PosterREPRINT

Waters

# THE USE OF SEARCH WORKFLOWS IN PEPTIDE ASSIGNMENT FROM MS/MS DATA

*Richard Denny, Kieran Neeson, Catriona Rennie, Keith Richardson, Steve Leicester, Neil Swainston, Jennifer Worroll, Phillip Young.*
*Waters Corporation, Manchester, UK*

**Presented at ABRF 2002, Austin, Texas, USA, 9th - 12th March 2002**

## Introduction

Whilst significant advances have been made in the identification of tryptic peptides and thus proteins from tandem mass spectrometry data, it is apparent that many spectra remain unassigned. Reasons for lack of assignment of the spectra include databank incompleteness, point mutations, post translational modifications not selected by the user, and non-specific cleavage by trypsin.

In an effort to increase the assignment rate of tandem mass spectrometry data to peptides, we have implemented a new algorithm known as Automod, which on input of protein sequences, generates all possible peptides from a sequence, and then using a library of post-translational modifications and point mutations attempts to assign tandem mass spectrometry data to the generated peptides. Furthermore, we have extended our de novo peptide sequencing algorithm to include batch processing, so that more than more spectrum can be interpreted in a search.

By using the same peptide fragmentation model and scoring scheme for databank searching, automod and de novo sequencing (**Figure 1**), we are able to chain these different search strategies together into a data interpretation workflow. In this poster we present example data illustrating the power of this approach.

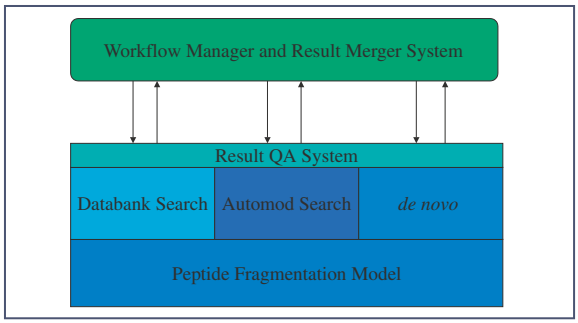


*Figure 1. Schematic of workflow system*

## Method

A two protein mix consisting of pyruvate kinase and fructose bi-phosphate was digested using trypsin.

The tryptic digest mixture was directly injected, with 1uL of the solution applied to the LC-MS/MS set-up. The Micromass CapLC system initially flowed at a flow rate of 30µL/min whilst the sample was injected and pre-concentrated on a 320µMx1mm pre-column. The peptides were then eluted onto an analytical $C_{18}$ PepMap column (150mmx75µM) and a gradient was run from 5% acetonitrile: 95% water (both containing 0.1% formic acid) to 60% acetonitrile after 40 minutes at a flow rate of 200nL/min. The LC eluent from this column then entered the Nanoflow source of the mass spectrometer.

The Q-Tof *Ultima* API mass spectrometer was operated in positive ion mode with a source temperature of 80°C, a counter current gas flow rate of 50L/hr and with a potential of 3400 V applied to the Nanoflow probe body. Both MS and MS/MS data was acquired with the TOF analyser with data integrated every second.

MICROMASS MS TECHNOLOGIES

Analysis of the sample by MS/MS was achieved in the data directed analysis mode (DDA). In this mode of operation the Q-Tof automatically selected precursor ions based upon ions rising above a predetermined intensity threshold and possessing a charge state of 2+ or 3+. At one particular decision time interval the instrument would consider selecting 2 precursor ions for MS/MS. In the MS/MS mode the quadrupole was used in resolving mode to select the precursor ion of interest, for fragmentation in the hexapole collision cell. MS/MS was performed with argon gas in the collision cell at a pressure of $6x10^{-5}$ mbar measured in the analyser. MS/MS data was acquired for a maximum of 4 seconds per component.

