Package 'prodata'

May 24, 2017

Title Builds Protien Sequence And Subregion Molecular Weight Database.

Version 1.1

Description Given a list of protiens this package will build a database of the protien sequence, the sub region sequence, the M and CC PTMS, and the molecular weights.

Depends R (>= 3.3.3),
Peptides (>= 2.0.0),
reutils (>= 0.2.3),
Biostrings (>= 2.42.1),
data.table (>= 1.10.4),
plyr (>= 1.8.4),
biomaRt (>= 2.30.0),
devtools (>= 1.12.0),
stringr (>= 1.2.0),
biofiles (>= 0.0),
UniProt.ws (>= 2.16.0),
ape (>= 4.1),
tibble (>= 1.2)

Remotes github::gschofl/biofiles

License GNU GENERAL PUBLIC LICENSE V3.0

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

R topics documented:

| conversion | *Uniprot PTM names with standarized names and PTM molecular weight effects* |

Description

This data was assembled of the author of this package from various sources. The Old ptm names come from the Uniprot PTM names. The new PTM names are more computer friendly and merge PTMs that are essentially the same

Usage

```
data(conversion)
```

Format

CSV file with standardized PTM names and effect of the molecular weights of the PTM.

Examples

```
data(conversion)
```

| countPTMs | *A function that takes a protein dataframe and ptm data frame.* |

Description

This function will allow you to count the number of PTMs in a protein data frame generated by the fetchGenebank function. It will use the data from the PTM dataframe created by the PTM function. Using the geneid it will link the data and count the PTMs.

Usage

```
countPTMs(proteindf, ptmdf)
```

Arguments

proteindf	Data frame of protein data generated by the fetchGenebank function.
ptmdf	Data frame of PTM data generated by the ptm function

Examples

```
proteindf <- countPTMs(proteindf, ptmdf)
```

---

fetchGenebank	*A function that takes a list of UIDs and returns a genebank files*

Description

This function will allow you to enter a list of protien/gene names and convert the into NCBI unique IDs. It will return a list. The first element will be your UIDss the second element will have the protien/gene names that were not matched to an ID.

Usage

```
fetchGenebank(UIDs)
```

Arguments

UIDss	List of NCBI UIDss.

Examples

```
df <- fetchGenebankl(UIDss)
```

---

getUID	*A function that takes a protien/genelist and returns the Unique ID number of NCBI*

Description

This function will allow you to enter a list of protien/gene names and convert the into NCBI unique IDs. It will return a list. The first element will be your UIDs the second element will have the protien/gene names that were not matched to an ID.

Usage

```
getUID(protien_gene_list,
  baseterm = "[Protein Name] AND Mus musculus[Organism]", db = "protien")
```

Arguments

protien_gene_list

> List of protiens or genes.

baseterm   The term that you would like attached to your protien/gene names to make
the search as accurate as possible. I.e. "[Protein Name] AND Mus muscu-
lus[Organism]"

db   The database of the NCBI that you would like to access. I.e. protien

Examples

```
uids <- getUID(protien_gene_list, baseterm = "[Protein Name] AND Mus musculus[Organism]", db = "protien" )
```

---

humanMouse   *A function that takes a list of human genes and finds the mouse ho-
mologs*

Description

This function will accept a list of human gene names and finds their mouse homologs. This gene
list can then be used with the getUID function of this package.

Usage

```
humanMouse(genelist)
```

Arguments

genelist   A list of human genes

Examples

```
mousegenelist <- humanMSouse(humangenelist)
```

---

keyplayers	*List of keyplayers for protiens active in a diabetic mouse pancreas*

Description

List of protiens that are active in a diabetic mouse pancreas. This can be used to test the function of this package.

