Package 'Peptides'

October 16, 2014

Version 1.0.1

Date 2014-08-30

Title Calculate indices and theoretical physicochemical properties of peptides and protein sequences

Author Daniel Osorio, Paola Rondon-Villarreal and Rodrigo Torres.

Maintainer Daniel Osorio <daniel.osorio@correo.uis.edu.co>

URL https://github.com/dosorio/Peptides/

Depends seqinr , R (>= 2.10.0)

Suggests RUnit

Description Calculate physicochemical properties and indices from aminoacid sequences of peptides and proteins. Include also utilities for read and plot GROMACS output files .XVG.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-10-16 07:52:43

R topics documented:

`Peptides-package` *Calculate indices and theoretical physicochemical properties of peptides and proteins sequences.*

Description

Calculate physicochemical properties and indices from amino acid sequences of peptides and proteins. Include also utilities for read and plot GROMACS output files .XVG.

Details

Package:	Peptides
Type:	Package
Version:	1.0.1
Date:	2014-10-15
License:	GPL-2

Author(s)

Daniel Osorio, Paola Rondon-Villarreal and Rodrigo Torres.

Maintainer: Daniel Osorio <daniel.osorio@correo.uis.edu.co>

`aacomp` *Compute the amino acid composition of a protein sequence*

Description

This function calculates the amount of amino acids of a particular class and classified as: Tiny, Small, Aliphatic, Aromatic, Non-polar, Polar, Charged, Basic and Acidic based on their size and R-groups using same function implemented in EMBOSS "pepstat". The output is a matrix with the number and percentage of amino acids of a particular class

Usage

```
aacomp(seq)
```

Arguments

seq	amino acid sequence string

Value

The output is a matrix with the number and percentage of amino acids of a particular class

Tiny	(A + C + G + S + T)
Small	(A + B + C + D + G + N + P + S + T + V)
Aliphatic	(A + I + L + V)
Aromatic	(F + H + W + Y)
Non-polar	(A + C + F + G + I + L + M + P + V + W + Y)
Polar	(D + E + H + K + N + Q + R + S + T + Z)
Charged	(B + D + E + H + K + R + Z)
Basic	(H + K + R)
Acidic	(B + D + E + Z)

Note

This function was originally written by Alan Bleasby (ajb@ebi.ac.uk) for EMBOSS package. Further information: http://emboss.sourceforge.net/apps/cvs/emboss/apps/pepstats.html

References

Rice, Peter, Ian Longden, and Alan Bleasby. "EMBOSS: the European molecular biology open software suite." Trends in genetics 16.6 (2000): 276-277.

Examples

```
# COMPARED TO PEPSTATS
# http://emboss.bioinformatics.nl/cgi-bin/emboss/pepstats
# Property      Residues              Number  Mole%
# Tiny          (A+C+G+S+T)             4   19.048
# Small         (A+B+C+D+G+N+P+S+T+V)   4   19.048
# Aliphatic     (A+I+L+V)               5   23.810
# Aromatic      (F+H+W+Y)               5   23.810
# Non-polar     (A+C+F+G+I+L+M+P+V+W+Y) 11  52.381
# Polar         (D+E+H+K+N+Q+R+S+T+Z)   9   42.857
# Charged       (B+D+E+H+K+R+Z)         8   38.095
# Basic         (H+K+R)                 8   38.095
# Acidic        (B+D+E+Z)               0   00.000

## AA composition of PDB: 1D9J Cecropin Peptide
aacomp("KWKLFKKIGIGKFLHSAKKFX")
```

```
## Output
#          Number Mole%
# Tiny           4 19.05
# Small          4 19.05
# Aliphatic      5 23.81
# Aromatic       5 23.81
# Non Polar     11 52.38
# Polar          9 42.86
# Charged        8 38.10
# Basic          8 38.10
# Acidic         0  0.00
```

`aindex` *Compute the aliphatic index of a protein sequence*

Description

This function calculates the Ikai (1980) aliphatic index of a protein. The aindex is defined as the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins.

