186

Evaluation of different interpretation strategies to discover PTM in MS/MS peptide fragmentation data

Daniel C. Chamrad[1]; Gerhard Körting[1]; Ken Fantom[2]; Andy West[2]; Klaus Schneider[2]; Ulrike Schweiger-Hufnagel[3]; Herbert Thiele[3]; Martin Blüggel[1]

[1]Protagen AG, Dortmund, Germany; [2]GlaxoSmithKline, Harlow, Great Britain; [3]Bruker Daltonik GmbH, Bremen, Germany

Overview

The phenomenon of acquired high quality MS/MS spectra that can not be explained within typical sequence database searches is well known. Although protein identification was successful it is manually very laborious and in most cases even impossible to match these spectra with any suggested protein sequence.

The procedure of second pass searches has been developed to overcome this problem. Here we report from our in house developed tool PTM-Explorer (Protagen AG, Dortmund, Germany; Bruker Daltonik GmbH, Bremen, Germany). Two LC-MS/MS datasets of recombinant kinase proteins (MAPK 12, and MAPK 13) have been used for the evaluation.

Introduction

Nowadays high throughput identification of peptides and proteins from large LC based MS/MS datasets is a standard procedure. However usually only a small portion of spectra can be explained and proteins are identified with limited sequence coverage. Besides the success of identification of a large amount of protein species with low sequence coverage, the challenge remains to gain as much primary sequence information as possible including detection of post-translational modification (PTM). Exploring PTM is a prerequisite for understanding the mechanisms and functions of biological processes.

The manifold reasons why spectra remain unidentified in standard database searches have been sumarized in Figure 1.

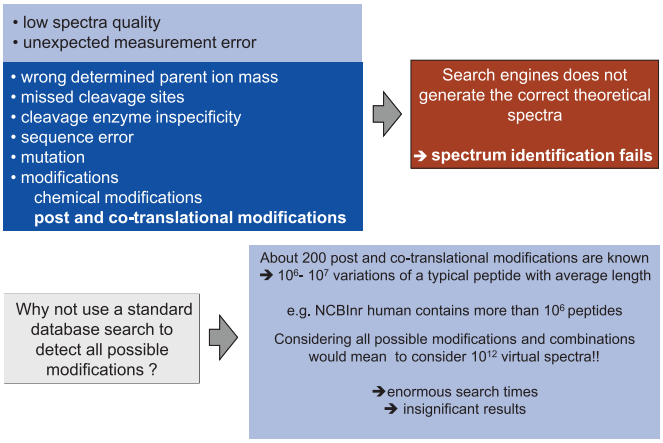

Figure 1. Reasons for unexplained MS/MS spectra after protein identification. One important reason are PTMs. Including all possible PTM (about 200 have been described in the literature) in a single step during a search remains a combinatorial problem.

There lies a significant potential in the numerous currently not interpreted peptide fragmentation spectra to improve primary structure definition of proteins.

After the protein identification, most of this valuable information is lost, simply because of the time consuming process of manual evaluation which is still required using current available software.

To address the needs for more exact primary structure definition and PTM detection, promising software tools have become available performing a second pass search (Figure 2).

Here we use the software tool PTM-Explorer to analyse MS/MS data of kinase proteins and compare the performance with standard database search software (Mascot).

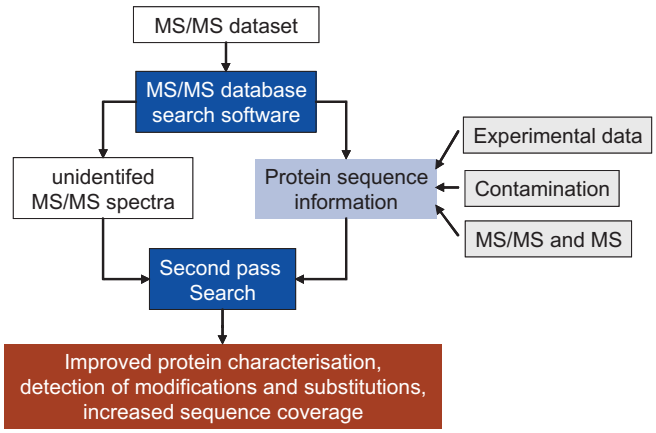

Figure 2. Overview on the second pass search strategy. After a typical MS/MS database search, the identified protein sequences and sequence information available from other experiments is used to perform a second pass search which can detect PTM, amino acid substitution and unspecific cleavage products.

