



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Peptide Pattern Recognition

Busk, Peter Kamp; Lange, Lene

Publication date:
2013

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Busk, P. K., & Lange, L. (2013). Peptide Pattern Recognition. Computer programme

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Peptide Pattern Recognition

The tool for finding short, conserved peptides is called Peptide Pattern Recognition (PPR) and consist of the executable file "ppr.exe".

The tool was written for Windows and has been tested under Windows XP and Windows 7.

Manual installation:

- 1) Download "ppr.exe" to a folder on your computer.
- 2) Optional: Download the example file "gh61_gh45.txt" and this manual called "READ ME" to the same folder.

Automatic installation on the C: drive depending on your browser:

With Chrome:

- 1) Click on the document called "PPR installer".
- 2) Double click on the downloaded file called "2installer_ppr.exe" at the bottom of the browser window.
- 3) Allow the program to make changes to your computer.
- 4) Follow the instructions.

With Firefox:

- 1) Click on the document called "PPR installer".
- 2) Save the file.
- 3) Double click on the downloaded file called "2installer_ppr.exe" in the "Downloads" window.
- 4) Allow the program to make changes to your computer.
- 5) Follow the instructions.

With Internet Explorer:

- 1) Click on the document called "PPR installer".
- 2) Click on "Run" at the bottom of the browser window.
- 3) Click on "Run" once more at the bottom of the browser window or, if your computer comes up with some security warnings, grant the necessary permissions to run the installer.
- 4) Follow the instructions.

To use PPR:

1) Place a file containing the sequences you want to analyze in the same folder as "ppr.exe".

a) The sequences should be in strict fasta format:

```
>name of sequence  
sequence
```

b) The file should be a ".txt" file.

c) The preferred name for the file is *protein_family_unique_sequences.txt* or *protein_family.txt*. The file "gh61_gh45.txt" is an example of the format of the input.

d) If there is doubt about the format of the input it may be a good idea to download "format2fasta.exe" from [http://vbn.aau.dk/en/publications/format2fastaexe\(32d3be8d-d204-4780-970c-c5dad652d53\).html](http://vbn.aau.dk/en/publications/format2fastaexe(32d3be8d-d204-4780-970c-c5dad652d53).html) and format the input with this tool before the PPR analysis.

2) When the input is ready the PPR analysis can be started by double clicking the "ppr.exe" icon. This will open a small dos window.

3) Enter the requested information (file name and parameters) and run the analysis.

4) The result will be stored in a number of text file in a subfolder named after the input file name and the parameters used for the PPR analysis.

5) Press "enter" to close the dos window.

The PPR analysis can be aborted at any time by closing the dos window. This may leave a number of intermediate text files in the folder with "ppr.exe". The files will be deleted automatically when PPR is executed again.

Interpretation of the result:

The result of a PPR analysis consist of a number of files containing groups of protein sequences and a number of files containing lists of peptides corresponding to the protein groups.

The protein files are named "[protein name]_group_1.txt", "[protein name]_group_2.txt" and so on. (For example "gh61_gh45_group_1.txt"). These files can be opened with MS Word or another text editor.

Each protein in the file has been assigned a score, which is the number of peptides from the corresponding peptide list, that can be found in the protein sequence.

The peptide files are named "[protein name]_group_1_peps.txt", "[protein name]_group_2_peps.txt" and so on. (For example "gh61_gh45_group_1_peps.txt"). These files can be opened with MS Excel, Open Office Calc or similar.

The peptides are listed from the fourth row and downwards.

The first column indicates the median position of the first amino acid in the peptide in all the proteins that contain the peptide.

The second row contain the sequence of the peptide in one letter code.

The third row contain the frequency of the peptide calculated as number of proteins that contain the peptide divided by number of proteins in the group.

The fourth row contain the number of proteins that contain the peptide.

Good luck!

Peter Busk, 25.February.2013

The software is provided 'as is', without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.