

On Finding putative PTM (pPTM) Marker Ion in HCD scans using PTM_MarkerFinder

Christian Panse*

Functional Genomics Center Zurich

Paolo Nanni[†]

Functional Genomics Center Zurich

Abstract

Glycopeptides as well as acetylated, methylated and other modified peptides release specific fragment ions during CID (collision-induced dissociation) and HCD (higher energy collisional dissociation) fragmentation. These fragment ions can be used to validate the presence of the PTM (post translational modifications) on the peptides. `PTM_MarkerFinder`, an R function of the `protViz` package that takes advantage of such marker ions. `PTM_MarkerFinder` scans the MS/MS spectra in the output of a peptide spectrum match search, e.g., Mascot, for marker ions specific for selected PTMs.

While the software tool has been described by [Nanni, Panse, Gehrig, Mueller, Grossmann, and Schlapbach \(2013\)](#) here we provide a step-by-step guide on how the software can be used.

Keywords: MarkerFinder, putative post translational modifications, R.

1. Howto get the software and data

The method for finding the marker ions is contained in the R package `protViz` available through CRAN using <https://cran.r-project.org/package=protViz>. The package requires R ([R Development Core Team 2008](#)) installed.

The minimal data structure requirement for the `PTM_MarkerFinder` function looks as follow.

```
R> library(protViz)
R> data(HexNAc)
R> str(HexNAc[[1]], nchar.max = 30)
```

List of 12

```
$ peptideSequence      : chr "STMQELNSR"
$ mascotScore          : num 49.5
$ modification         : chr "000000000000"
$ MonoisotopicAAMass  : num [1:9] 0 0 0 0 0 0 0 0 0
$ proteinInformation  : chr "zz|ZZ_FGCZCont0219|"
$ title                : chr "NGlycoFASP_NH"| __truncated__
```

*Correspondence: Christian Panse, Functional Genomics Center Zurich, Winterthurerstr. 190, CH-8057, Zürich, Switzerland, Telephone: +41-44-63-53910, E-mail: cp@fgcz.ethz.ch

[†]Paolo Nanni, Functional Genomics Center Zurich, Winterthurerstr. 190, CH-8057, Zürich, Switzerland, Telephone: +41-44-63-53930, E-mail: paolo.nanni@fgcz.uzh.ch

```

$ pepmass      : num 533
$ charge       : num 2
$ scans        : num 2659
$ rtinseconds  : num 1846
$ mZ           : num [1:150] 101 104 105 110 112 ...
$ intensity    : num [1:150] 369.3 2860 37.3 103.8 190.7 ...

```

Here we have listed the HexNAc data which is included in **protViz**.

protViz also provides a perl script `protViz_mascotDat2RData.pl`¹ taking mascot server dat files as input and producing RData output.

```

$ /usr/local/lib/R/site-library/protViz/exec/protViz_mascotDat2RData.pl \
  -d=/usr/local/mascot/data/20130116/F178287.dat \
  -m=$HOME/mod_file

```

`mascotDat2RData.pl` requires the mascot server `mod_file` keeping all the configured modification of the mascot server.

In theory `PTM_MarkerFinder` can process the output of any search engine for peptide identification. It is up to the R user writing a wrapper script converting the output of any particular peptide identification search engine to the data structure listed above.

2. Finding the Marker Ions

2.1. HexNAc – Example

`PTM_MarkerFinder` can search for any Marker ion series. The next lines define the `HexNAc_MarkerIons`.

```

R> HexNAc_MarkerIons <- c(126.05495, 138.05495, 144.06552,
+   168.06552, 186.07608, 204.08665)

```

The lines below configure the modification information used by the search engine. The HexNAc modification below is described on unimod http://www.unimod.org/modifications_view.php?editid1=43.

```

R> ptm.0 <- cbind(AA = "-",
+   mono = 0.0, avg = 0.0, desc = "unmodified", unimodAccID = NA)
R> ptm.1 <- cbind(AA='N',
+   mono = 317.122300, avg = NA, desc = "HexNAc",
+   unimodAccID=2)
R> ptm.2 <- cbind(AA='M',
+   mono = 147.035400, avg = NA, desc = "Oxidation",
+   unimodAccID=1)
R> m <- as.data.frame(rbind(ptm.0, ptm.1, ptm.2))

```

¹The prefix `protViz_` is used to benefit from the `bash` tab completion.