Usage

```
aindex(seq)
```

Arguments

`seq`	amino acid sequence string in upper case

References

Ikai (1980). Thermostability and aliphatic index of globular proteins. Journal of Biochemistry, 88(6), 1895-1898.

Examples

```
# COMPARED TO ExPASy ALIPHATIC INDEX
# http://web.expasy.org/protparam/

# SEQUENCE: SDKEVDEVDAALSDLEITLE
# Aliphatic index: 117.00

aindex("SDKEVDEVDAALSDLEITLE")
# [1] 117
```

boman *Compute the Boman (Potential Protein Interaction) index*

Description

This function computes the potential protein interaction index proposed by Boman (2003) based in the amino acid sequence of a protein. The index is equal to the sum of the solubility values for all residues in a sequence, it might give an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors, to normalize it is divided by the number of residues. A protein have high binding potential if the index value is higher than 2.48.

Usage

```
boman(seq)
```

Arguments

seq	aminoacid sequence string

References

Boman, H. G. (2003). Antibacterial peptides: basic facts and emerging concepts. Journal of internal medicine, 254(3), 197-215.

Radzicka, A., & Wolfenden, R. (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochemistry, 27(5), 1664-1670.

Examples

```
# COMPARED TO YADAMP DATABASE
# http://yadamp.unisa.it/showItem.aspx?yadampid=845&x=0,4373912
# SEQUENCE: FLPVLAGLTPSIVPKLVCLLTKKC
# BOMAN INDEX  -1.24

boman("FLPVLAGLTPSIVPKLVCLLTKKC")
# [1] -1.24
```

charge *Compute the theoretical net charge of a protein sequence*

Description

This function computes the net charge of a protein sequence based on the Henderson-Hasselbalch equation described by Moore, D. S. (1985). The net charge can be calculated at defined pH using one of the 9 pKa scales availables: Bjellqvist, EMBOSS, Murray, Sillero, Solomon, Stryer, Lehninger, Dawson or Rodwell

Usage

```
charge(seq, pH, pKscale)
```

Arguments

seq	amino acid sequence as string
pH	pH value
pKscale	a character string specifying the pKa scale to be used; must be one of "Bjellqvist", "EMBOSS", "Murray", "Sillero", "Solomon", "Stryer", "Lehninger", "Dawson" or "Rodwell"

References

Kiraga, J. (2008) Analysis and computer simulations of variability of isoelectric point of proteins in the proteomes. PhD thesis, University of Wroclaw, Poland.

Bjellqvist, B., Hughes, G.J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.Ch., Frutige,r S., Hochstrasser D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis, 14:1023-1031.

EMBOSS data are from http://emboss.sourceforge.net/apps/release/5.0/emboss/apps/iep.html.

Murray, R.K., Granner, D.K., Rodwell, V.W. (2006) Harper's illustrated Biochemistry. 27th edition. Published by The McGraw-Hill Companies.

Sillero, A., Maldonado, A. (2006) Isoelectric point determination of proteins and other macromolecules: oscillating method. Comput Biol Med., 36:157-166.

Solomon, T.W.G. (1998) Fundamentals of Organic Chemistry, 5th edition. Published by Wiley.

Stryer L. (1999) Biochemia. czwarta edycja. Wydawnictwo Naukowe PWN.

