Biological Structure Analysis ‘Bio3D’ Package

Brianna Cabrera

# Introduction

Bio3d is an R package that contains utilities that are able to process, organize, and explore structure and sequence data. It is used in the Bioinformatics field which develops and improves on methods for storing, retrieving, organizing, and analyzing biological data. In the field of genetics and genomics, bioinformatics aids in sequencing and annotating genomes and their observed mutations. It also plays a role in the analysis of gene and protein expression and regulation. This R package was designed specifically to facilitate the analysis of multiple structures from both experiment and simulation. It makes it easy to work with structures of different compositions.

## Features

* Read, write, and process biomolecular structure, sequence and dynamics trajectory data
* Perform atom:
  + Selection
  + Alignment
  + Conformational clustering
  + Correlation network analysis
  + Torsion analysis
  + Distance matrix analysis
  + Principal component analysis of structure data
* Conservation analysis of sequence and structural data
* Integration of major protein structure and sequence databases associated search tools
* Utility fuctions provided to enable the statistical and graphical power of the R environment to work with biological sequence and structural data

## Installation

To install this R package, run this command at your R prompt:

#install.packages("bio3d")  
library(bio3d) #loads the Bio3D package at the R console prompt

Once installed, you can use this package at the R console, within R markdown documents.

## Basic Usage

Protein Data Bank files (PDB files) are the most common format for the distribution and storage of biomolecular coordinate data. At their most basic, PDB coordinate files contain a list of all the atoms of one or more molecular structures. A basic usage of this R package is reading a single PDB file using the read.pdb() function. For example, we can read and inspect the on-line file with PDB ID 4q21:

pdb <- read.pdb("4q21") #reads in PDB file and saves as an object "pdb"

## Note: Accessing on-line PDB file

pdb #prints file onto screen

##   
## Call: read.pdb(file = "4q21")  
##   
## Total Models#: 1  
## Total Atoms#: 1447, XYZs#: 4341 Chains#: 1 (values: A)  
##   
## Protein Atoms#: 1340 (residues/Calpha atoms#: 168)  
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)  
##   
## Non-protein/nucleic Atoms#: 107 (residues: 80)  
## Non-protein/nucleic resid values: [ GDP (1), HOH (78), MG (1) ]  
##   
## Protein sequence:  
## MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG  
## QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDL  
## AARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQHKL  
##   
## + attr: atom, xyz, seqres, helix, sheet,  
## calpha, remark, call

With being able to read a PDB file, we are able to view the number of atoms, it’s protein sequence, and more that you can analyze and manipulate.

## Next Steps

After reading a PDB file, the next step would be using different analyses for multiple PDB files at once. In order to achieve this, there are other vignettes within this package that provide different analyses for PDB files. You can find these different analyses here: <http://thegrantlab.org/bio3d/tutorials> These are more basic functions of this package.