PTM_MarkerFinder is called.

```
R> S <- PTM_MarkerFinder(data = HexNAc,
+   modification = m$mono,
+   modificationName = m$desc,
+   minMarkerIntensityRatio = 3,
+   itol_ppm = 20,
+   mZmarkerIons = HexNAc_MarkerIons)
```

The content of S can be seen in the Table below.

scans	mZ	markerIonMZ	markerIonIntensity	markerIonMzError	markerIonPpmError	query	pepmass	peptideSequence	modification
3687	126.06	126.05	9945.00	-0.00	-0.64257649497898	4	713.36	IMNVTTDSLTK	000100000000
3687	138.06	138.05	1933.00	-0.00	-2.49175522390729	4	713.36	IMNVTTDSLTK	000100000000
3687	144.07	144.07	412.30	-0.00	-1.59649326794302	4	713.36	IMNVTTDSLTK	000100000000
3687	168.07	168.07	810.20	-0.00	-2.36811844277867	4	713.36	IMNVTTDSLTK	000100000000
3687	204.09	204.09	3273.00	-0.00	-1.74435407225623	4	713.36	IMNVTTDSLTK	000100000000
2540	126.06	126.05	2945.00	-0.00	-0.825036336847078	6	490.56	HSFNGNQSTFK	0000001000000
2540	138.06	138.05	759.20	-0.00	-10.3725737215287	6	490.56	HSFNGNQSTFK	0000001000000
2540	144.07	144.07	195.40	-0.00	-0.118001850879316	6	490.56	HSFNGNQSTFK	0000001000000
2540	168.07	168.07	262.90	-0.00	-0.916308466469431	6	490.56	HSFNGNQSTFK	0000001000000
2540	186.08	186.08	188.50	-0.00	-2.95577150125756	6	490.56	HSFNGNQSTFK	0000001000000
2540	204.09	204.09	998.40	-0.00	-1.5189603491234	6	490.56	HSFNGNQSTFK	0000001000000
4393	126.06	126.05	13620.00	-0.00	-1.03922824020165	9	891.41	EASGLSDNETEWLK	000000010000000
4393	138.06	138.05	3798.00	-0.00	-0.420122390602973	9	891.41	EASGLSDNETEWLK	000000010000000
4393	168.07	168.07	1526.00	-0.00	-0.642606113437682	9	891.41	EASGLSDNETEWLK	000000010000000
4393	186.08	186.08	1014.00	-0.00	-0.983467730223809	9	891.41	EASGLSDNETEWLK	000000010000000
4393	204.09	204.09	5041.00	-0.00	-1.06817259804309	9	891.41	EASGLSDNETEWLK	000000010000000
2739	126.06	126.05	7327.00	-0.00	-0.690174721011021	10	665.59	NA	NA
2739	138.05	138.05	1963.00	-0.00	-0.311470082107949	10	665.59	NA	NA
2739	144.07	144.07	468.60	-0.00	-0.5344787486255	10	665.59	NA	NA
2739	168.07	168.07	624.30	-0.00	-0.642606113437682	10	665.59	NA	NA
2739	204.09	204.09	2496.00	-0.00	-0.622284313992652	10	665.59	NA	NA

