Identifying and locating PTMs in complex peptides or proteins based on acquired high resolution Full-Scan and MS/MS data

Ray Fyhr[1]

[1]Discovery & Pre Clinical IT, Merck Research Laboratories, Rahway, NJ 07065

Abstract

Up until recently most mass spectrometry biomarker discovery strategy focused on small peptide fragments ignoring the post translational landscape of larger peptides and intact proteins. Top down proteomics analyzes the intact protein and all its post translational modification in one single run. Here we describe an extension to a new top-down proteomics algorithm developed at Merck called **MAR** (1). The software runs on Linux clusters, relies only on a predefined list of 'differential' modifications (2) (e.g., phosphorylation) and a FASTA-formatted protein database, and is not constrained to full-length proteins for identification. The added functionality to the recently published work (Mazur, Fyhr, RCMS, 2011) elaborates on techniques that locate post-translational modifications within high scoring candidate polypeptide matches. These candidates are then further scored to determine the location of the modified residue. Currently the software is equipped to find a single PTM location within a polypeptide but the design is capable of being expanded to find multiple modifications. The application of these new developments for protein id may be very useful in areas such as neuroproteomics and neurology.

Objective

Create a software tool to convert high resolution MS/MS data into peptide identifications with PTM ID and residue location capabilities

- Leverage off the software architecture of the MAR algorithm (1)
- Employ a simple fasta-formatted protein database structure
- Allow pre-defined "differential" modifications for searching (2)
- Eliminate intact protein or enzymatic restrictions
- Consider high mass accuracy data for scoring (4)
- Perform full-scan surveying to determine high probability PTMs
- Make it parallelizable for high performance
- Develop new functionality to locate the residue within the peptide

Methods

Raw Data Files: 9 yeast samples were prepared with a Top Down protocol and analyzed using a 12 T LTQ-FT Ultra mass spectrometer (Thermo Fisher)(5).

LC-MS spectral de-isotoping: Full-scan MS and MS/MS spectra contained within the yeast .raw files were converted to lists (_isos.csv format) of monoisotopic, neutral masses using the Horn transformation function of the publically available program Decon2LS(3). (PNNL)

MAR Algorithm: Software and Structure

mar_index – is used repeatedly in MAR to take a .dat file and create a binary file of the amino acids in the proteins and an index file to enhance performance.

mar_ptmdiff – reduces computation time by discovering matches between the experimental data from the full-scans and the theoretical list of PTMs.

mar_NoE – uses the precursor mass of the fragmented ion and extracts all polypeptide sequences that match this value (within a given mass tolerance).

mar_ions - matches the experimental ms2 fragment ions with in silico generated theoretical fragments of the candidates, then scored(1) (4)

mar_locate – takes the individual winning candidates that contain a PTM and generates all permutations of potential new candidates replacing occurrences of the amino acid being modified with a synthetic residue equal to the sum of mw of the amino acid and the mw of the PTM. This .dat file containing all the permutations is run through mar_index and mar_ions and scored like before.

