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. How to 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 "00000000000"
 $ MonoisotopicAAmass: num [1:9] 0 0 0 0 0 0 0 0 0
 $ proteinInformation: chr "zz|ZZ_FGCZCont0219|"__truncated__
 $ 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
Here we have listed the HexNAc data which is included in protViz. protViz also provides and perl script protViz_mascotDat2RData.pl\(^1\) taking mascot server dat files as input and producing RData output.

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

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.

\[
\text{R> HexNAc_MarkerIons <- c(126.05495, 138.05495, 144.06552,} \\
\text{168.06552, 186.07608, 204.08665)}
\]

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](http://www.unimod.org/modifications_view.php?editid1=43).

\[
\text{R> ptm.0 <- cbind(AA = "-",} \\
\text{ mono = 0.0, avg = 0.0, desc = "unmodified", unimodAccID = NA)}
\]

\[
\text{R> ptm.1 <- cbind(AA='N',} \\
\text{ mono = 317.122300, avg = NA, desc = "HexNAc",} \\
\text{ unimodAccID=2)}
\]

\[
\text{R> ptm.2 <- cbind(AA='M',} \\
\text{ mono = 147.035400, avg = NA, desc = "Oxidation",} \\
\text{ unimodAccID=1)}
\]

\[
\text{R> m <- as.data.frame(rbind(ptm.0, ptm.1, ptm.2))}
\]

\(^1\)The prefix protViz_ is used to benefit from the bash tab completion.
PTM_MarkerFinder is called.

```r
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.

<table>
<thead>
<tr>
<th>scan</th>
<th>mZ</th>
<th>markerIonMZ</th>
<th>markerIonIntensity</th>
<th>markerIonMzError</th>
<th>markerIonPpmError</th>
<th>query</th>
<th>pepmass</th>
<th>peptideSequence</th>
<th>modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>3687</td>
<td>126.06</td>
<td>126.05</td>
<td>9945.00</td>
<td>-0.00</td>
<td>-0.64260611347682</td>
<td>4</td>
<td>713.36</td>
<td>IMNVTTDSL</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>3687</td>
<td>138.06</td>
<td>138.05</td>
<td>1033.00</td>
<td>-0.00</td>
<td>-2.4917522896729</td>
<td>4</td>
<td>713.36</td>
<td>IMNVTTDSL</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>3687</td>
<td>144.07</td>
<td>144.07</td>
<td>412.30</td>
<td>-0.00</td>
<td>-1.56491287493402</td>
<td>4</td>
<td>713.36</td>
<td>IMNVTTDSL</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>3687</td>
<td>159.07</td>
<td>159.07</td>
<td>1010.20</td>
<td>-0.00</td>
<td>-2.3661142477667</td>
<td>4</td>
<td>713.36</td>
<td>IMNVTTDSL</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>3687</td>
<td>204.09</td>
<td>204.09</td>
<td>3273.00</td>
<td>-0.00</td>
<td>-1.7435670225623</td>
<td>4</td>
<td>713.36</td>
<td>IMNVTTDSL</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2540</td>
<td>126.06</td>
<td>126.05</td>
<td>2045.00</td>
<td>-0.00</td>
<td>-0.825036336847078</td>
<td>6</td>
<td>490.56</td>
<td>HSFNGNQSTFK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2540</td>
<td>138.06</td>
<td>138.05</td>
<td>759.30</td>
<td>-0.00</td>
<td>-10.3725737215287</td>
<td>6</td>
<td>490.56</td>
<td>HSFNGNQSTFK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2540</td>
<td>144.07</td>
<td>144.07</td>
<td>193.30</td>
<td>-0.00</td>
<td>-1.1800150979156</td>
<td>6</td>
<td>490.56</td>
<td>HSFNGNQSTFK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2540</td>
<td>168.07</td>
<td>168.07</td>
<td>262.30</td>
<td>-0.00</td>
<td>-0.2639486409031</td>
<td>6</td>
<td>490.56</td>
<td>HSFNGNQSTFK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2540</td>
<td>186.08</td>
<td>186.08</td>
<td>180.00</td>
<td>-0.00</td>
<td>-2.95373516125756</td>
<td>6</td>
<td>490.56</td>
<td>HSFNGNQSTFK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>4393</td>
<td>126.06</td>
<td>126.05</td>
<td>13620.00</td>
<td>-0.00</td>
<td>-1.03922824020165</td>
<td>9</td>
<td>891.41</td>
<td>EASGLSDNETEWLK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>4393</td>
<td>138.06</td>
<td>138.05</td>
<td>3798.00</td>
<td>-0.00</td>
<td>-0.420122300082973</td>
<td>9</td>
<td>891.41</td>
<td>EASGLSDNETEWLK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>4393</td>
<td>168.07</td>
<td>168.07</td>
<td>1526.00</td>
<td>-0.00</td>
<td>-0.64260611347682</td>
<td>9</td>
<td>891.41</td>
<td>EASGLSDNETEWLK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>4393</td>
<td>186.08</td>
<td>186.08</td>
<td>1014.00</td>
<td>-0.00</td>
<td>-0.93540773022849</td>
<td>9</td>
<td>891.41</td>
<td>EASGLSDNETEWLK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>4393</td>
<td>204.09</td>
<td>204.09</td>
<td>5041.00</td>
<td>-0.00</td>
<td>-0.8673258603409</td>
<td>9</td>
<td>891.41</td>
<td>EASGLSDNETEWLK</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2739</td>
<td>126.06</td>
<td>126.05</td>
<td>7327.00</td>
<td>-0.00</td>
<td>-0.690174172040102</td>
<td>10</td>
<td>665.59</td>
<td>NA</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2739</td>
<td>138.05</td>
<td>138.05</td>
<td>1063.00</td>
<td>-0.00</td>
<td>-0.311470062107949</td>
<td>10</td>
<td>665.59</td>
<td>NA</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2739</td>
<td>144.07</td>
<td>144.07</td>
<td>456.60</td>
<td>-0.00</td>
<td>-0.5344787485255</td>
<td>10</td>
<td>665.59</td>
<td>NA</td>
<td>HexNAc_MarkerIons</td>
</tr>
<tr>
<td>2739</td>
<td>168.07</td>
<td>168.07</td>
<td>624.30</td>
<td>-0.00</td>
<td>-0.622284313992652</td>
<td>10</td>
<td>665.59</td>
<td>NA</td>
<td>HexNAc_MarkerIons</td>
</tr>
</tbody>
</table>

Table 1: Result