The resulting spectra were processed using the MaxEntLite algorithm which performs de-convolution, centering and de-isotoping in one step. Using a small databank consisting of pyruvate kinase and trypsinogen we designed a search workflow consisting of a databank search, Automod search and de novo peptide sequencing search (**Figures 2 - 4**). At each stage the results of the previous stage were filtered on peptide score. Any precursor masses that remained unmatched to peptides were used in searching.

The workflow was submitted to the ProteinLynx search engine. The sequences obtained in the *de novo* step were then searched for homology in the Swissprot databank using BLAST at NCBI.

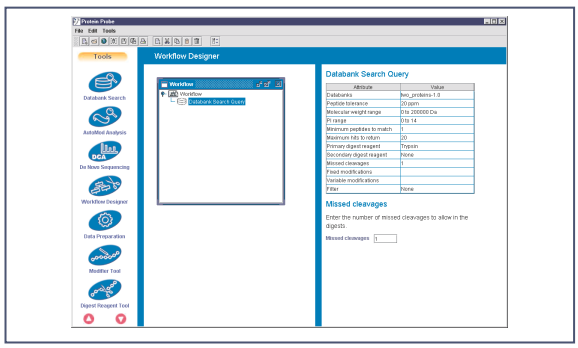


*Figure 2. Setting up the databank search in ProteinProbe*

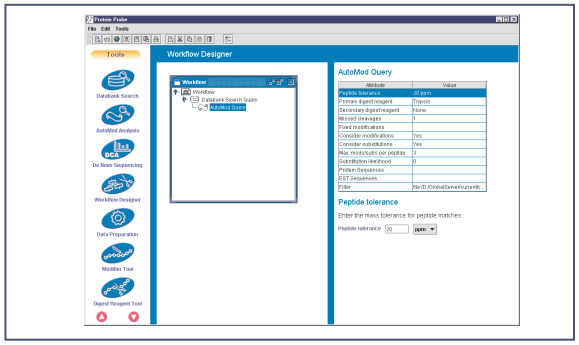


*Figure 3. Setting up the Automod search in ProteinProbe*

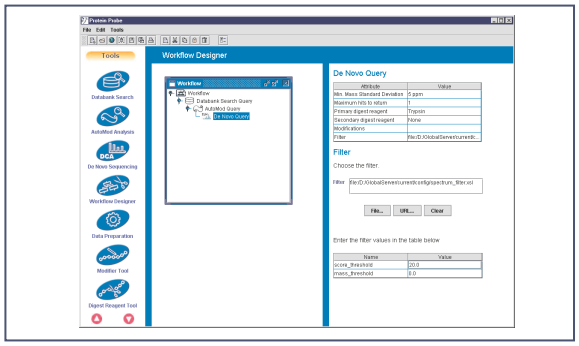


*Figure 4. Setting up the de novo query in ProteinProbe*

## Results

Analysis of the raw data with MaxEntLite produced 250 MS/MS spectra.

The resulting workflow search results are shown in **Figure 5**. **Tables 1-3** show the number of peptides identified at each stage of the workflow.

In the first databank search stage 1 peptide from Trypsin was identified along with 20 peptides from pyruvate kinase (**Table 1**).

The second stage Automod search revealed a further 12 new peptides most of which were the result of non-specific cleavage by Trypsin (**Figure 6**). An additional peptide previously identified in the databank search stage was found with a de-amidation modification (**Table 2**).

# Waters

The third *de novo* peptide sequencing stage produced a set of 4 peptides not found in the databank search or Automod stages. Analysis of the sequences with BLAST found significant homology to fructose bi-phosphate (**Table 3**).

Manual inspection of the unassigned spectra at the end of the workflow revealed that the remaining data was inconclusive.