Examples

```
# COMPARED TO EMBOSS PEPSTATS
# http://emboss.bioinformatics.nl/cgi-bin/emboss/pepstats
# SEQUENCE: FLPVLAGLTPSIVPKLVCLLTKKC
# Charge   = 3.0

charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Bjellqvist")
# [1] 2.737
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="EMBOSS")
# [1] 2.914
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Murray")
# [1] 2.908
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Sillero")
# [1] 2.920
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Solomon")
# [1] 2.844
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Stryer")
# [1] 2.877
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Lehninger")
# [1] 2.873
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Dawson")
```

```
# [1] 2.844
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="Rodwell")
# [1] 2.820

# COMPARED TO YADAMP
# http://yadamp.unisa.it/showItem.aspx?yadampid=845&x=0,7055475
# SEQUENCE: FLPVLAGLTPSIVPKLVCLLTKKC
# CHARGE pH5: 3.00
# CHARGE pH7: 2.91
# CHARGE pH9: 1.09

charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=5, pKscale="EMBOSS")
# [1] 3.037
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=7, pKscale="EMBOSS")
# [1] 2.914
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=9, pKscale="EMBOSS")
# [1] 0.718

# JUST ONE COMMAND
charge(seq="FLPVLAGLTPSIVPKLVCLLTKKC",pH=seq(from = 5,to = 9,by = 2), pKscale="EMBOSS")
# [1] 3.037 2.914 0.718
```

H *22 Hydrophobicity values for amino acids form ExPASy "protscale"*

Description

A list with 22 hydrophobicity scales form ExPASy "protscale"

Usage

```
data(H)
```

Format

A list with 22 Hydrophobicity scales.

Details

Hydrophobicity scales are values that define relative hydrophobicity of amino acid residues.

Source

ExPASy-Protscale (http://web.expasy.org/protscale/)

References

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook (pp. 571-607). Humana Press.

Eisenberg D., Schwarz E., Komarony M., Wall R. Normalized consensus hydrophobicity scale. J. Mol. Biol. 179:125-142(1984).

Sweet R.M., Eisenberg D. Optimized matching hydrophobicity (OMH). J. Mol. Biol. 171:479-488(1983).

Hopp T.P., Woods K.R. Hydrophilicity. Proc. Natl. Acad. Sci. U.S.A. 78:3824-3828(1981).

Kyte J., Doolittle R.F. Hydropathicity. J. Mol. Biol. 157:105-132(1982).

Manavalan P., Ponnuswamy Average surrounding hydrophobicity. P.K. Nature 275:673-674(1978).

Abraham D.J., Leo A.J. Hydrophobicity (delta G1/2 cal). Proteins: Structure, Function and Genetics 2:130-152(1987).

Black S.D., Mould D.R. Hydrophobicity of physiological L-alpha amino acids. Anal. Biochem. 193:72-82(1991).

Bull H.B., Breese K. Hydrophobicity (free energy of transfer to surface in kcal/mole). Arch. Biochem. Biophys. 161:665-670(1974).

Fauchere J.-L., Pliska V.E. Hydrophobicity scale (pi-r). Eur. J. Med. Chem. 18:369-375(1983).

Guy H.R. Hydrophobicity scale based on free energy of transfer (kcal/mole). Biophys J. 47:61-70(1985).

Janin J. Free energy of transfer from inside to outside of a globular protein. Nature 277:491-492(1979).

Miyazawa S., Jernigen R.L. Hydrophobicity scale (contact energy derived from 3D data). Macromolecules 18:534-552(1985).

Rao M.J.K., Argos P. Membrane buried helix parameter. Biochim. Biophys. Acta 869:197-214(1986).

Roseman M.A. Hydrophobicity scale (pi-r). J. Mol. Biol. 200:513-522(1988).

Tanford C. Hydrophobicity scale (Contribution of hydrophobic interactions to the stability of the globular conformation of proteins). J. Am. Chem. Soc. 84:4240-4274(1962).

Wolfenden R.V., Andersson L., Cullis P.M., Southgate C.C.F. Hydration potential (kcal/mole) at 25C. Biochemistry 20:849-855(1981).

Wilson K.J., Honegger A., Stotzel R.P., Hughes G.J. Hydrophobic constants derived from HPLC peptide retention times. Biochem. J. 199:31-41(1981).

Parker J.M.R., Guo D., Hodges R.S. Hydrophilicity scale derived from HPLC peptide retention times. Biochemistry 25:5425-5431(1986).