Methods

The analysed LC-MS/MS data stem from tryptic digested MAP-Kinase 12 and MAP-Kinase 13 proteins (GlaxoSmithKline, Harlow, Great Britain). These datasets have been imported to the proteome bioinformatics platform ProteinScape v1.3 (Bruker Daltonik GmbH, Bremen Germany; Protagen AG, Dortmund, Germany). From there standard sequence database searches have been started using Mascot version 2.0 (www.matrixscience.com). The obtained protein identifications (MAPK 12 and 13) were included in a second pass search using the “error tolerant search” of Mascot and the spezialised software PTM-Explorer (Figure 3), which was developed in cooperation with Bruker Daltonik GmbH.

PTM-Explorer is part of ProteinScape 1.3 and uses so called predefined search strategies in an autonomous manner to detect modifications, amino acid substitutions, unsuspected large mass measurement errors, enzyme non-specificity and unknown mass shifts. Possible peptide matches undergo a significance analysis to prevent random matching results. The results are presented in different views. The protein view clusters the results by the amino acids positions in the protein. The peptide view clusters the spectra by all different peptides (Figure 4).

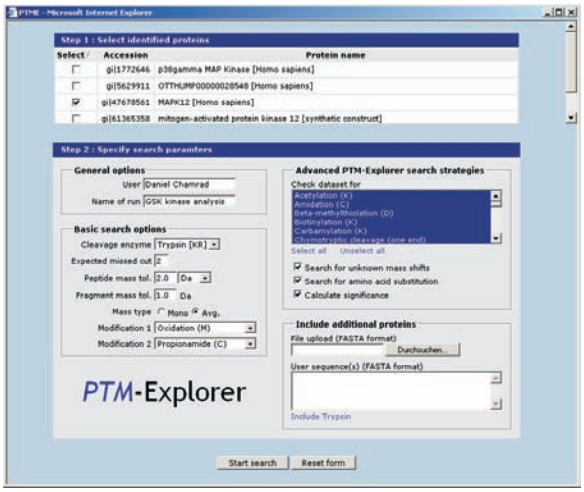

Figure 3. Screenshot of the search parameter input of PTM-Explorer. Using this form the regarded protein sequences and applied search strategies are defined. Proteins that were already identified within ProteinScape are automatically transferred to the form and can be selected. Most search strategies have been addopted from the Unimod (www.unimod.org).

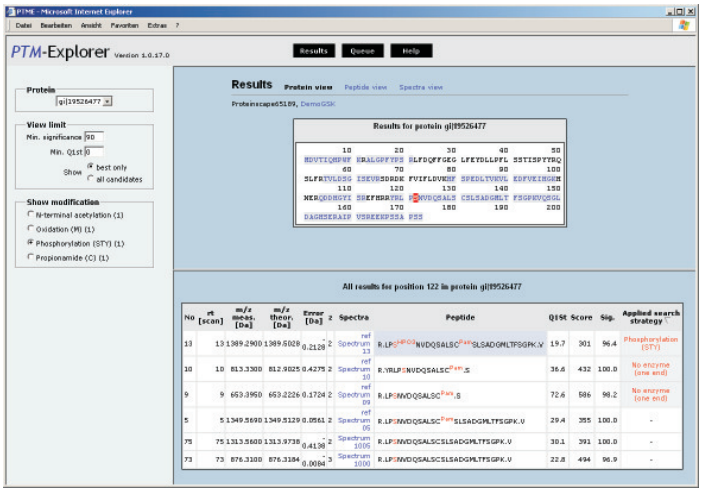

Figure 4. Screenshot of the protein view in PTM-Explorer. A click on an amino acid in the protein sequence displays all spectra which contain information on the specified amino acid. Features as phosphoryation are accessible quickly using a filter (left frame).

