RBMRB notes

Jean-Yves Sgro

February 22, 2018

Contents

This is a summary of notes during the presentation of guest speaker Dr Kumaran Baskaran, author of the RBMRB package

Note: the interactive plots are only visible on the HTML version of this document. PDF and Docx versions will not show them.

1 Biological Magnetic Resonance Data Bank

RBRMB is an R package to access data from the **Biological Magnetic Resonance Data Bank** - BMRB, a repository for data from NMR spectroscopy on proteins, peptides, nucleic acids, and other biomolecules.

This short document is a set of notes from a "hands-on" session, but more complete "how to" and descriptions can be found on the RBMRB documentation[1].

2 Installation

The RBMRB can be installed from the R line command with:

```
install.packages('RBMRB)
```

---

[1] http://www.bmrb.wisc.edu/RBMRB_Doc/index.shtml

or from the `Packages` tab within **[RStudio]https://www.rstudio.com/products/rstudio/download/)**.

Note: There are many dependencies that will be installed automatically at the same time.

3   Activating RBMRB

Before it can be used the package needs to be loaded within `R` with the command:

```
library(RBMRB)
```

4   Data access.

The RBMRB package contains commands to access data directly from the within the web page of **Biological Magnetic Resonance Data Bank** - BMRB online database.

At the top of the web page is a search field to request accession code of compounds of interest.

The database contains on **chemical shift** data, while 3D coordinates are available from the Protein Data Bank{^2].

Therefore wtih the RBMRB package we will only be interested in ***chemical shift*** *meta data. These can be fetched by accession code (see above.) For example here we get data for entry* **15060** *(*NMR structure of the murine DLC2 (deleted in liver cancer -2) SAM (sterile alpha motif) domain*.)*

We'll save the chemical shift data in the `R` object `cs`:

```
cs <- fetch_entry_chemical_shifts(15060)
```

You can explore the data table with these commands:

To see a spreadsheet like format:

```
View(cs)
```

The see the first 6 lines of data:

```
head(cs)
```

You will see that the data contains 24 variables (columns.) If you explore further you may also find that there are 974 observations (rows.) This can be confirmed by checking the dimension:

```
dim(cs)
```

```
[1] 974  24
```

We are mostly interested in the column `Val` containing the **value** of the chemical shift. This is the 11th column in the data. We could access this column simply by using the command:

```
cs$Val
```

which would display all values within that column on the screen.

The following command is meant to simplify the data by keeping only the relevant columns. The command `convert_cs_to_n15hsqc` is part of the RBMRB package. See `?convert_cs_to_n15hsqc` for help.
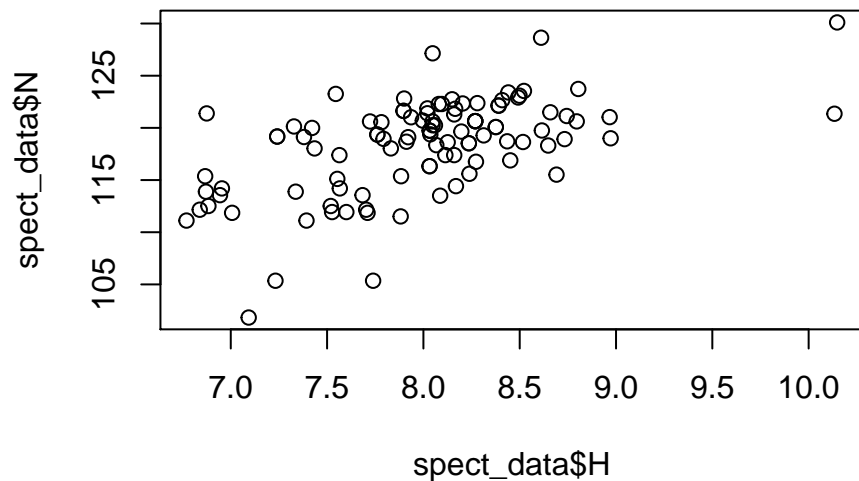
```
spect_data <- convert_cs_to_n15hsqc(cs)
```

5 Plots

5.1 Standard graphics

The usualy "base `R`" commands are available, by choosing the columns we need. For example:

```
plot(spect_data$H, spect_data$N)
```


Which could also be written as

```
with(spect_data, plot(H,N))
```

5.2 Built-in graphical functions

The package contains built-in graphical functions that take advantage of modern plotting packages installed as dependencies such as `ggplot2` and `plotly` providing interactive plots. This is one reason why a plot is sent first towards an `R` object, here `plt`.

```
plt <- HSQC_15N(15060)
plt
```

5.2.1 TOCSY PLOT

See Wikipedia definition of TOCSY[2].

This plot will connect all oxygen atoms of one residue to try to identify amino acids.

```
plt2 <- TOCSY(15060)
plt2
```

However this plot is very busy and not so useful here.

5.2.2 Compare 2 chemical shift spectra

In BMRB assigments have already been done and can be used for comparison.

---

[2]https://en.wikipedia.org/wiki/Two-dimensional_nuclear_magnetic_resonance_spectroscopy#Total_correlation_spectroscopy_(TOCSY)

5.2.2.1 Example: compare 2 Ubiquitins

```
plt3 <- HSQC_15N(c(11505, 11547))
plt3
```

Each entry will have a different color

The plot can be modified by adding a connecting line between related residues:

```
plt3 <- HSQC_15N(c(11505, 11547), 'line')
plt3
```

> Important note: those far away on the plot represent change in frequency *NOT* spacial distance! An example where this can happen could be closeness to water.

6 pH titration.

```
plt4 <- HSQC_15N(18857)
plt4
```

Modify to connect by line to see the changes:

```
plt4 <- HSQC_15N(18857, 'line')
plt4
```

7 Example from help

The command `HSQC_15N` simulates H1-N15 HSQC spectra for a given entry or list of entries from BMRB. More information is available in the help:

```
?HSQC_15N
```

One example within the help page can be used to try to find 'active sites' by interaction with ligand(s).

```
plot_hsqc<-HSQC_15N(c(17074,17076,17077),'line')
plot_hsqc
```

8 Histograms

The plot can be done in 2 ways:

- download the data first
- or plot directly with built-in command

  Note: many of the command download ALL the data from the database and may require time and…. lots computer RAM!

Example: all CB atoms for all ALA residues in the whole database:

```
cs_hist <- chemical_shift_hist('ALA','CB')
cs_hist
```

A similar command will apply to all atoms of ALA residues by replacing `CB` with `*`.

```
cs_hist <- chemical_shift_hist('ALA','*')
cs_hist
```

Here we plot for the `CB` atoms of all residues within the whole database:

```
cs_hist <- chemical_shift_hist('*','CB')
cs_hist
```

9 FILTER

WARNING - huge data size that takes long time and lots of RAM.

```
# DO NOT RUN BY DEFAULT
# df <- filter_residue(fetch_res_chemical_shifts('*'))
```

10 HELP

When asking for help as *e.g.* we did above with `?HSQC_15N` there is a link at the very bottom that looks like:

```
[Package RBMRB version 2.1.2 Index]
```

where the word "Index" is a link to the page containing links to all commands from the package.