Cowan R., Whittaker R.G. Hydrophobicity indices at ph 3.4 determined by HPLC. Peptide Research 3:75-80(1990).

Cowan R., Whittaker R.G. Hydrophobicity indices at ph 7.5 determined by HPLC. Peptide Research 3:75-80(1990).

Examples

```
data(H)
```

hmoment *Compute the hydrophobic moment of a protein sequence*

Description

This function compute the hmoment based on Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). Hydriphobic moment is a quantitative measure of the amphiphilicity perpendicular to the axis of any periodic peptide structure, such as the a-helix or b-sheet. It can be calculated for an amino acid sequence of N residues and their associated hydrophobicities Hn. If the secuence length is < 11 AA, the window length is equal to the AA sequence length, if it is > 11, windows of 11 residues are evaluated

Usage

```
hmoment(seq,angle)
```

Arguments

seq	amino acid sequence as string
angle	Protein rotational angle

Value

The max hydrophobic moment (uH) as a numerical vector of length one

References

Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. Proceedings of the National Academy of Sciences, 81(1), 140-144.

Examples

```
# COMPARED TO EMBOSS:HMOMENT
# http://emboss.bioinformatics.nl/cgi-bin/emboss/hmoment
# SEQUENCE: FLPVLAGLTPSIVPKLVCLLTKKC
# ALPHA-HELIX ANGLE=100 : 0.56
# BETA-SHEET  ANGLE=160 : 0.25

# ALPHA HELIX VALUE
hmoment("FLPVLAGLTPSIVPKLVCLLTKKC",100)
# [1] 0.56

# BETA SHEET VALUE
hmoment("FLPVLAGLTPSIVPKLVCLLTKKC",160)
# [1] 0.25
```

`hydrophobicity` *Compute the hydrophobicity index*

Description

This function calculates the GRAVY hydrophobicity index of an amino acids sequence using one of the 24 scales availables on ExPASy "protscale"

Usage

```
hydrophobicity(seq,scale)
```

Arguments

`seq`	amino acid sequence string in upper case
`scale`	a character string specifying the hydophobicity scale to be used; must be one of `"KyteDoolittle"`, `"AbrahamLeo"`, `"BullBreese"`,`"Guy"`, `"Miyazawa"`, `"Roseman"`, `"Wolfenden"`, `"Wilson"`, `"Cowan3.4"`, `"Aboderin"`, `"Sweet"`, `"Eisenberg"`, `"HoppWoods"`, `"Manavalan"`, `"BlackMould"`, `"Fauchere"`, `"Janin"`, `"Rao"`, `"Tanford"`, `"Cowan7.5"`, `"Chothia"` or `"Rose"`.

Source

http://web.expasy.org/protscale/

References

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook (pp. 571-607). Humana Press.

Eisenberg D., Schwarz E., Komarony M., Wall R. Normalized consensus hydrophobicity scale. J. Mol. Biol. 179:125-142(1984).

Sweet R.M., Eisenberg D. Optimized matching hydrophobicity (OMH). J. Mol. Biol. 171:479-488(1983).

Hopp T.P., Woods K.R. Hydrophilicity. Proc. Natl. Acad. Sci. U.S.A. 78:3824-3828(1981).

Kyte J., Doolittle R.F. Hydropathicity. J. Mol. Biol. 157:105-132(1982).

Manavalan P., Ponnuswamy Average surrounding hydrophobicity. P.K. Nature 275:673-674(1978).

Abraham D.J., Leo A.J. Hydrophobicity (delta G1/2 cal). Proteins: Structure, Function and Genetics 2:130-152(1987).

Black S.D., Mould D.R. Hydrophobicity of physiological L-alpha amino acids. Anal. Biochem. 193:72-82(1991).

Bull H.B., Breese K. Hydrophobicity (free energy of transfer to surface in kcal/mole). Arch. Biochem. Biophys. 161:665-670(1974).