*Figure 5. Result of workflow in ProteinProbe*

| m/z | charge | mw | Sequence | Protein |
|---|---|---|---|---|
| 420.7678 | 2 | 839.5228 | APIIAVTR | KPY1_RABIT Pyruvate Kinase |
| 434.7427 | 2 | 867.4749 | MQHLIAR | KPY1_RABIT Pyruvate Kinase |
| 439.2411 | 2 | 876.4705 | LFEELAR | KPY1_RABIT Pyruvate Kinase |
| 473.2585 | 2 | 944.5114 | VNLAMNVGK | KPY1_RABIT Pyruvate Kinase |
| 495.7581 | 2 | 989.5029 | GSGTAEVELK | KPY1_RABIT Pyruvate Kinase |
| 503.2892 | 2 | 1004.565 | KLFEELAR | KPY1_RABIT Pyruvate Kinase |
| 510.263 | 2 | 1018.508 | GDYPLEAVR | KPY1_RABIT Pyruvate Kinase |
| 559.807 | 2 | 1117.598 | GSGTAEVELKK | KPY1_RABIT Pyruvate Kinase |
| 571.3073 | 2 | 1140.603 | GDLGIEIPAEK | KPY1_RABIT Pyruvate Kinase |
| 577.2921 | 2 | 1152.566 | SSGTSYPDVLK | TRY1_BOVIN Trypsinogen |
| 586.3216 | 2 | 1170.624 | LDIDSAPITAR | KPY1_RABIT Pyruvate Kinase |
| 588.9966 | 3 | 1763.978 | KGVNLPGAAVDLPAVSEK | KPY1_RABIT Pyruvate Kinase |
| 599.2922 | 2 | 1196.575 | ITLDNAYMEK | KPY1_RABIT Pyruvate Kinase |
| 607.9769 | 3 | 1820.909 | RFDEILEASDGIMVAR | KPY1_RABIT Pyruvate Kinase |
| 634.6208 | 3 | 1900.853 | MNFSHGTHEYHAETIK | KPY1_RABIT Pyruvate Kinase |
| 708.0269 | 3 | 2121.074 | QKGPDFLVTEVENGGFLGSK | KPY1_RABIT Pyruvate Kinase |
| 724.9045 | 2 | 1447.792 | VYVDDGLISLQVK | KPY1_RABIT Pyruvate Kinase |
| 818.9482 | 2 | 1635.883 | GVNLPGAAVDLPAVSEK | KPY1_RABIT Pyruvate Kinase |
| 818.9496 | 2 | 1635.883 | GVNLPGAAVDLPAVSEK | KPY1_RABIT Pyruvate Kinase |
| 821.8917 | 2 | 1641.763 | DPVQEAWAEDVDLR | KPY1_RABIT Pyruvate Kinase |
| 833.4104 | 2 | 1664.808 | FDEILEASDGIMVAR | KPY1_RABIT Pyruvate Kinase |
| 933.4625 | 2 | 1864.921 | GPDFLVTEVENGGFLGSK | KPY1_RABIT Pyruvate Kinase |

*Table 1. Results of databank search stage*

| m/z | Charge | mw | Sequence | Protein | Sub / Mod |
|---|---|---|---|---|---|
| 420.711 | 2 | 839.402 | GSVEASYK | KPY1_RABIT Pyruvate Kinase | |
| 424.72 | 2 | 847.422 | SDGIMVAR | KPY1_RABIT Pyruvate Kinase | |
| 437.754 | 2 | 873.492 | IVLTESGR | KPY1_RABIT Pyruvate Kinase | |
| 457.272 | 2 | 912.528 | PVAVALDTK | KPY1_RABIT Pyruvate Kinase | |
| 482.736 | 2 | 963.456 | TGWRPGSGF | KPY1_RABIT Pyruvate Kinase | |
| 521.751 | 2 | 1041.48 | AMGSVEASYK | KPY1_RABIT Pyruvate Kinase | |
| 529.812 | 2 | 1057.613 | ALIVLTESGR | KPY1_RABIT Pyruvate Kinase | |
| 534.25 | 2 | 1066.487 | PGSGFTNTMR | KPY1_RABIT Pyruvate Kinase | |
| 542.753 | 2 | 1083.491 | TLDNAYMEK | KPY1_RABIT Pyruvate Kinase | |
| 785.894 | 2 | 1569.767 | TATESFASDPILYR | KPY1_RABIT Pyruvate Kinase | |
| 799.412 | 2 | 1596.803 | FLVTEVEDGGFLGSK | KPY1_RABIT Pyruvate Kinase | D for N (8) |
| 933.962 | 2 | 1865.905 | GPDFLVTEVEDGGFLGSK | KPY1_RABIT Pyruvate Kinase | D for N (11) |
| 1053.035 | 2 | 2104.048 | QKGPDFLVTEVENGGFLGSK | KPY1_RABIT Pyruvate Kinase | Pyrrolidone carboxylic acid N-TERM (N-term) |