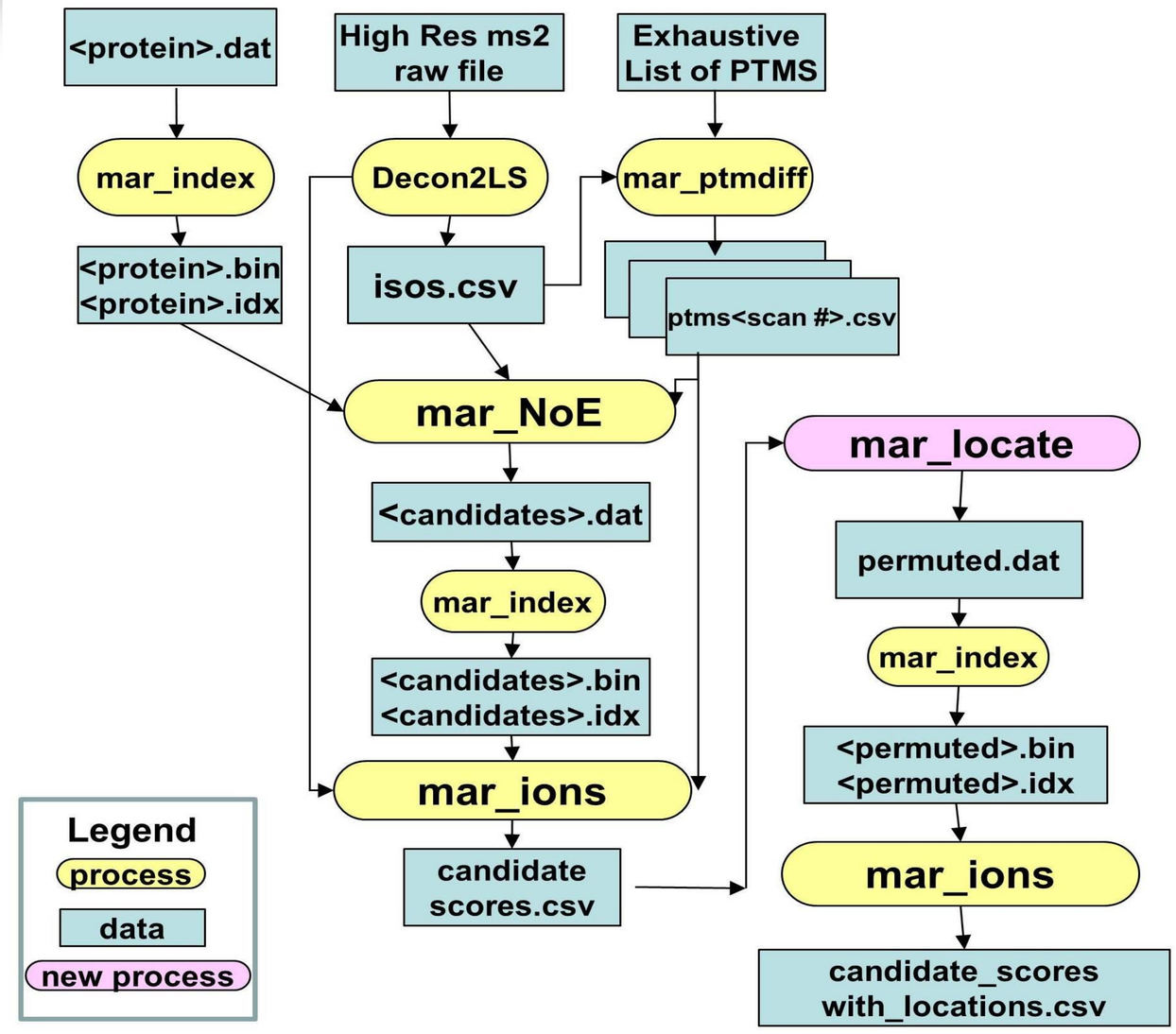

Figure 1. MAR Algorithm Architecture

Results

Table 1. Abbreviated list of some top polypeptide identifications

P-score	B/C ion	Y/Z ion	polypeptide	mw	mw diff	PTM mw	PTM description	aa RES
3.87E-59	29	50	TRX2_YEAST_NoE_2_104	11063.5735	0.0273	-2.0146	C 1 disulfide bridge	30
6.06E-57	38	45	G3P3_YEAST_NoE_240_332	10149.2956	-0.9718	0.9840	N Deamidated	9
4.03E-41	22	43	RS28B_YEAST_NoE_1_67	7602.1650	-0.0191	42.0106	K Acetyl	11
7.49E-33	26	31	G3P3_YEAST_NoE_240_332	10150.2809	0.0135	0.9840	N Deamidated	9
8.90E-28	30	23	HSP12_YEAST_NoE_2_109	11596.6444	0.0154	42.0106	K Acetyl	11
1.82E-20	17	34	G3P3_YEAST_NoE_208_332	13372.0599	-0.0033	14.0157	L methyl	101
6.95E-16	19	16	SDO1L_YEAST_NoE_2_111	11913.1289	0.0169	42.0106	K Acetyl	5
1.57E-14	4	40	MAL12_YEAST_NoE_496_584	10427.5542	-0.6747	-17.0265	Q Q pyroglutamic ac	18
4.52E-10	7	23	ENO1_YEAST_NoE_320_437	12646.6507	-0.9759	0.9840	Q Gln->Glu	42
1.34E-08	16	16	G3P3_YEAST_NoE_283_332	5573.7878	0.0095	14.0157	L methyl	43

Table 2. The top 10 scores for highlighted scan above

P-score	B/C ions	Y/Z ion	polypeptide	mw	mw diff	PTM mw	PTM description	aa RES
6.06E-57	38	45	G3P3_YEAST_NoE_240_332	10149.2956	-0.9718	0.9840	N Deamidated	9
2.07E-22	44	8	G3P3_YEAST_NoE_240_332	10149.2956	-0.9718	0.9840	N Deamidated	75
1.86E-18	9	39	G3P3_YEAST_NoE_240_332	10149.2956	0.0122	0.0000		
5.38E+03	8	14	G3P3_YEAST_NoE_240_332	10149.2956	-0.9718	0.9840	Q Gln->Glu	61
9.24E+04	8	12	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	S water loss	7
9.24E+04	8	12	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	T water loss	2
3.57E+05	8	11	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	S water loss	33
3.57E+05	8	11	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	S water loss	41
3.57E+05	8	11	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	S water loss	45
3.57E+05	8	11	SLM2_YEAST_NoE_129_217	10149.2956	-1.0228	-18.0106	T water loss	32

Conclusions

- Developed a first version of the PTM locator extension to the MAR algorithm (1).
- Results:
  - 9 raw files with total of 628 ms2 scans
  - 123 Identifications with P-Scores < $1e^{-5}$
  - 54 of those had PTMs which were algorithmically located.
- Learned how to handle disulfide bridges by adding them as pseudo PTMs.
- Learned how to adjust the run tolerances to find additional IDs with good P-scores.
- Better scoring method may exist that also makes use of the ion matching variances.
- Results demonstrate a promising complement to existing software tools for protein ID

References

1. Mazur MT, Fyhr R. Rapid Commun Mass Spectrom. 2011 Dec 15;25(23):3617-26. doi: 10.1002/rcm.5257
2. Creasy, D. M.; Cottrell, J. S. Unimod: Protein modifications for mass spectrometry. Proteomics 2004, 4 (6), 1534-1536.
3. Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics 2009, 10:87.
4. Meng, F.; Cargile, B.J.; Miller, L.M.; Forbes, A.J.; Johnson, J.R.; Kelleher, N.L. Nat.Biotechnol. 2001, 19 (10), 952-957
5. Kellie JF, Catherman AD, Durbin KR, Tran JC, Tipton JD, Norris JL, Witkowski CE 2nd, Thomas PM, Kelleher NL. Anal Chem. 2012 Jan 3;84(1):209-15. Epub 2011 Dec 14.