In-silico cleavage of polypeptides using the cleaver package

Sebastian Gibb*

April 24, 2017

Contents

1 Introduction 1
2 Simple Usage 1
3 Insulin & Somatostatin Example 3
4 Isotopic Distribution Of Tryptic Digested Insulin 4
5 Session Information 5

1 Introduction

Most proteomics experiments need protein (peptide) separation and cleavage procedures before these molecules could be analyzed or identified by mass spectrometry or other analytical tools. cleaver allows in-silico cleavage of polypeptide sequences to e.g. create theoretical mass spectrometry data. The cleavage rules are taken from the ExPASy PeptideCutter tool\(^4\).

2 Simple Usage

Loading the cleaver package:

```
> library("cleaver")
```

Getting help and list all available cleavage rules:

```
*mail@sebastiangibb.de
In-silico cleavage of polypeptides using the `cleaver` package

```r
> help("cleave")

Cleaving of *Gastric juice peptide 1 (P01358)* using *Trypsin*:

```r
> ## cleave it
> cleave("LAAGKVEDSD", enzym="trypsin")
```

$LAAGKVEDSD
[1] "LAAGK" "VEDSD"

```r
> ## get the cleavage ranges
> cleavageRanges("LAAGKVEDSD", enzym="trypsin")
```

$LAAGKVEDSD

```r
start end
[1,] 1 5
[2,] 6 10
```

```r
> ## get only cleavage sites
> cleavageSites("LAAGKVEDSD", enzym="trypsin")
```

$LAAGKVEDSD

```r
[1] 5
```

Sometimes cleavage is not perfect and the enzyme miss some cleavage positions:

```r
> ## miss one cleavage position
> cleave("LAAGKVEDSD", enzym="trypsin", missedCleavages=1)
```

$LAAGKVEDSD

```r
[1] "LAAGKVEDSD"
```

```r
> cleavageRanges("LAAGKVEDSD", enzym="trypsin", missedCleavages=1)
```

$LAAGKVEDSD

```r
start end
[1,] 1 10
```

```r
> ## miss zero or one cleavage positions
> cleave("LAAGKVEDSD", enzym="trypsin", missedCleavages=0:1)
```

$LAAGKVEDSD

```r
[1] "LAAGK" "VEDSD" "LAAGKVEDSD"
```

```r
> cleavageRanges("LAAGKVEDSD", enzym="trypsin", missedCleavages=0:1)
```

$LAAGKVEDSD

```r
start end
[1,] 1 5
[2,] 6 10
[3,] 1 10
```

Combine `cleaver` and the *Biostrings* R package:

```r
> ## create AAStringSet object
> p <- AAStringSet(c(gaju="LAAGKVEDSD", pnm="AGEPKLDAGV"))
```
In-silico cleavage of polypeptides using the `cleaver` package

```r
> ## cleave it
> cleave(p, enzym="trypsin")

AAStringSetList of length 2
[["gaju"]]
LAAGK VEDSD
[["pnm"]]
AGEPK LDAGV

> cleavageRanges(p, enzym="trypsin")

IRangesList of length 2
$gaju
IRanges object with 2 ranges and 0 metadata columns:

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

$pnm
IRanges object with 2 ranges and 0 metadata columns:

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

> cleavageSites(p, enzym="trypsin")

$gaju
[1] 5

$pnm
[1] 5

3 Insulin & Somatostatin Example

Downloading Insulin (P01308) and Somatostatin (P61278) sequences from the UniProt database using the `UniProt.ws` R package.

