Descriptors of Drug Targets

By: Noureddin Sadawi

Nov 2014

1 Basic/Statistical Descriptors from the R Package `Peptides`

Functions here perform calculates using the sequence as a `String` in uppercase. Observe that each of them returns one real value. More info about the `Peptides` package can be found at:
http://cran.r-project.org/web/packages/Peptides/

1. aindex(seq): 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.

2. boman(seq): 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

3. charge(seq, pH, pKscale): 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 , Dawson , EMBOSS , Lehninger , Murray , Rodwell , Sillero , Solomon or Stryer*
(Example: charge(seq="KLVCLLTKKC",pH=7, pKscale="Murray"))

4. hmoment(seq,angle,window): 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. An 11 residues window is used as default.

Example: hmoment(seq = "FLPVLAGLTPSIVPKLVCLLTKKC", angle = 100, window = 11)

5. hydrophobicity(seq,scale): This function calculates the GRAVY hydrophobicity index of an amino acids sequence using one of the 38 scales from different sources.
   A character string specifying the hydophobicity scale to be used; must be one of *"Aboderin","AbrahamLeo","Argos","BlackMould","BullBreese","Casari","C "Eisenberg","Engelman","Fasman","Fauchere","Goldsack","Guy","HoppWoods"," "Rao","Rose","Roseman","Sweet","Tanford","Welling","Wilson","Wolfenden"or"*
   Example: hydrophobicity("QWGRRCCGWGPGRRYCVRWC","Aboderin")
   or hydrophobicity("QWGRRCCGWGPGRRYCVRWC","AbrahamLeo")

6. instaindex(seq): 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.

7. lengthpep(seq): This function counts the number of amino acids in a protein sequence

8. mw(seq): 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

9. pI(seq): This function computes the isoelectric point (pI), which is the pH at which a particular molecule or surface carries no net electrical charge.

2 Amino Acid Sequence-based Descriptor from the R Package `ProtR`

More info here: `http://cran.r-project.org/web/packages/protr/`
The dipeptide composition gives a fixed pattern length of 400. The dipeptide composition encapsulates information about the fraction of amino acids as well as their local order. The dipeptide composition is calculated using the following equation:

$$\text{Fraction of dep(i)} = \frac{\text{total number of dep(i)}}{\text{Total number of all possible dipeptides}} \qquad (1)$$

Where $dep(i)$ is a dipeptide i out of 400 dipeptides
The required function from this package is called `extractDC(seq)` where *seq* is a character vector representing the input protein sequence.

More information about this descriptor can be found in this paper:
"M. Bhasin, G. P. S. Raghava. Classification of Nuclear Receptors Based on Amino Acid Composition and Dipeptide Composition. Journal of Biological Chemistry, 2004, 279, 23262"

Also, I have understood from reading through several papers that Andreas Bender (Uni of Cambridge) uses it frequently in his experiments. One paper says "these descriptors are simple, yet are known to be quite good at predicting protein structural classes, functional classes and subcellular localizations, Bhasin and Raghava 2004"