Table 1: Result

```
R> summary(S)
```

```
scans      mZ      markerIonMZ      markerIonIntensity
2540:6   Min.    :126.1   Min.    :126.1   Min.    : 188.5
2739:5   1st Qu.:138.1   1st Qu.:138.1   1st Qu.: 624.3
3687:5   Median :144.1   Median :144.1   Median : 1526.0
4393:5   Mean    :159.5   Mean    :159.5   Mean    : 2838.1
          3rd Qu.:186.1   3rd Qu.:186.1   3rd Qu.: 3273.0
          Max.    :204.1   Max.    :204.1   Max.    :13620.0
```

```
markerIonMzError      markerIonPpmError      query
Min.    : -0.0014320   -0.642606113437682: 2      10:5
1st Qu. : -0.0003100   -0.118001850879316: 1      4 :5
Median  : -0.0001310   -0.311470082107949: 1      6 :6
Mean    : -0.0002436   -0.420122390602973: 1      9 :5
3rd Qu. : -0.0000870   -0.5344787486255 : 1
Max.    : -0.0000170   -0.622284313992652: 1
          (Other)      :14
```

```
pepmass      peptideSequence      modification
Min.    :490.6   EASGLSDNETEWLK:5      0000000010000000:5
1st Qu. :490.6   HSFNGNQSTFK :6      0000001000000 :6
Median  :665.6   IMNVTTDSLTK :5      0001000000000 :5
```

```

Mean      :680.7   NA           :5       NA           :5
3rd Qu.  :713.4
Max.     :891.4

```

Some overview graphics just an overview of the sample data set HexNAc.

```

R> op <- par(mfrow = c(2, 2), mar=c(4, 4, 4, 1))
R> dump <- lapply(split(S, S$query),
+   function(x){
+     plot(x$mZ, x$markerIonIntensity,
+       type = 'h',
+       col = 'lightblue',
+       cex = 2,
+       ylab = 'intensity', xlab='m/z',
+       xlim = range(c(HexNAc_MarkerIons,
+         max(HexNAc_MarkerIons)
+         + 0.1 * (max(HexNAc_MarkerIons) - min(HexNAc_MarkerIons)),
+         min(HexNAc_MarkerIons)
+         - 0.1 * (max(HexNAc_MarkerIons) - min(HexNAc_MarkerIons))))),
+       ylim = range(S$markerIonIntensity),
+       log = 'y',
+       main = paste("scan=", unique(x$scans),
+         "/query=", unique(x$query), sep=''));
+     text(x$mZ, x$markerIonIntensity,
+       round(x$mZ,2), col='red', cex=0.7)
+   })
R> par(op)

```

Figure 1 displays the output of PTM_MarkerFinder.

2.2. Reshaping the output and export

The R method `reshape` transforms the data frame `S` from a long format to a wide format.

```

R> names(S)[4] <- "mII"
R> S.wide <- reshape(S[,c(1,7,3,4)],
+   direction = 'wide',
+   timevar = "markerIonMZ",
+   idvar = c('scans', 'query'))
R>

```

export as comma separated file

```

R> write.table(S.wide,
+   file = file.path(tempdir(), "HexNAc_PTm_markerFinder.csv"),

