

script manual: **hp_plotter.rb**

Summary

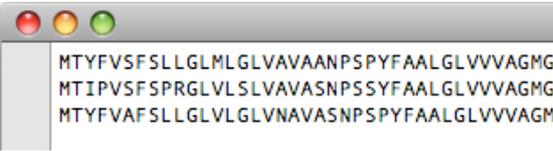
This script calculates the hydropathicity of a batch of amino acid sequences according to Kyte & Doolittle (1982). It prepares a tab separated file that can be read by Excel. You are then able to plot all trajectories into one graph, or to calculate statistics on your hydropathicity data (e.g. whether one or more of the sequences differs significantly).

Method

The method is best described in Kyte and Doolittle (1982). Hydropathicity values are attributed to each amino acid. The most hydrophobic amino acid, isoleucine, has a hydropathicity of 4.5, while the most hydrophilic, arginine, has -4.5. Using a sliding window analysis, the sum of all amino acids within the window is assigned to the central position of the window.

Input file

hp_plotter.rb expects a plain vanilla text file with nothing but amino acid sequences, one sequence per line. Sequences may vary in length, but no gaps should be included. Just make sure, that it looks a bit like this:



```
MTYFVSFSL LGLMLGLVAVAAANPSYFAALGLVVVAGMG
MTIPVSFSPRGLVLSLVAVASNPSSYFAALGLVVVAGMG
MTYFVAFSL LGLV LGLVNAVASNPSPYFAALGLVVVAGM
```

For simplicity sake, put the input file in the folder same folder as hp_plotter.rb.

Output file

A simple tab-delimited file named 'hp_plotter.out' is written, in the folder where hp_plotter is. Old files of the same name are overwritten. The file can be read with MS Excel and should look like the image on page 2. The first column has the sequence positions, while the hydropathicity values of sequences 1 to 3 are written in columns 2 to 4. The first position is always the center of the start position of the sliding window. If you choose the default window size of 9, the first entry will be that of sequence position 5. In the case of the example image on page 2, the window size was obviously set to 5, since the first entry is that of position 3.

Plots can then be done in traditional Excel style.

	A	B	C	D
1	Position	Sequence 1	Sequence 2	Sequence 3
2	3	6.9	6.9	6.9
3	4	4.2	4.2	4.2
4	5	7.7	7.7	7.7
5	6	8.2	8.2	8.2
6	7	5.3	5.3	5.3
7	8	1	1	1
8	9	1.4	1.4	1.4
9	10	-1.5	-1.5	-1.5
10	11	3.5	3.5	3.5
11	12	3.5	3.5	3.5
12	13	3.2	3.2	3.2
13	14	3.5	3.5	3.5
14	15	7.8	7.8	7.8
15	16	5.4	5.4	5.4
16	17	9.7	9.7	9.7
17	18	11.9	11.9	11.9
18	19	11.2	11.2	11.2
19	20	3.5	3.5	3.5
20	21	0.1	0.1	0.1
21	22	-4.9	-4.9	-4.9
22	23	-8.3	-8.3	-8.3
23	24	-8.8	-8.8	-8.8
24	25	-2.5	-2.5	-2.5
25	26	0.9	0.9	0.9

Application

As `hp_plotter.rb` is written in the interpreted script language Ruby, this language must be installed on your system. No problem if you're on a Mac, Ruby is preinstalled since Mac OS 10.4 (Tiger). If your system runs on Windows, you'll have to find out how to install Ruby.

Make sure that `hp_plotter.rb` and your sequence file are in the same directory. Use the Terminal or a similar command-line tool to navigate to that directory. Then, start the script by typing

```
ruby hp_plotter.rb yourfilename
```

If you don't want to use the default size of the sliding window (which is 9), then type, for example

```
ruby hp_plotter.rb yourfilename 11
```

The script will only run properly if the window size is an odd number. This is not a bug, since window needs a central position its size must thus be odd.

Kyte, J., and R. F. Doolittle (1982) A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* 157:105-132