Installing Bio3D

Lars Skjaerven, Xin-Qiu Yao and Barry J. Grant
University of Michigan, Ann Arbor

November 14, 2013

1 Background

Bio3D is an R package that provides interactive tools for structural bioinformatics. The primary focus of Bio3D is the analysis of biomolecular structure, sequence and trajectory data [1].

What can I do with Bio3D?

Features include the ability to read and write biomolecular structure, sequence and dynamic trajectory data, query and search online sequence and structure databases, perform atom selection, re-orientation, superposition, rigid core identification, clustering, distance matrix analysis, alignment, conservation analysis, normal mode analysis, principal component analysis, and many other common sequence and structural analysis tasks.

Why an R package?

Bio3D aims to leverage the extensive graphical and statistical capabilities of the R environment (http://www.r-project.org) and thus provide a useful integrated framework for the exploratory interactive analysis of biomolecular sequence and structure data.

What is the purpose of this document?

The aim of this vignette[2] is to provide Bio3D package installation instructions.

Where can I find more information?

The latest version of the package, full documentation and furhter vignettes can be obtained from the main Bio3D website: http://thegrantlab.org/bio3d/ and wiki: http://bio3d.pbwiki.com/.

What other vignettes are available?

Available Bio3D package vignettes can be found online http://thegrantlab.org/bio3d/html/ and within R once the Bio3D package is installed. To see available vignettes from within R use the R command:

```
vignette(package = "bio3d")
```

At the time of writing these include:

- `Installing Bio3D`
- `Getting started with Bio3D`
- `Comparative protein structure analysis with Bio3D`

[1] Grant, B.J. et al. (2006) *Bioinformatics*, **22**, 2695-2696
[2] This vignette contains executable examples, see `help(vignette)` for further details.

- `Sequence conservation analysis with Bio3D`
- `Beginning trajectory analysis with Bio3D`
- `Dynamic network analysis with Bio3D`
- `Normal mode analysis with Bio3D`

Side-note: We are always interested in adding additional functionality and documentation to the Bio3D package. If you have ideas or suggestions for improvements, or indeed code that you would like to distribute as part of this package, please contact us – we would like to hear from you!

2 Quick Installation for Linux/Ubuntu Users

Most required packages and programs are available from the official Ubuntu repository:

```
apt-get install r-base r-base-core netcdf-bin libnetcdf-dev libxml2-dev seaview muscle pymol
```

DSSP however is not, but can be installed by:

```
wget ftp://ftp.cmbi.ru.nl/pub/software/dssp/dssp-2.0.4-linux-amd64 -O /usr/local/bin/dssp
chmod a+x /usr/local/bin/dssp
```

Download the source code of the latest Bio3D version:

```
wget http://thegrantlab.org/bio3d/bio3d_2.0-1.tar.gz
```

Start R by issuing the command `R` and then from the R prompt install required packages, and finally the Bio3D package:

```
install.packages(c("ncdf", "lattice", "grid", "bigmemory", "multicore", "XML"),
    dependencies = TRUE)
install.packages("bio3d_2.0-1.tar.gz")
```

If everything worked as expected you can skip ahead to section 5. If you experienced errors you should continue to read section 3 and the installation instructions in section 4.1.

3 Installation Perquisites

Before you attempt to install Bio3D you should have a relatively recent version of R installed and working on your system (we recommend at least R version 3.0.1). Detailed instructions for obtaining and installing R on various platforms can be found on the R home page http://www.r-project.org.

Do I need to know R?

To get the most out of Bio3D you should be quite familiar with basic R usage. Some newcomers to R find this a steep learning curve. However, once you have mastered basic operations with vectors and matrices in R you should feel confident about getting stuck into using the Bio3D package.

Sidenote: There are now numerous on–line resources that can help you get started using R. Again they can be found from the main R website at http://www.r-project.org. We also maintain a list of R resources at http://bio3d.pbworks.com/w/page/68764093/Use_R. However, google may be your best friend in this regard.

Additional R Packages

Bio3D makes use of a number of additional R packages including *ncdf*, *lattice*, *grid*, *bigmemory*, *multicore*, and *XML*. These can be most easily installed from within R with the command:

```
install.packages(c("ncdf", "lattice", "grid", "bigmemory", "multicore", "XML"),
    dependencies = TRUE)
```

4 Obtaining and Installing Bio3D

The Bio3D package is available in two forms from http://thegrantlab.org/bio3d/

- as platform independent source code (intended primarily for Mac and Unix systems),
- as a compiled binary for Windows.

To install from source requires that your machine has standard compilers and tools such as Perl 5.004 or later. If you run into problems with source installation please refer to section 6.1 of the R Installation and Administration Manual.

4.1 Linux/Unix Installation

If you are unable to use the quick installation instructions described in section 2 in a Unix environment then you should download the latest Bio3D source tar.gz file from above. Then within an R session type:

```
install.packages("bio3d_2.0-1.tar.gz")
```

Or from the command line:

```
R CMD INSTALL bio3d_2.0-1.tar.gz
```

Sidenote: This will only work if you have permission to write files to the standard package installation location. If you would rather install to a different location you can set the `R_LIBS` environment variable to a location of your choice. For example, if you use tcsh/bash then add a line similar to the following to your .tcshrc/.bashrc file:

```
# csh:
setenv R_LIBS /home/myname/R/lib/R/library

# bash:
export R_LIBS=/home/myname/R/lib/R/library
```

Obviously you will want to change the path above to a directory of your choice.