```

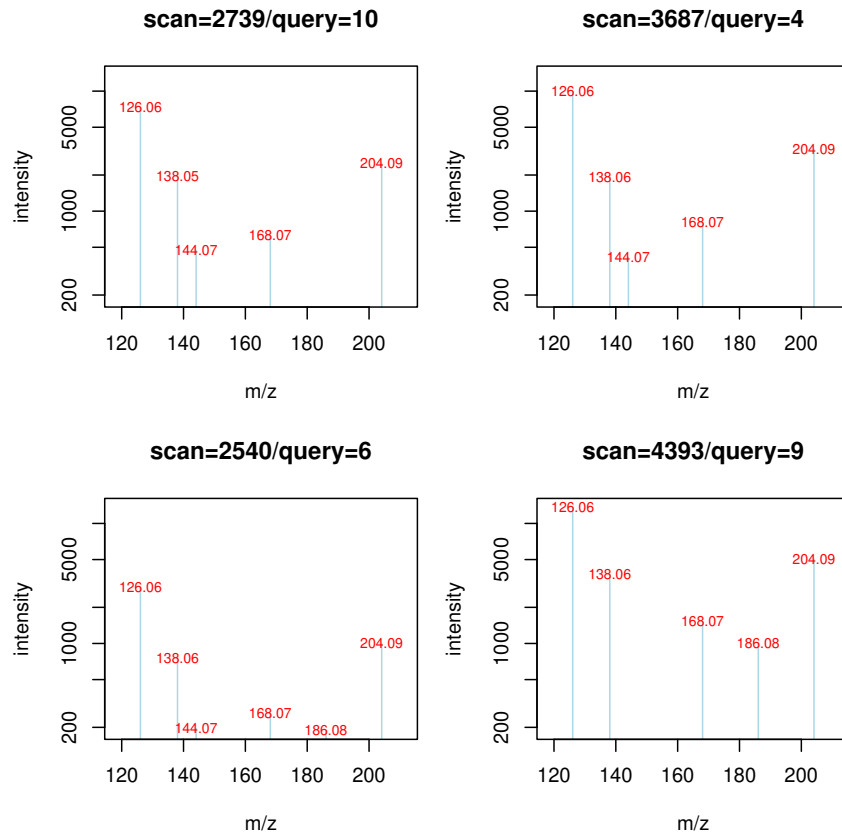


Figure 1: Overview of the marker ions.

scans	query	mII.126.05495	mII.138.05495	mII.144.06552	mII.168.06552	mII.204.08665	mII.186.07608
3687	4	9945.00	1933.00	412.30	810.20	3273.00	
2540	6	2945.00	759.20	195.40	262.90	998.40	188.50
4393	9	13620.00	3798.00		1526.00	5041.00	1014.00
2739	10	7327.00	1963.00	468.60	624.30	2496.00	

Table 2: Result

```

+         sep = ', ',
+         row.names = FALSE,
+         col.names = TRUE,
+         quote = FALSE)

```

2.3. Visualization of the Result

```

R> # prepare the input
R> d <- list(); d[[1]] <- HexNAc[[3]]; d[[2]] <- HexNAc[[4]]; d[[3]] <- HexNAc[[5]]
R> S <- PTM_MarkerFinder(data = d, modification = m$mono,
+         modificationName = m$desc,
+         minMarkerIntensityRatio = 3,

```

```
+      itol_ppm = 20,  
+      mZmarkerIons = HexNAc_MarkerIons)
```

The graphics can be seen in Figure 2.

3. Demonstartion

The user can call the demonstration with

```
R> demo(PTM_MarkerFinder)
```

3.1. Other examples

The following ADP-Ribose marker ions configuration was described by [Bilan, Leutert, Nanni, Panse, and Hottiger \(2017\)](#).

```
R> ADP_Ribose <- c(136.0618, 250.0935, 348.0704, 428.0367)
```

4. Session information

An overview of the package versions used to produce this document are shown below.

- R version 4.1.3 (2022-03-10), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=en_US.UTF-8, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Running under: Debian GNU/Linux 11 (bullseye)
- Matrix products: default
- BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
- LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p-r0.3.13.so
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: lattice 0.20-45, protViz 0.7.3, xtable 1.8-4
- Loaded via a namespace (and not attached): codetools 0.2-18, compiler 4.1.3, grid 4.1.3, Rcpp 1.0.8, tools 4.1.3