Fauchere J.-L., Pliska V.E. Hydrophobicity scale (pi-r). Eur. J. Med. Chem. 18:369-375(1983).

Guy H.R. Hydrophobicity scale based on free energy of transfer (kcal/mole). Biophys J. 47:61-70(1985).

Janin J. Free energy of transfer from inside to outside of a globular protein. Nature 277:491-492(1979).

Miyazawa S., Jernigen R.L. Hydrophobicity scale (contact energy derived from 3D data). Macromolecules 18:534-552(1985).

Rao M.J.K., Argos P. Membrane buried helix parameter. Biochim. Biophys. Acta 869:197-214(1986).

Roseman M.A. Hydrophobicity scale (pi-r). J. Mol. Biol. 200:513-522(1988).

Tanford C. Hydrophobicity scale (Contribution of hydrophobic interactions to the stability of the globular conformation of proteins). J. Am. Chem. Soc. 84:4240-4274(1962).

Wolfenden R.V., Andersson L., Cullis P.M., Southgate C.C.F. Hydration potential (kcal/mole) at 25C. Biochemistry 20:849-855(1981).

Welling G.W., Weijer W.J., Van der Zee R., Welling-Wester S. Antigenicity value X 10. FEBS Lett. 188:215-218(1985).

Wilson K.J., Honegger A., Stotzel R.P., Hughes G.J. Hydrophobic constants derived from HPLC peptide retention times. Biochem. J. 199:31-41(1981).

Parker J.M.R., Guo D., Hodges R.S. Hydrophilicity scale derived from HPLC peptide retention times. Biochemistry 25:5425-5431(1986).

Cowan R., Whittaker R.G. Hydrophobicity indices at ph 3.4 determined by HPLC. Peptide Research 3:75-80(1990).

Cowan R., Whittaker R.G. Hydrophobicity indices at ph 7.5 determined by HPLC. Peptide Research 3:75-80(1990).

Rose G.D., Geselowitz A.R., Lesser G.J., Lee R.H., Zehfus M.H. Mean fractional area loss (f) [average area buried/standard state area]. Science 229:834-838(1985)

Examples

```
# COMPARED TO GRAVY Grand average of hydropathicity (GRAVY) ExPASy
# http://web.expasy.org/cgi-bin/protparam/protparam
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# GRAVY: -0.950

hydrophobicity("QWGRRCCGWGPGRRYCVRWC","KyteDoolittle")
# [1] -0.95
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","AbrahamLeo")
# [1] 0.09
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","BullBreese")
# [1] 0.16
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Guy")
# [1] 0.19
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Miyazawa")
# [1] 5.74
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Roseman")
# [1] -0.5
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Wolfenden")
```

```
# [1] -6.31
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Wilson")
# [1] 3.16
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Cowan3.4")
# [1] 0.08
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Aboderin")
# [1] 3.84
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Sweet")
# [1] -0.11
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Eisenberg")
# [1] -0.33
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","HoppWoods")
# [1] -0.14
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Manavalan")
# [1] 13.04
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","BlackMould")
# [1] 0.5
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Fauchere")
# [1] 0.53
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Janin")
# [1] -0.1
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Rao")
# [1] 0.81
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Tanford")
# [1] -0.29
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Cowan7.5")
# [1] 0.06
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Chothia")
# [1] 0.26
hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Rose")
# [1] 0.76
```

`instaindex` *Compute the instability index of a protein sequence*

Description

This function calculates the instability index proposed by Guruprasad (1990). A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

Usage

```
instaindex(seq)
```

Arguments

`seq`	amino acid sequence string

References

Guruprasad K, Reddy BV, Pandit MW (1990). "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence". Protein Eng. 4 (2): 155 - 61. doi:10.1093/protein/4.2.155

Examples

```
# COMPARED TO ExPASy INSTAINDEX
# http://web.expasy.org/protparam/
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# The instability index (II) is computed to be 83.68

instaindex("QWGRRCCGWGPGRRYCVRWC")
# [1] 83.68
```