*Table 2. Results of Automod search*

| m/z | charge | mw | Sequence | Sequence from BLAST | Protein |
|---|---|---|---|---|---|
| 470.745 | 2 | 939.477 | ELSDLAHR | ELSDIAHR | ALFA_RABIT Fructose bi-phosphate |
| 666.853 | 2 | 1331.682 | LGLAADESTTELL | LAADEST | ALFA_RABIT Fructose bi-phosphate |
| 745.853 | 2 | 1489.69 | LQSLGTEEDEEKL | LQSIGTENTEE | ALFA_RABIT Fructose bi-phosphate |
| 469.234 | 2 | 936.43 | QEEYVRN | QEEYV | Sequence too short for determination |

*Table 3. Results of de novo search followed by BLAST*

```
SKSHSEAGSA  FIQTQQLHAA  MADTFLEHMC  RLDIDSAPIT  ARNTGIICTI
GPASRSVETL  KEMIKSGMNV  ARMNFSHGTH  EYHAETIKNV  RTATESFASD
PILYRPVAVA  LDTKGPEIRT  GLIKGSGTAE  VELKKGATLK  ITLDNAYMEK
CDENILWLDY  KNICKVVDVG  SKVYVDDGLI  SLQVKQKGPD  FLVTEVENGG
FLGSKKGVNL  PGAAVDLPAV  SEKDIQDLKF  GVEQDVDMVF  ASFIRKAADV
HEVRKILGEK  GKNIKIISKI  ENHEGVRRFD  EILEASDGIM  VARGDLGIEI
PAEKVFLAQK  MIIGRCNRAG  KPVICATQML  ESMIKKPRPT  RAEGSDVANA
VLDGADCIML  SGETAKGDYP  LEAVRMQHLI  AREAEAAMFH  RKLFEELARS
SSHSTDLMEA  MAMGSVEASY  KCLAAALIVL  TESGRSAHQV  ARYRPRAPII
AVTRNHQTAR  QAHLYRGIFP  VVCKDPVQEA  WAEDVDLRVN  LAMNVGKARG
FFKKGDVVIV  LTGWRPGSGF  TNTMRVVPVP
```

*Figure 6. Non-specific cleavage peptides found in Automod stage*

## Conclusion

We have shown that use of an extended peptide identification system incorporating separate PTM analysis, amino acid substitution and *de novo* sequencing can enhance overall protein identification from tandem mass spectrometry data.

PosterREPRINT

**Author to whom all correspondence should be addressed:**
Phillip Young
Waters Corporation
(Micromass UK Limited)
Floats Road, Wythenshawe
Manchester, M23 9LZ
**Tel:** + 44 (0) 161 946 2400
**Fax:** + 44 (0) 161 946 2480
**e-mail:** phillip.young@micromass.co.uk

WATERS CORPORATION
34 Maple St.
Milford, MA 01757 U.S.A.
T: 508 478 2000
F: 508 872 1990
www.waters.com

Made in the United Kingdom

Waters
RIGHT ON TIME.

MICROMASS®
MS TECHNOLOGIES



Certificate No: 951387