```r
> ## load UniProt.ws library
> library("UniProt.ws")

> ## select species Homo sapiens
> UniProt.ws <- UniProt.ws(taxId=9606)

> ## download sequences of Insulin/Somatostatin
> s <- select(UniProt.ws, keys=c("P01308", "P61278"), columns=c("SEQUENCE"))

Getting extra data for P01308,P61278

'select()' returned 1:1 mapping between keys and columns
In-silico cleavage of polypeptides using the `cleaver` package

```r
## fetch only sequences
sequences <- setNames(s$SEQUENCE, s$UNIPROTKB)

## remove whitespaces
sequences <- gsub(pattern='[[:space:]]', replacement='', x=sequences)

Cleaving using **Pepsin**:

```r
cleave(sequences, enzym="pepsin")
```

$P01308

```
[1] "MA" "L" "WMRLLP" "LL"
[5] "A" "WPDPAAA" "F" "VNQH"
[9] "CGSH" "VEA" "Y" "VCGERG"
[13] "FF" "YTPKTRREAED" "QVQVQVE" "GGGPGAGS"
[17] "LQP" "LA" "EGS" "QKRGIVEQCCTSICS"
[21] "YQ" "ENYCN"
```

$P61278

```
[1] "ML" "SCRL" "QCA"
[4] "L" "AA" "SIV"
[7] "A" "GCVTGAPSHPRL" "RQ"
[10] "FL" "QKS" "LAAAAGQKEL"
[13] "AKY" "AE" "SEPNQTENDA"
[16] "LEPED" "SQAAEQDEMRL" "EL"
[19] "QRSANSNPAMAPRERKAGCKN" "FF" "WKT"
[22] "FTSC"
```

### 4 Isotopic Distribution Of Tryptic Digested Insulin

A common use case of in-silico cleavage is the calculation of the isotopic distribution of peptides (which were enzymatic digested in the in-vitro experimental workflow). Here the `BRAIN` R package\(^2\) is used to calculate the isotopic distribution of `cleaver`'s output. (please note: it is only a toy example, e.g. the relation of intensity values between peptides isn't correct).

```r
## load BRAIN library
library("BRAIN")

## cleave insulin
cleavedInsulin <- cleave(sequences[1], enzym="trypsin")[[1]]

## create empty plot area
plot(NA, xlim=c(150, 4300), ylim=c(0, 1),
+     xlab="mass", ylab="relative intensity",
+     main="tryptic digested insulin - isotopic distribution")

## loop through peptides
for (i in seq(along=cleavedInsulin)) {
```

### count C, H, N, O, S atoms in current peptide
```r
apoms <- BRAIN::getAtomsFromSeq(cleavedInsulin[[i]])
```

### calculate isotopic distribution
```r
d <- useBRAIN(atoms)
```

### draw peaks
```r
lines(d$masses, d$isoDistr, type="h", col=2)
```

---

5 Session Information

- R version 3.4.0 (2017-04-21), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=C,
In-silico cleavage of polypeptides using the cleaver package

- Running under: Ubuntu 16.04.2 LTS
- Matrix products: default
- BLAS: /home/biocbuild/bbs-3.5-bioc/R/lib/libRblas.so
- LAPACK: /home/biocbuild/bbs-3.5-bioc/R/lib/libRlapack.so
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils
- Other packages: BRAIN 1.22.0, BiocGenerics 0.22.0, Biostrings 2.44.0, IRanges 2.10.0, PolynomF 0.94, RCurl 1.95-4.8, RSQLite 1.1-2, S4Vectors 0.14.0, UniProt.ws 2.16.0, XVector 0.16.0, bitops 1.0-6, cleaver 1.14.0, knitr 1.15.1, lattice 0.20-35
- Loaded via a namespace (and not attached): AnnotationDbi 1.38.0, Biobase 2.36.0, BiocStyle 2.4.0, DBI 0.6-1, Rcpp 0.12.10, backports 1.0.5, compiler 3.4.0, digest 0.6.12, evaluate 0.10, grid 3.4.0, highr 0.6, htmltools 0.3.5, magrittr 1.5, memoise 1.1.0, rmarkdown 1.4, rprojroot 1.2, stringi 1.1.5, stringr 1.2.0, tools 3.4.0, yaml 2.1.14, zlibbioc 1.22.0

References


