

ProPAS: standalone software to analyze protein properties

Songfeng Wu & Yunping Zhu*

State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, 27 Taiping Road, China; Beijing 100850, Yunping Zhu - Email: zhuyunping@gmail.com; Phone: 8610-80705225; *Corresponding author

Received December 29, 2011; Accepted January 07, 2012; Published February 03, 2012

Abstract:

Physicochemical properties of proteins/peptides are often used to separate proteins/peptides from complex biological samples in proteomics research, which is necessary for the further proteins/peptides identification. So, the analyses of these properties are very important for the design of proteomic experiments and the further analyses of identified results. Current available web-based tools are difficult to analyze large-scale datasets. Here we provide a standalone software ProPAS (Protein Properties Analyses Software) to calculate these properties in local computers. ProPAS could be used to calculate several physicochemical parameters of proteins/peptides, including the isoelectric point (pI), hydrophobicity (Hy) and mass weight (MW). Additionally, ProPAS provides the statistic results to get the number distribution along the parameter values. The results could be copied to other software such as Excel to get the distribution chart. The amino acid composition of all the input proteins/peptides could also be calculated from all input sequences. Detail description, updates and download link for ProPAS could be found at the website of <http://bioinfo.hupo.org.cn/tools/ProPAS/propas.htm>.

Background:

Protein identification is a basic task for proteomics research. Usually, the proteins/peptides need to be separated according to their physicochemical properties before the identification [1]. The common used physicochemical properties include isoelectric point, hydrophobicity and mass weight. So, the calculation of these physicochemical properties is very important and could be used for the design of proteomic experiments. Further analyses of physicochemical properties for the identified proteins might be helpful to discover the experimental bias, which would be important for improving the experiments. The analyses could also be used to discover the different distribution caused by some biological factors, for example the different isoelectric point and hydrophobicity in different sub-cellular localization [2, 3]. Most current available softwares for these physicochemical properties calculation are web-based, including the software in ExPASy [4]. The web-based softwares are difficult for the analyses of large-scale datasets. ProPAS provide an alternative choice to analyze the data in local computer, which could be used to process small or large-scale datasets. Additionally, ProPAS could also provide the statistic results to get the

protein/peptide number distribution of these physicochemical properties.

Methodology:

Calculation of isoelectric point

Isoelectric point (pI) was estimated as the pH at which a protein took on a neutral charge ($R=0$) [5]. (Please see supplementary material for definition of R).

Calculation of hydrophobicity

Hydrophobicity value for each protein is calculated as the mean value of hydrophobic index for all amino acids. In this software, Kyte-Doolittle scale [6] is used as the hydrophobic index.

Calculation of mass weight

Protein mass weight (MW) is calculated by the addition of average isotopic masses of amino acids in the protein and the average isotopic mass of one water molecule. The mass weight values for the amino acids are from ExPASy [7], including 20 standard amino acids and 2 non-standard amino acids (Selenocysteine and Pyrrolysine). Additionally, the

characters B (Asx: Aspartic acid or Asparagine), Z (Glx: Glutamine or Glutamic acid) and X (Xaa: Any amino acid) are also accepted for the calculation, whose mass values are calculated as the mean values of the corresponding amino acid masses weighted by their respective frequencies. In this program, their mass values are: B (114.6686), Z (128.7531) and X (111.1138).

Calculation of amino acid composition

Amino acid composition is calculated as the count and the percentage for each amino acid from all the input proteins/peptides. The software was developed by Perl script. **Figure 1** shows the interface and its functional modules of the software.

