

RaFoSA: Random Forests Secondary Structure Assignment When All-Atom information is Lacking

Concise User Guide

1.0 How to Obtain and Use RaFoSA

1. User should download RaFoSA from <http://bioinformatics.center/RaFoSA> (Fig. 1a).
2. User should then, extract contents of the zip file to obtain the file and folders listed in Fig. 1b.
3. RaFoSA can then be used to assign secondary structure (SS) to residues of a protein molecule (even when all-atom information is lacking) by invoking the command
`"python full/path/to/RaFoSA.py full/path/to/pdbFile.pdb"`
as shown in Fig. 1c. RaFoSA is written in Python Programming Language.

NOTE: Depending on your current python installation, you may need to install some important python libraries that may not come pre-installed in your python distribution. You can use `"easy_install nameOfPackage"` or `"pip install nameOfPackage"` to install each of the packages. If you do not know which of the additional packages you already have, just try to RaFoSA first and you will find out.

Additional Python Packages that may be required are

1. NumPy (<http://www.numpy.org>)
Installation Command: `easy_install numpy`
Alternative Installation Command: `pip install numpy`
2. BioPython (http://biopython.org/wiki/Main_Page)
Installation Command: `easy_install Biopython`
Alternative Installation Command: `pip install Biopython`
3. SciKitLearn (<http://scikit-learn.org/stable>)
Installation Command: `easy_install scikit-learn`
Alternative Installation Command: `pip install scikit-learn`

More information about easy_install can be found here <http://peak.telecommunity.com/DevCenter/EasyInstall>.
More information about pip can be found here <https://pip.pypa.io/en/stable/>. For further assistance, please email us through emmanuel@gapp.nthu.edu.tw.

4. Three important output files (`pdbId.rafosa8`, `pdbId.rafosa3`, and `pdbId.vmd`) would be generated, as shown in the example in Fig. 1d.
5. The (text) files contain the SS assignment using 7-class SS ("HBEGITS", `pdbId.rafosa8`), SS assignment using 3-class SS ("hcscccc", `pdbId.rafosa3`) based on M1 mapping ("HBEGITS" > "hcscccc", see the main text for more details), and VMD visualization state (`pdbId.vmd`, for visualization of the molecule in VMD). **It is `pdbId.rafosa3` that most users need. Helix is represented by "h", coil by "c", and sheet by "s".**
6. Example of the contents of `pdbId.rafosa8` and `pdbId.rafosa3` are shown in Fig. 1e.
7. To use `pdbId.vmd`, the user should start VMD as shown in Fig. 1f.
8. Then, click "File" menu and select "Load Visualization State" as shown in Fig. 1g.
9. The user should then select the VMD visualization state file (`pdbId.vmd`) and click "OK" as shown in Fig. 1h.
10. The molecule is then rendered, as shown in Fig. 1i, based on information from RaFoSA's SS assignment.