References

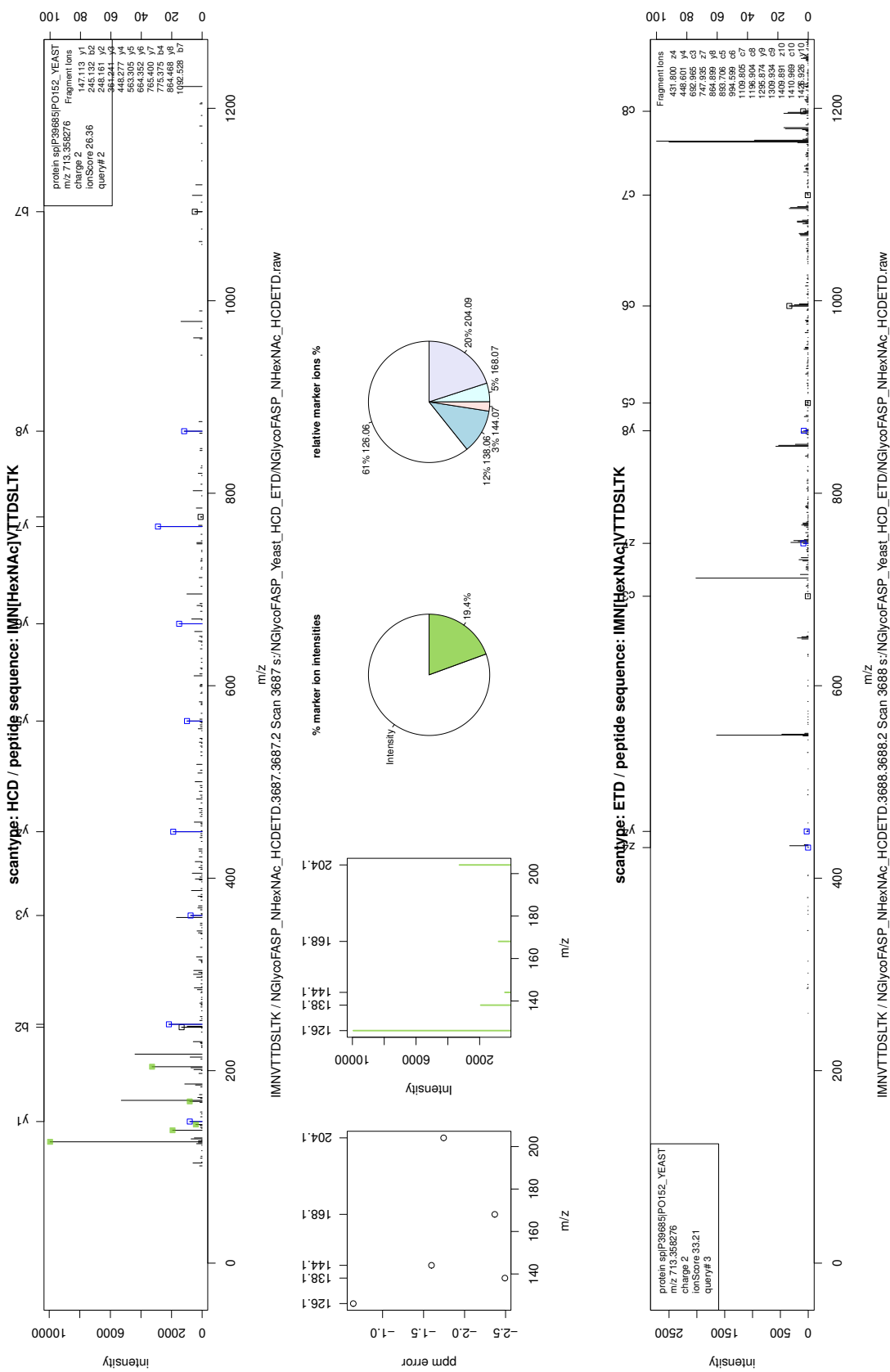


Figure 2: Graphical output of the method.

- Bilan V, Leutert M, Nanni P, Panse C, Hottiger MO (2017). “Combining Higher-Energy Collision Dissociation and Electron-Transfer/Higher-Energy Collision Dissociation Fragmentation in a Product-Dependent Manner Confidently Assigns Proteomewide ADP-Ribose Acceptor Sites.” *Anal. Chem.*, **89**(3), 1523–1530. doi:10.1021/acs.analchem.6b03365.
- Nanni P, Panse C, Gehrig P, Mueller S, Grossmann J, Schlapbach R (2013). “PTM MarkerFinder, a software tool to detect and validate spectra from peptides carrying post-translational modifications.” *Proteomics*, **13**(15), 2251–2255. doi:10.1002/pmic.201300036.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Affiliation:

Paolo Nanni and Christian Panse
UZH|ETH Zürich
Functional Genomics Center Zurich
Winterthurerstr. 190
CH-8057, Zürich, Switzerland
Telephone: +41/44/63-53910