---

lengthpep *Compute the amino acid length of a protein sequence*

Description

This function counts the number of amino acids in a protein sequence

Usage

```
lengthpep(seq)
```

Arguments

seq	amino acid sequence string

Value

Return the residue count as a numeric vector of length one

Examples

```
# COMPARED TO ExPASy ProtParam
# http://web.expasy.org/protparam
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# Number of amino acids: 20

lengthpep("QWGRRCCGWGPGRRYCVRWC")
# [1] 20
```

membpos *Compute theoretically the class of a protein sequence*

Description

This function calculates the theoretical class of a protein sequence based on the relationship between the hydrophobic moment and hydrophobicity scale proposed by Eisenberg (1984).

Usage

```
membpos(seq,angle)
```

Arguments

seq	amino acid sequence string
angle	Protein rotational angle

References

Eisenberg, David. "Three-dimensional structure of membrane and surface proteins." Annual review of biochemistry 53.1 (1984): 595-623.

Examples

```
membpos("ARQQNLFINFCLILIFLLLI",100)

#            Pep     H   uH       MembPos
#  1  ARQQNLFINFC -0.01 0.40      Globular
#  2  RQQNLFINFCL  0.03 0.40      Globular
#  3  QQNLFINFCLI  0.39 0.21      Globular
#  4  QNLFINFCLIL  0.56 0.29 Transmembrane
#  5  NLFINFCLILI  0.77 0.29       Surface
#  6  LFINFCLILIF  0.95 0.21       Surface
#  7  FINFCLILIFL  0.95 0.16 Transmembrane
#  8  INFCLILIFLL  0.93 0.15 Transmembrane
#  9  NFCLILIFLLL  0.90 0.15 Transmembrane
#  10 FCLILIFLLLI  1.10 0.16       Surface

membpos("ARQQNLFINFCLILIFLLLI",160)

#            Pep     H   uH       MembPos
#  1  ARQQNLFINFC -0.01 0.53      Globular
#  2  RQQNLFINFCL  0.03 0.44      Globular
#  3  QQNLFINFCLI  0.39 0.33      Globular
#  4  QNLFINFCLIL  0.56 0.33 Transmembrane
#  5  NLFINFCLILI  0.77 0.35       Surface
#  6  LFINFCLILIF  0.95 0.35       Surface
#  7  FINFCLILIFL  0.95 0.39       Surface
#  8  INFCLILIFLL  0.93 0.39       Surface
```

```
#  9  NFCLILIFLLL  0.90 0.18 Transmembrane
#  10 FCLILIFLLLI  1.10 0.18       Surface
```

---

mw *Compute the molecular weight of a protein sequence*

Description

This function calculates the molecular weight of a protein sequence. It is calculated as the sum of the mass of each amino acid using the scale available on Compute pI/Mw tool.

Usage

```
mw(seq)
```

Arguments

seq	amino acid sequence string

Source

The formula and amino acid scale are the same available on ExPASy Compute pI/Mw tool http://web.expasy.org/

References

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook (pp. 571-607). Humana Press. Chicago

Examples

```
# COMPARED TO ExPASy Compute pI/Mw tool
# http://web.expasy.org/compute_pi/
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# Theoretical pI/Mw: 9.88 / 2485.91

mw("QWGRRCCGWGPGRRYCVRWC")
# [1] 2485.9
```

pepdata *Physicochemical properties and indices from 100 amino acid sequences*

Description

Physicochemical properties and indices from 100 amino acid sequences (50 antimicrobial and 50 non antimicrobial)

Usage

```
data(pepdata)
```

Format

A data frame with 100 observations on the following 23 variables.

