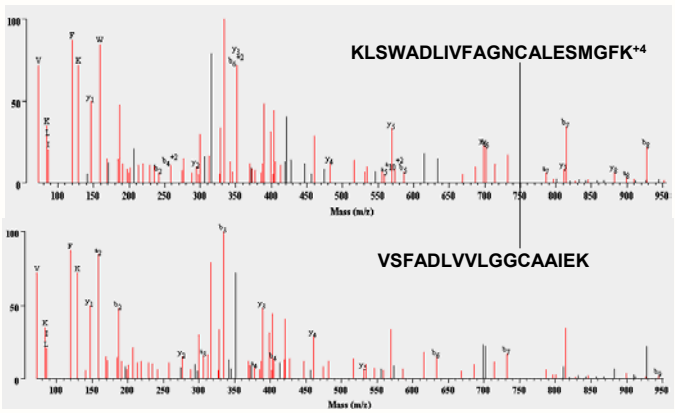
Enzyme Catalyzed Rearrangement Reaction [1, 2]



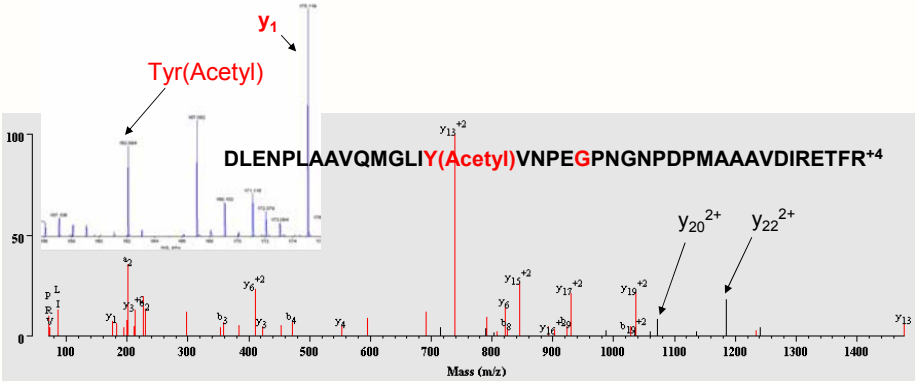
Normally the Arg is adjacent to the Met as in the first spectrum. The second example however shows an additional Arg at the C-terminus.

Disulfide Bridges

The new algorithm can be used in conjunction with Protein Prospector program MS-Bridge to find peptides with disulfide bridges. Mass shifts between 0-2000 Da were considered. 3 disulphide combinations were found. An example is given below (1023.3276^{+4}).



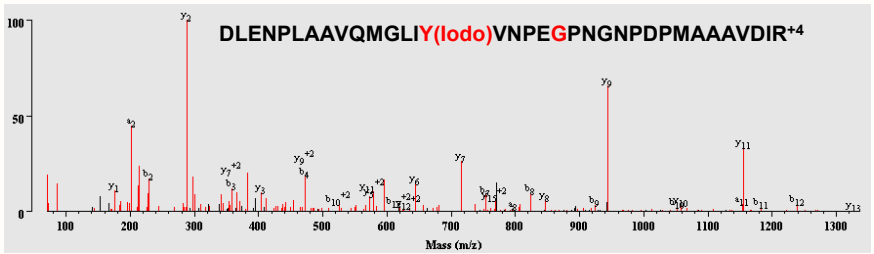
Mutant Peptides Sample 1



926.23 (4+) is DLENPLAAVQMGLIYVNPEGPNGNPDPMAAAVDIR + 40.02
1059.56(4+) is DLENPLAAVQMGLIYVNPEGPNGNPDPMAAAVDIRETFR + 40.02
That is "acetyl"-Tyr, ie. Y – OH + CH_3CO = +C_2H_2 = +26.015 Da
Gly → Ala mutation + CH_2 = +14.015 Da
→ +40.03 Da

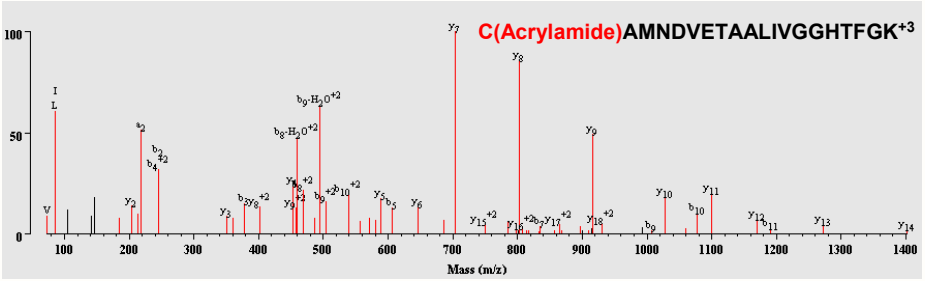
In this case y_{20}^{+2} and y_{22}^{+2} don't match as the algorithm is only currently considering a single modification per peptide.

