Package ‘ampir’

December 6, 2019

Type Package
Title Predict Antimicrobial Peptides
Version 0.1.0
Date 2019-11-26
Description A toolkit to predict antimicrobial peptides from protein sequences on a genome-wide scale. It incorporates a support vector machine model trained on publicly available antimicrobial peptide data using calculated physico-chemical and compositional sequence properties described in Meher et al. (2017) <doi:10.1038/srep42362>. In order to support genome-wide analyses this model is designed to accept any type of protein as input and calculation of compositional properties has been optimised for high-throughput use.

URL https://github.com/Legana/ampir
License GPL-2
Encoding UTF-8
LazyData true
Depends R (>= 3.5.0)
Imports Peptides, caret (>= 6.0.0), kernlab, Rcpp
RoxygenNote 7.0.0
Suggests testthat, knitr, rmarkdown
VignetteBuilder knitr
LinkingTo Rcpp
NeedsCompilation yes
Author Legana Fingerhut [aut, cre], Ira Cooke [aut], Jinlong Zhang [ctb] (R/read_faa.R), Nan Xiao [ctb] (R/calc_pseudo_comp.R)
Maintainer Legana Fingerhut <legana.fingerhut@my.jcu.edu.au>
Repository CRAN
Date/Publication 2019-12-06 09:50:13 UTC
R topics documented:

aaseq_is_valid  .................................................. 2
calculate_features ............................................. 3
calc_amphiphilicity ........................................... 3
calc_hydrophobicity .......................................... 4
calc_mw ................................................................. 4
calc_net_charge .................................................... 5
calc_pl ................................................................. 5
calc_pseudo_comp ............................................... 6
df_to_faa ............................................................. 6
predict_amps ....................................................... 7
read_faa .............................................................. 8

Index 9

| aaseq_is_valid | Check protein sequences for non-standard amino acids |

Description
Any proteins that contains an amino acid that is not one of the 20 standard amino acids is flagged as invalid.

Usage
aaseq_is_valid(seq)

Arguments

seq A vector of protein sequences

Value
A logical vector where TRUE indicates a valid protein sequence and FALSE indicates a sequence with invalid amino acids
**calculate_features**  
*Calculate a set of numerical features from protein sequences*

**Description**

This function calculates set physicochemical and compositional features from protein sequences.

**Usage**

```r
calculate_features(df, min_len = 20)
```

**Arguments**

- `df`: A dataframe which contains protein sequence names as the first column and amino acid sequence as the second column.
- `min_len`: Minimum length sequence for which features can be calculated. It is an error to provide sequences with length shorter than this.

**Value**

A dataframe containing numerical values related to the protein features of each given protein.

**Note**

This function depends on the Peptides package.

**References**


---

**calc_amphiphilicity**  
*Calculate amphiphilicity (or hydrophobic moment)*

**Description**

Calculate amphiphilicity (or hydrophobic moment)

**Usage**

```r
calc_amphiphilicity(seq)
```

**Arguments**

- `seq`: A protein sequence.
References
Osorio, D., Rondon-Villarreal, P. & Torres, R. Peptides: A package for data mining of antimicrobial
peptides. The R Journal. 7(1), 4–14 (2015). The imported function originates from the Peptides
package (https://github.com/dosorio/Peptides/).

calc_hydrophobicity Calculate the hydrophobicity

Description
Calculate the hydrophobicity

Usage
calc_hydrophobicity(seq)

Arguments
seq A protein sequence

References
Osorio, D., Rondon-Villarreal, P. & Torres, R. Peptides: A package for data mining of antimicrobial
peptides. The R Journal. 7(1), 4–14 (2015). The imported function originates from the Peptides
package (https://github.com/dosorio/Peptides/).

calc_mw Calculate the molecular weight

Description
Calculate the molecular weight

Usage
calc_mw(seq)

Arguments
seq A protein sequence

References
Osorio, D., Rondon-Villarreal, P. & Torres, R. Peptides: A package for data mining of antimicrobial
peptides. The R Journal. 7(1), 4–14 (2015). The imported function originates from the Peptides
package (https://github.com/dosorio/Peptides/).
### calc_net_charge

**Description**

Calculate the net charge

**Usage**

```
calc_net_charge(seq)
```

**Arguments**

- `seq` A protein sequence

**References**


---

### calc_pI

**Description**

Calculate the isoelectric point (pI)

**Usage**

```
calc_pI(seq)
```

**Arguments**

- `seq` pI

**References**

**calc_pseudo_comp**  
*Calculate the pseudo amino acid composition*

**Description**
This function is adapted from the extractPAAC function from the protr package (https://github.com/nanxstats/protr)

**Usage**
```r
calc_pseudo_comp(seq, lambda_min = 4, lambda_max = 19)
```

**Arguments**
- `seq`: A vector of protein sequences as character strings
- `lambda_min`: Minimum allowable lambda. It is an error to provide a protein sequence shorter than `lambda_min+1`
- `lambda_max`: For each sequence lambda will be set to one less than the sequence length or `lambda_max`, whichever is smaller

**References**

---

**df_to_faa**  
*Save a dataframe in FASTA format*

**Description**
This function writes a dataframe out as a FASTA format file

**Usage**
```r
df_to_faa(df, file = "")
```

**Arguments**
- `df`: a dataframe containing two columns: the sequence name and amino acid sequence itself
- `file`: file path to save the named file to

**Value**
A FASTA file where protein sequences are represented in two lines: The protein name preceded by a greater than symbol, and a new second line that contains the protein sequence
predict_amps

Examples

my_protein <- read_faa(system.file("extdata/bat_protein.fasta", package = "ampir"))

# Write a dataframe to a FASTA file
df_to_faa(my_protein, tempfile("my_protein.fasta", tempdir()))

predict_amps(faa_df, min_len = 5)

Arguments

faa_df A dataframe obtained from read_faa) containing two columns: the sequence name (seq_name) and amino acid sequence (seq_aa)

min_len The minimum protein length for which predictions will be generated

Value

The original input data.frame with a new column added called prob_AMP with the probability of that sequence to be an antimicrobial peptide. Any sequences that are too short or which contain invalid amin acids will have NA in this column

Examples

my_bat_faa_df <- read_faa(system.file("extdata/bat_protein.fasta", package = "ampir"))
predict_amps(my_bat_faa_df)
# seq_name   prob_AMP
# [1] G1P6H5_MYOLU 0.9723796
read_faa  
Read FASTA amino acids file into a dataframe

Description
This function reads a FASTA amino acids file into a dataframe

Usage
read_faa(file = NULL)

Arguments
file file path to the FASTA format file containing the protein sequences

Value
Dataframe containing the sequence name (seq_name) and sequence (seq_aa) columns

Note
This function was adapted from `read.fasta.R` by Jinlong Zhang (jinlongzhang01@gmail.com) for the phylotools package (http://github.com/helixcn/phylotools)

Examples

read_faa(system.file("extdata/bat_protein.fasta", package = "ampir"))

## Output
# seq_name  seq_aa
# [1] G1P6H5_MYOLU MALTVRIQAACL...
Index

aaseq_is_valid, 2

calc_amphiphilicity, 3
calc_hydrophobicity, 4
calc_mw, 4
calc_net_charge, 5
calc_pI, 5
calc_pseudo_comp, 6
calculate_features, 3
df_to_faa, 6

predict_amps, 7

read_faa, 8