`sequence` a character vector with the sequences of 100 peptides (50 antimicrobial and 50 non-antimicrobial)

`group` Integrer vector with the group code `"0"` for non antimicrobial and `"1"` for antimicrobial

`length` a numeric vector with the length of the amino acid sequence

`mw` a numeric vector with the molecular weight of the amino acid sequence

`tinyAA` A numeric vector with the fraction (as percent) of tiny amino acids that make up the sequence

`smallAA` A numeric vector with the fraction (as percent) of small amino acids that make up the sequence

`aliphaticAA` A numeric vector with the fraction (as percent) of aliphatic amino acids that make up the sequence

`aromaticAA` A numeric vector with the fraction (as percent) of aromatic amino acids that make up the sequence

`nonpolarAA` A numeric vector with the fraction (as percent) of non-polar amino acids that make up the sequence

`polarAA` A numeric vector with the fraction (as percent) of polar amino acids that make up the sequence

`chargedAA` A numeric vector with the fraction (as percent) of charged amino acids that make up the sequence

`basicAA` A numeric vector with the fraction (as percent) of basic amino acids that make up the sequence

`acidicAA` A numeric vector with the fraction (as percent) of acid amino acids that make up the sequence

`charge` a numeric vector with the charge of the amino acid sequence

`pI` a numeric vector with the isoelectric point of the amino acid sequence

`aindex` a numeric vector with the aliphatic index of the amino acid sequence

`instaindex` a numeric vector with the instability index of the amino acid sequence

`boman` a numeric vector with the potential peptide-interaction index of the amino acid sequence

`hydrophobicity` a numeric vector with the hydrophobicity index of the amino acid sequence

`hmoment` a numeric vector with the hydrophobic moment of the amino acid sequence

`transmembrane` A numeric vector with the fraction of Transmembrane windows of 11 amino acids that make up the sequence

`surface` A numeric vector with the fraction of Surface windows of 11 amino acids that make up the sequence

`globular` A numeric vector with the fraction of Globular windows of 11 amino acids that make up the sequence

Examples

```
data(pepdata)
```

`pI` *Compute the isoelectic point (pI) of a protein sequence*

Description

The isoelectric point (pI), is the pH at which a particular molecule or surface carries no net electrical charge.

Usage

```
pI(seq,pKscale)
```

Arguments

`seq`	amino acid sequence string in upper case
`pKscale`	a character string specifying the pK scale to be used; must be one of `"Bjellqvist"`, `"EMBOSS"`, `"Murray"`, `"Sillero"`, `"Solomon"`, `"Stryer"`, `"Lehninger"`, `"Dawson"` or `"Rodwell"`

Examples

```
# COMPARED TO ExPASy ProtParam
# http://web.expasy.org/cgi-bin/protparam/protparam
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# Theoretical pI: 9.88

pI("QWGRRCCGWGPGRRYCVRWC","Bjellqvist")
# [1] 9.881

# COMPARED TO EMBOSS PEPSTATS
# http://emboss.bioinformatics.nl/cgi-bin/emboss/pepstats
```

```
# SEQUENCE: QWGRRCCGWGPGRRYCVRWC
# Isoelectric Point = 9.7158

pI("QWGRRCCGWGPGRRYCVRWC","EMBOSS")
# [1] 9.716

# OTHER SCALES

pI("QWGRRCCGWGPGRRYCVRWC","Murray")
# [1] 9.818
pI("QWGRRCCGWGPGRRYCVRWC","Sillero")
# [1] 9.89
pI("QWGRRCCGWGPGRRYCVRWC","Solomon")
# [1] 9.582
pI("QWGRRCCGWGPGRRYCVRWC","Stryer")
# [1] 9.623
pI("QWGRRCCGWGPGRRYCVRWC","Lehninger")
# [1] 9.931
pI("QWGRRCCGWGPGRRYCVRWC","Dawson")
# [1] 9.568
pI("QWGRRCCGWGPGRRYCVRWC","Rodwell")
# [1] 9.718
```

`pKscales` *9 pKa scales for the side chain of charged amino acids from various sources*

Description

9 pKa scales for the side chain of charged amino acids from various sources

Usage

```
data("pKscales")
```

Format

A data frame with the charged amino-acid, cTerm and nTerm in row and nine sources in column. The rownames are the one-letter code for amino-acids.