Mutant Peptides Sample 4



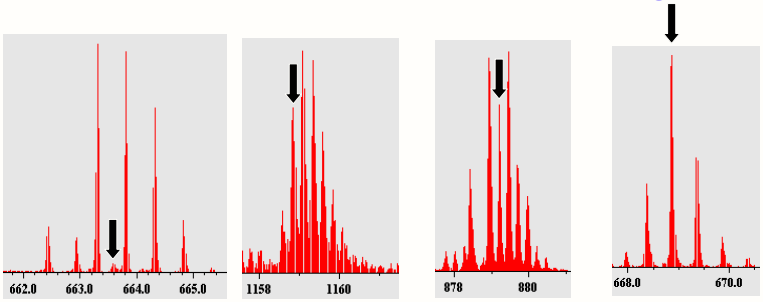
947.24 (4+) is DLENPLAAVQMGLIYVNPEGPNGNPDPMAAAVDIR + 124.06
That is "iodo"-Tyr, ie. OH + I = +109.9017 Da
Gly → Ala mutation = +CH_2 = +14.015 Da
→ +123.917 Da

Mutant Peptides Sample 5



702.37 (3+) is MAMNDVETAALIVGGHTFGK + 43.021 Da
Met → Cys mutation: $-C_2H_4$
Acrylamide addition: $+C_3H_5NO$
→ +CHNO = +43.006 Da

Incorrect Monoisotopic Mass Assignment

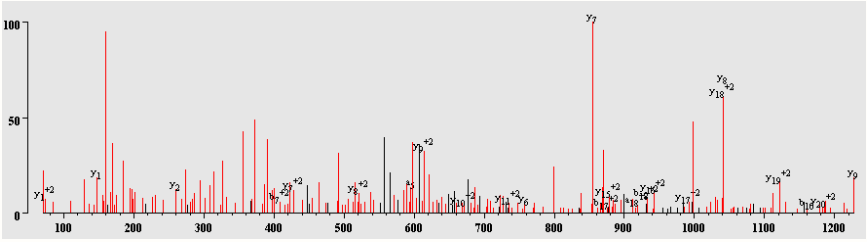


The arrows show the monoisotopic precursor m/z value returned by the Q-STAR precursor centroiding algorithm. The last case is a mixture of a deamidated and non-deamidated peptide.

The new algorithm was able to confidently match all the corresponding MSMS spectra.

Incorrect Charge Assignment

If a large enough mass range is allowed the new algorithm can also match peptides where the precursor centroiding algorithm assigned an incorrect charge. For example it obtained matches for peptides with charges of 6 and 7 which the centroiding algorithm can't assign automatically.



MPEQHPPITETTTGAASNGC(185.0888)PVVGHMKYPVEGGGNQDWWPNRLNLK^{+6}

In this case only the y ions are matching so the exact hit needs to be worked out manually.

Conclusions and Future Work

The algorithm was able to increase the number of spectra matched from around 30% to around 80%. It can be used to identify novel modifications and experimental problems.

Once the modifications have been catalogued they can be programmed into a more efficient algorithm for routine searches.

We need to investigate how the mass tolerance used in the search affects the results returned.

Currently the algorithms output is a peptide and a mass shift. We hope to use a multi-stage search to allow several modifications per peptide and to improve the modification site assignment.

Acknowledgements

- Funding: NIH NCRR grant 01614 and the Vincent Coates Foundation.
- Dr. Reza Ghiladi provided the KatG samples

- Refs.:

1) Lippincott, J. *et al.* (1997) Mapping of recombinant hemoglobin using immobilized trypsin cartridges. *Anal. Biochem.* **252,** 314-325.

2) Fodor, S. and Zhang, Z. Protease-catalyzed rearrangement of terminal amino acid residues in peptides. 51st ASMS Conference, Montreal, Quebec, Canada (2003)

http://www.inmerge.com/aspfolder/ASMSAbstracts.html. (Abstract: 406)