Results

Using PTM-Explorer several modifications were detected in the samples of the MAPK 12 and MAPK 13 proteins (Figure 5 and Figure 6). They can be divided in the three groups a) PTM (phosphorylation, methylation), b) experimental (Pyroglutamate formation, methionine oxidation) and c) unknown i.e. spectra that obviously contain kinase specific sequence tag information but the measured parent mass remains unexplained. Besides modified peptides a large amount of peptides were found that have been cleaved unspecifically on one side (20% of all detected peptides).

MAPK 12 - gi|61363685

```
        10         20         30         40         50
MSSPPPARSG FYRQEVTKTA WEVRAVYRDL QPVGSGAYGA VCSAVDGRTG
        60         70         80         90        100
AKVAIKKLYR PFQSELFAKR AYRELRLLKH MRHENVIGLL DVFTPDETLD
       110        120        130        140        150
DFTDFYLVMP FMGTDLGKLM KHEKLGEDRI QFLVYQMLKG LRYIHAAGII
       160        170        180        190        200
HRDLKPGNLA VNEDCELKIL DFGLARQADS EMTGYVVTRW YRAPEVILNW
       210        220        230        240        250
MRYTQTVDIW SVGCIMAEMI TGKTLFKGSD HLDQLKEIMK VTGTPPAEFV
       260        270        280        290        300
QRLQSDEAKN YMKGLPEWEK KDFASILTNA SPLAVNLLEK MLVLDAEQRV
       310        320        330        340        350
TAGEALAHPY FESLHDTEDE PQVQKYDDSF DDVDRTLDEW KRVTYKEVLS
       360        370        380        390        400
FKPPRQLGAR VSKEAPL
```

Pos.	Modification
C-42	Carbamidomethylation
C-165	Carbamidomethylation
Y-185	Phosphorylation
D-272	Beta-methyl thiolation or Gamma carboxylation
M-291	Methionine oxidation
T-365	Phosphorylation

Figure 5. Overview on the MAPK 12 sequence. Modifications detected by PTM-Explorer are marked red and explained in the table.

Two phosphorylations were detected by PTM-Explorer in the MAPK12 dataset, namely Y-185 and T-365. The phosphorylation at T-365 has not been decribed in the literature previously. A mass shift was also found at position D-273 (44 Da), which can be interpreted as a gamma-carboxylation or beta-methyl-thiolation, e.g. heterogenous mass shifts where detected between positions 25 and 49 (Figure 7). PTM-Explorer interprets these as amino acids “substitutions” or as “unknown mass shift”, but a final conclusion cannot be elucidated from the data.

MAPK 13 - gi|30583281

```
        10         20         30         40         50
MSLIRKKGFY KCDVNKTAWE LPKTYVSPTH VGSGAYGSVC SAIDKRSGEK
        60         70         80         90        100
VAIKKLSRPF QSEIFAKRAY RELLLLKHMQ HENVIGLLDV FTPASSLRNF
       110        120        130        140        150
YDFYLVMPFM QTDLQKIMGM EFSEEKIQYL VYQMLKGLKY IHSAGVVHRD
       160        170        180        190        200
LKPGNLAVNE DCELKILDFG LARHADAEMT GYVVTRWYRA PEVILSWMHY
       210        220        230        240        250
NQTVDIWSVG CIMAEMLTGK TLFKGKDYLD QLTQILKVTG VPGTEFVQKL
       260        270        280        290        300
NDKAAKSYIQ SLPQTPRKDF TQLFPRASPQ AADLLEKMLE LDVDKRLTAA
       310        320        330        340        350
QALTHPFFEP FRDPEEETEA QQPFDDSLEH EKLTVDEWKQ HIYKEIVNFS
       360        370        380        390        400
PIARKDSRRR SGMKL
```

Pos.	Modification
Q-12	Pyro-glutamate
Y-25	Phosphorylation
C-40	Carbamidomethylation
C-162	Carbamidomethylation
Y-182	Phosphorylation
T-265	Phosphorylation
Q-340	Pyro-glutamate
K-344	Methylation
E-345	Pryo-glutamate

Figure 6. Overview on the MAPK 13 sequence. Modifications detected by PTM-Explorer are marked red and explained in the table. Using PTM-Explorer the sequence coverage was increased in comparison to the standard Mascot search (49%->62%).