Source

Table 2 in Kiraga (2008)

References

Goloborodko, A.A.; Levitsky, L.I.; Ivanov, M.V.; and Gorshkov, M.V. (2013) "Pyteomics - a Python Framework for Exploratory Data Analysis and Rapid Software Prototyping in Proteomics", Journal of The American Society for Mass Spectrometry, 24(2), 301-304.

Kiraga, J. (2008) Analysis and computer simulations of variability of isoelectric point of proteins in the proteomes. PhD thesis, University of Wroclaw, Poland.

Bjellqvist, B., Hughes, G.J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.Ch., Frutige,r S., Hochstrasser D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis, 14:1023-1031.

EMBOSS data are from http://emboss.sourceforge.net/apps/release/5.0/emboss/apps/iep.html.

Murray, R.K., Granner, D.K., Rodwell, V.W. (2006) Harper's illustrated Biochemistry. 27th edition. Published by The McGraw-Hill Companies.

Sillero, A., Maldonado, A. (2006) Isoelectric point determination of proteins and other macromolecules: oscillating method. Comput Biol Med., 36:157-166.

Solomon, T.W.G. (1998) Fundamentals of Organic Chemistry, 5th edition. Published by Wiley.

Stryer L. (1999) Biochemia. czwarta edycja. Wydawnictwo Naukowe PWN.

Aronson, J. N. The Henderson-Hasselbalch equation revisited. Biochemical Education, 1983, 11 (2), 68.

Moore, D. S.. Amino acid and peptide net charges: A simple calculational procedure. Biochemical Education, 1986, 13 (1), 10-12.

Nelson, D. L.; Cox, M. M. Lehninger Principles of Biochemistry, Fourth Edition; W. H. Freeman, 2004; p. 1100.

Dawson, R. M. C.; Elliot, D. C.; Elliot, W. H.; Jones, K. M. Data for biochemical research. Oxford University Press, 1989; p. 592.

Rodwell, J. Heterogeneity of component bands in isoelectric focusing patterns. Analytical Biochemistry, 1982, 119 (2), 440-449.

Examples

```
data(pKscales)
```

`plot.xvg` *Plot time series from GROMACS XVG files*

Description

Read and plot output data from a XVG format file.

Usage

```
## S3 method for class 'xvg'
plot(x, ...)
```

Arguments

`x`	a .XVG output file from GROMACS.
`...`	ignored

Details

XVG is the default format file from GROMACS molecular dynamics package, contains data formatted for import into the Grace 2-D plotting program.

References

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29 (7), 845-854.

Examples

```
# PLOT ONE FILE
file <- system.file("xvg-files/epot.xvg",package="Peptides")
plot.xvg(file)

# PLOTTING MULTIPLE FILES
POPG <- system.file("xvg-files/POPG.xvg",package="Peptides")
POPE <- system.file("xvg-files/POPE.xvg",package="Peptides")
POPC <- system.file("xvg-files/POPC.xvg",package="Peptides")
par(mfcol=c(1,3))
plot.xvg(POPG)
plot.xvg(POPE)
plot.xvg(POPC)
```

---

read.xvg *Read XVG files from GROMACS Molecular Dynamics package*

Description

Read output data from a XVG format file.

Usage

```
read.xvg(file)
```

Arguments

file	a .XVG output file from GROMACS.

Details

XVG is the default format file from GROMACS molecular dynamics package, contains data formatted for import into the Grace 2-D plotting program.

References

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29 (7), 845-854.

Examples

```
# READING FILE
file <- system.file("xvg-files/epot.xvg",package="Peptides")
read.xvg(file)

#    Time (ps)  Potential
#  1         1 6672471040
#  2         2 6516461568
#  3         3 6351947264
#  4         4 6183133184
#  5         5 6015310336
#  6         6 5854271488
```

Index