```r
R> summary(S)
```
Some overview graphics  just an overview of the sample data set HexNAc.

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
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
R> write.table(S.wide,
+ file = file.path(tempdir(), "HexNAc_PTM_markerFinder.csv"),
```
Figure 1: Overview of the marker ions.

Table 2: Result

<table>
<thead>
<tr>
<th>scans</th>
<th>query</th>
<th>m/z126.05</th>
<th>m/z138.05</th>
<th>m/z144.06</th>
<th>m/z168.07</th>
<th>m/z204.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>3687</td>
<td>4</td>
<td>9945.00</td>
<td>1933.00</td>
<td>412.30</td>
<td>810.20</td>
<td>3273.00</td>
</tr>
<tr>
<td>2540</td>
<td>6</td>
<td>2945.00</td>
<td>759.20</td>
<td>195.40</td>
<td>262.90</td>
<td>998.40</td>
</tr>
<tr>
<td>4393</td>
<td>9</td>
<td>13620.00</td>
<td>3798.00</td>
<td>1526.00</td>
<td>5041.00</td>
<td>1014.00</td>
</tr>
<tr>
<td>2739</td>
<td>10</td>
<td>7327.00</td>
<td>1963.00</td>
<td>468.60</td>
<td>624.30</td>
<td>2496.00</td>
</tr>
</tbody>
</table>

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

\[ \textbf{R} \geq \text{demo(PTM\_MarkerFinder)} \]

3.1. Other examples

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

\[ \textbf{R} \geq \text{ADP\_Ribose} \leftarrow c(136.0618, 250.0935, 348.0704, 428.0367) \]

4. Session information

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

- R version 4.3.2 (2023-10-31), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=C, LC_NUMERIC=C, LC_TIME=en\_US.UTF-8, LC_COLLATE=C,
  LC_MONETARY=en\_US.UTF-8, LC_MESSAGES=en\_US.UTF-8,
  LC_PAPER=en\_US.UTF-8, LC_NAME=C, LC_ADDRESS=C
- Time zone: Europe/Zurich
- TZcode source: system (glibc)
- Running under: Debian GNU/Linux trixie/sid
- Matrix products: default
- BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.11.0
- LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.11.0
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: protViz 0.7.9, xtable 1.8-4
- Loaded via a namespace (and not attached): Rcpp 1.0.11, codetools 0.2-19,
  compiler 4.3.2, tools 4.3.2
Figure 2: Graphical output of the method.
References


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