Three different phosphorylations, Pyro-Glutamate formation from N-terminal peptide residues, and Methylation were deteted in the MAPK13 dataset (Figure 8). The phosphorylations at Y-25 and T-265 have not been described in the literature previously. These two phosphorylations, and the methylation at K-344 should be confirmed by additional experiments.

Using Mascot error tolerant search none of the phosphorylations that were detected by PTM-Explorer were found. In a standard search including phosphorylation Mascot found at least one tyrosin phosphorylation in case of both proteins (MAPK12: Y-185, MAPK 13: Y-182).

In the Mascot error tolerant search various mass shifts below 2 Da were detected (mostly substitution). PTM-Explorer identifies these spectra also, but the mass shift is not reported as it is below the mass accuracy of the measurement.

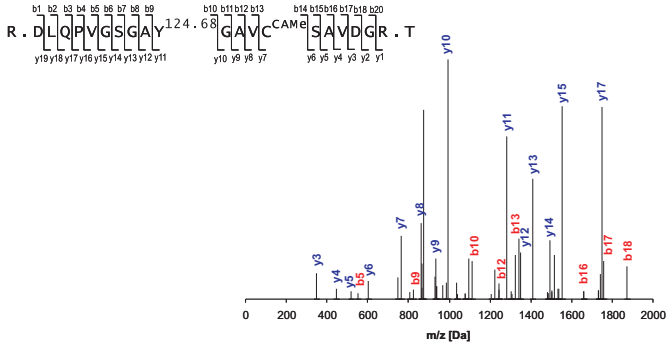

Figure 7. The spectrum shows an observed unknown mass shift detected in the analysed MAPK 12 at position 38. Possibly this can be explained by transpeptidation [2] or a sequence error. Other possible reasons are combination of different modifications in the peptides or sequence errors.

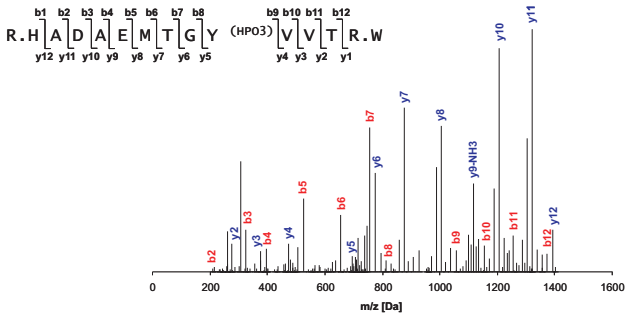

Figure 8. MS/MS spectrum showing the phosphorylation at Y-182 in MAPK 13.

Conclusion

PTM-Explorer successfully allows elucidation of unexplained MS/MS spectra in a high throughput manner. Screening for various chemical and posttranslational modifications, substitutions, sequence errors, unspecific cleavage products and unknown mass shifts can be done in one step now.

In case of the MAPK 12 and MAPK 13 datasets various modifications were detected by PTM-Explorer including 5 different phosphorylation sites. Using a Mascot phosphorylation search reveals only two of the phosphorylation sites.

PTM-Explorer as well as Mascot “error tolerant search” report various spectra that contain kinase specific sequence tags, but the measured parent mass can not be fully explained due to incomplete fragmentation or low spectrum quality.

Time efficient evaluation of MS/MS spectra results is critical. The unique visualisation concept of PTM-Explorer, which uses several views to cluster the spectra, prooved to be suited for large datasets (LC-MS/MS with thousands of spectra).

References

1. Schaefer H, Chamrad DC, Marcus K, Reidegeld KA, Bluggel M, Meyer HE. Proteomics. 2005 Mar;5(4):846-52.