4.2 MacOS X Installation

R on Mac OS X can be used either on the command-line, like on other Unix systems, or via the R.app GUI. If you prefer to use the command line based R system then simply follow the Unix installation instructions above.

Alternatively, you can use the `Packages and Data` menu of the GUI, in particular the sub-item Package Installer: Download the source tar.gz file from above. In the R GUI select `Packages and Data` → `Package Installer` → `Local Source Package`, and press the `Install` button. Select the Bio3D tar file and press `Open`.

4.3 Windows Installation

To install the Bio3D package on Windows download the compiled binary .zip file from above.

Start R and from GUI click `Packages` → `Install Package(s) from local zip file` then simply select your downloaded Bio3D zip file and click `Open` to finish the installation.

4.4 Installing the development version of Bio3D

For the majority of users we recommend the use of the last stable release available from the main Bio3D website. The development version is available from our bitbucket repository and typically contains new functions and bug fixes that have not yet been incorporated into the latest stable release.

There are several ways to download and install the development version of Bio3D. The simplest method is to install directly from our bitbucket repository using the R function `install_bitbucket()` from the `devtools` package.

```
install.packages("devtools")
library(devtools)
install_bitbucket("bio3d", username = "Grantlab", subdir = "ver_devel/bio3d/")
```

Alternative installation methods and additional instructions are posted to the wiki section of our bitbucket repository.

5 Additional utilities

There are a number of additional packages and programs that will either interface directly with Bio3D (MUSCLE, DSSP and STRIDE), or that we consider generally invaluable for working with biomolecular structure and sequence data (e.g. VMD, PyMOL, and SEAVIEW). A brief description of how to obtain these additional packages is given below.

5.1 Required for full Bio3D functionality

MUSCLE: Muscle is a fast multiple sequence alignment program available from the muscle home page `http://www.drive5.com/muscle`. The Bio3D functions `seqaln()` and `pdbaln()` currently calls the MUSCLE program, hence MUSCLE must be installed on your system and in the search path for executables if you wish to use this function.

A note for Mac and Unix users: After downloading MUSCLE, it should be unzipped and renamed to just "muscle" and placed in a directory such as "/usr/local/bin/".

DSSP: DSSP is another secondary structure analysis which should be installed on your system as an executable called "dssp" and be in the search path for executables. DSSP is available from a number of sources including:

- `http://www.cmbi.ru.nl/dssp.html`
- `http://swift.cmbi.ru.nl/gv/dssp/DSSP_5.html`.

5.2 Optional

STRIDE: STRIDE is a secondary structure analysis program available from the EMBL-Heidelberg. Stride is similar in functionality to the more prevalent DSSP (see below). However, stride is often much easer to setup on different computer systems as you may be able to simply copy or link to the stride executable distributed within every version of VMD (see above).

SEAVIEW: SEAVIEW is a graphical multiple sequence alignment editor. Download information and documentation are available from PBIL http://pbil.univ-lyon1.fr/software/seaview.html. I use Seaview to manually check and edit protein sequence alignment files pior to detailed analysis. I believe this should be done with every alignment regardless of how accurate the various automatic tools are supposed to be.

VMD: VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics. Visit the VMD website for download information and documentation http://www.ks.uiuc.edu/Research/vmd/. Along with the standard documentation you may find my VMD cheat sheet useful. I have also included a link on this page to my .vmdrc file which includes a number of timesaving customizations (see the cheat sheet for full details).

PyMOL: PyMOL is another visualization program with overlapping functionality with VMD. Bio3D functions `view.dccm()` and `view.modes()` require PyMOL in the search path. PyMOL is open-source software and available from http://www.pymol.org.

6 Testing your installation

You should now be able to load the Bio3D library into your current R session by typing the usual `library(bio3d)` command at the R Console.

```
library(bio3d)
help(package = "bio3d")
vignette(package = "bio3d")
```

We now suggest you use the command `demo("pdb")`, `demo("pca")` and `demo("md")` to get a quick feel for some of the tasks that we will be introducing in subsequent vignettes:

```
library(bio3d)
demo("pdb")
demo("pca")
demo("md")
```

7 Where to next

If you have read this far, congratulations! We are ready to have some fun and move to other package vignettes that describe various analysis including basic Molecular Dynamics Trajectory Analysis, Correlation Network Analysis (where we will build and dissect dynamic networks form different correlated motion data), enhanced methods for Normal Mode Analysis (where we will explore the dynamics of large protein families and superfamilies), and advanced Comparative Structure Analysis (where we will mine available experimental data and supplement it with simulation results to map the conformational dynamics and coupled motions of proteins).

8 Session Information

The version number of R and packages loaded for generating the vignette were:

```
sessionInfo()

## R version 3.0.2 (2013-09-25)
## Platform: x86_64-apple-darwin10.8.0 (64-bit)
##
```

```
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] tools     stats     graphics  utils     datasets  grDevices methods
## [8] base
##
## other attached packages:
## [1] knitr_1.5   bio3d_2.0-1
##
## loaded via a namespace (and not attached):
## [1] evaluate_0.5.1 formatR_0.9    highr_0.2.1    stringr_0.6.2
```