Usage

```
data(ketplayers)
```

Format

A list of proteins

Examples

```
data(keyplayers)

ids <- getUID(keyplayers, "[Protein Name] AND Mus musculus[Organism]", "protien" ))

proteindf <- fetchGenebank(ids$UIDs)
res <- unip(id$UIDs)

ptmdf <- readPTM(file = 'output.txt')
ptmdf <- ptm(ptmdf, res)
ptmdf <- ptmClean(ptmdf)

proteindf <- countPTMs(proteindf, ptmdf)
finaldf <- predictedMW(proteindf, ptmdf)
```

---

molecW	*A function that takes a dataframe produced by the fetchGenebank() function and returns PTMS and molecular weight*

Description

This function takes a dataframe produced by the fetchGenebank() function and returns PTMS and molcular weights via extra columns in the data frame. I can take other data frames if the sequences columns is named seq

Usage

```
molecW(df, monoisotopic = FALSE)
```

Arguments

df   df produced by the fetchGenebank() function or where the sequence column is
                named seq.

monoisotopic   A logical value TRUE or False indicating if monoistopic weights of amino-acids
                should be used

Examples

```
finaldf <- molecw(df)
```

| predictedMW | *A function that takes a protein dataframe produces by the countPTMs function along with the PTM data frame to return predicted molecular weight.* |

Description

This function takes a protein dataframe produces by the countPTMs function along with the PTM data frame to return predicted molecular weight. It uses the conversion data frame to map the PTMs to the correct molecular weight and to produce a predicted molecular weight based on the experimentally verfied PTM data from Uniprot.

Usage

```
predictedMW(proteindf, ptmdf)
```

Arguments

proteindf   Data frame of protein data generated by the fetchGenebank function.

ptmdf   Dataframe of PTM data generated by the PTM function.

Examples

```
finaldf <- predictedMW(proteindf, ptmdf)
```

ptm	*A function that takes a ptm data frame and a data frame of UIDs and UNIPROT IDs.*

Description

This function creates a PTM dataframe that links geneid from NCBI to the UNIPROT PTM data. This will allow this data to link experimintally verified PTM data from UNIPROT to protien data from NCBI.

Usage

```
ptm(ptmdf, res)
```

Arguments

ptmdf	Data frame of PTM data generated by the readPTM function
res	Data frame of UNIPROT IDs and UIDs produced by the unip function

Examples

```
ptmdf <- ptm(ptmdf, res)
```

ptmClean	*A function that takes a data frame of PTM modifications created by the PTM function.*

Description

This function will allow you to enter a list of dataframe of PTM data. It will remove spaces and normalize the names for the PTMs. It will also subset the ptmdf, so that it only includes PTMs that are in your dataset.

Usage

```
ptmClean(ptmdf)
```

Arguments

ptmdf	Dataframe of PTM data generated by the PTM function.

Examples

```
ptmdf <- ptmClean(ptmdf)
```

---

readPTM   *A function that takes an output file from the Uniprot command line tool*

Description

This function takes an output file from the command line tool built to access uniprot. This tool
takes a file generated by the unip function. You must request the data be returned as gff format.
Sometimes there is a header that must be removed before the gff data. This tool can be found here
https://github.com/jdrudolph/uniprot. It will return this as an R dataframe.

Usage

```
readPTM(ptmfile = "output.txt")
```

Arguments

ptmfile   Output file from uniprot command line tool

Examples

```
Command line Usage:
uniprot retrieve -f gff uniprot.txt output.txt

ptmdf <- readPTM(ptmfile = 'output.txt')
```

---

searchpro   *A function that searches your output based on MW*

Description

This function will allow you to search your data set based on a range of molecular weights

Usage

```
searchpro(df, range = 1:max(df$molecularweight_predicted), predicted = TRUE)
```

Arguments

df   Data frame following the formatting of data frames produced by the prodata
                 package

range   Range of values for the molecular weight that you would like to search for.
                 Default returns whole dataset.

predicted   If TRUE (default) it will use the molecular weight including the PTM calcula-
                 tions. If FALSE it will use the molecular weight without PTMs.

Examples

```
filtered <- searchpro(df, range = 1000:2000, predicted = TRUE)
```

| unip | *A function that takes a list of UID and returns a list of UNIPROT IDs* |

Description

This function converts UID into UNIPROT IDs for the retrival of PTM data from UNIPROT. It writes this to a file to be used in the comman line UNIPROT tool. It also returns a list of UIDs attached to UNIPROT ID. This will be used to build the ptmdf.

Usage

```
unip(UIDs)
```

Arguments

UIDs   List of NCBI UIDs.

Examples

```
res <- unip(UIDs)
```

Index