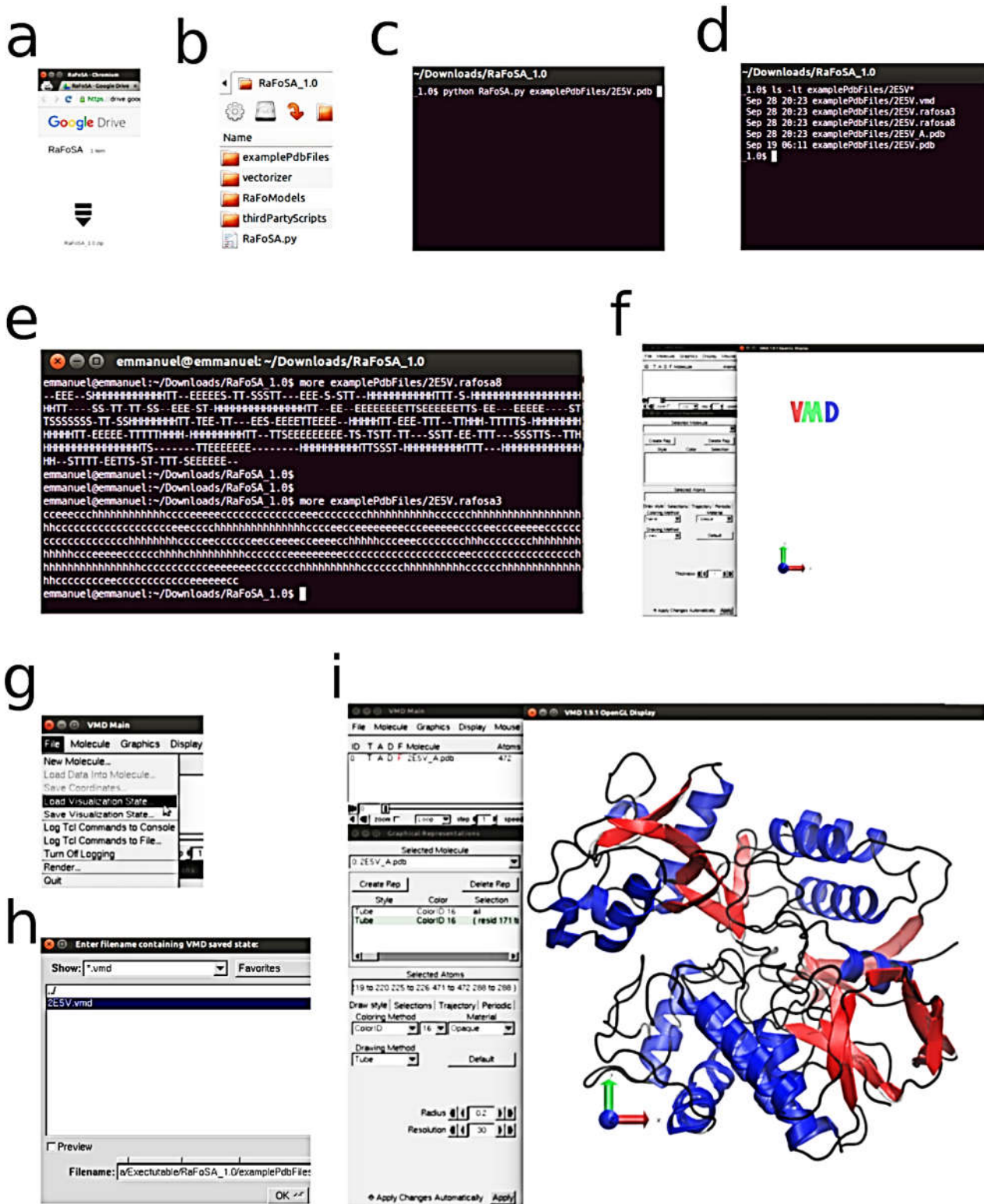


Fig. 1 How to obtain and use RaFoSA. **a.** User should download RaFoSA from <http://bioinformatics.center/RaFoSA>. Then, extract contents of the zip file to obtain the file and folders listed in **b.** RaFoSA can then be used to assign secondary structure (SS) to residues of a protein molecule (even when all-atom information is lacking) by invoking the command `python full/path/to/RaFoSA.py full/path/to/pdbFile.pdb` as shown in **c.** Three important output files (`pdbId.rafosa8`, `pdbId.rafosa3`, and `pdbId.vmd`) are generated as shown in the example in **d.** The (text) files contain the SS assignment using 7-class SS ("`HBEGITS`", `pdbId.rafosa8`), SS assignment using 3-class SS ("`hcsccec`", `pdbId.rafosa3`) based on M1 mapping ("`HBEGITS`"-> "`hcsccec`", see the main text for more details), and VMD visualization state (`pdbId.vmd`, for visualization of the molecule in VMD). The contents of `pdbId.rafosa8` and `pdbId.rafosa3` are shown in **e.** To use `pdbId.vmd`, the user should start VMD as shown in **f.** Then, click "File" menu and select "Load Visualization State" as shown in **g.** The user should then select the VMD visualization state file (`pdbId.vmd`) and click "OK" as shown in **h.** The molecule is then rendered, as shown in **i,** based on information from RaFoSA's SS assignment.

2.0 How to Get Further Assistance

If you require further assistance, please contact the lead developer through the following contact details

Email: **emmanuel@gapp.nthu.edu.tw**
Tel: **+18133888836**