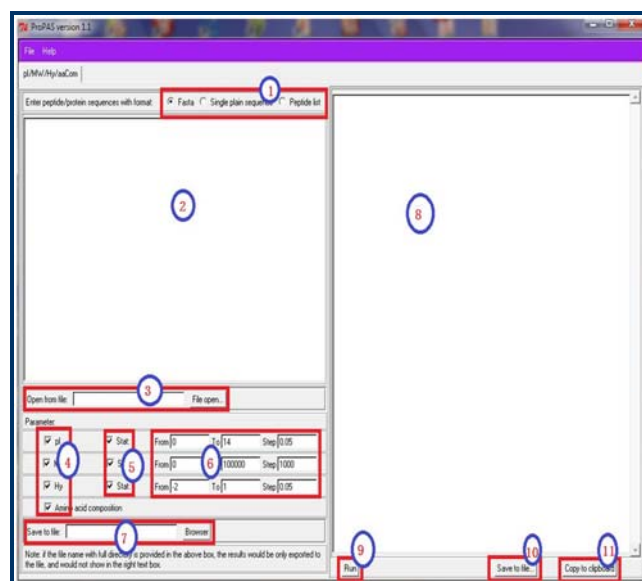


Figure 1: The interface of ProPAS. The different functional modules are showed in the interface: (1) input format options; (2) text box for input data; (3) text box for input file and its directory; (4) parameters that could be analyzed; (5) need statistic or not; (6) range and step of the parameter for the statistic; (7) file and directory that the results will be exported to; (8) text box for output data; (9) button to calculate the properties; (10) Button to save the results showed in text box "8" to file; (11) Button to copy the results showed in text box "8" to clipboard.

Input

The input data for the software includes

(a) One or more protein sequences with Fasta format ("Fasta" in the software); (b) One protein/peptide with only amino acid sequence ("Single plain sequence" in the software); (c) One or more peptide sequences, one peptide for each line ("Peptide list" in the software). The data could be input with two different ways. Firstly, the data could be copied to the clipboard and then pasted to the text box in the software. Secondly, the data could be stored as a file and then the user

would find and select the file by clicking the "File open..." button. If the data is provided by both of the above ways, only the data from the file would be used for the calculation.

Parameter setting:

Sometimes, the user would only need to calculate some of the parameters. It could be set in the parameter plate. De-select any parameter would skip the calculation. For the distribution statistic, the ranges and their steps could also be changed in the parameter plate upon the requirement of the research.

Output:

The calculation results could be showed in the right text box of the software, or be exported to the file if there is the file name and its directory in the box following "Save to file:" when running the program.

The output results include the following sections

(a) Result of the physicochemical properties for each protein/peptide. It would list the values of pI, Hy and MW for each protein. (b) Result of the pI statistic; (c) Result of the MW statistic; (d) Result of the Hy statistic; (e) Result of the amino acid composition statistic; in the statistic results, the number and percent of each step of the parameters are provided to get their distribution.

Caveat & future development:

The post-translational modifications are not considered in the calculation, which would be added in the further development. Some other parameters of protein properties would also be considered if needed (for example, the protein stability and so on).

Acknowledgment:

This work was supported by the Chinese National Key Program of Basic Research (Grant No. 2011CB910601, 2010CB912700). We thank Ping Wan, Xiaofang Xue for the help of the software development.

References:

- [1] Ly L & Wasinger VC, *Proteomics*. 2011 **11**: 513 [PMID: 21241016].
- [2] Chan P & Warwicker J, *BMC Biol*. 2009 **7**: 69 [PMID: 19849832].
- [3] Zhang T *et al*. *Comput Biol Chem*. 2006 **30**: 367 [PMID: 16963318].
- [4] Wilkins MR *et al*. *Methods Mol Biol*. 1999 **112**: 531 [PMID: 10027275].
- [5] Bjellqvist B *et al*. *Electrophoresis*. 1993 **14**: 1023 [PMID: 8125050].
- [6] Kyte J & Doolittle RF, *J Mol Biol*. 1982 **157**: 105 [PMID: 7108955].
- [7] http://web.expasy.org/findmod/findmod_masses.html#AA
- [8] Wu S *et al*. *Proteomics*. 2006 **6**: 449 [PMID: 16317776].

Edited by P Kangueane

Citation: Wu *et al*. *Bioinformatics* 8(3): 167-169 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Methodology:

Calculation of isoelectric point

Isoelectric point (pI) was estimated as the pH at which a protein took on a neutral charge (R=0 in the following formula) [5]. For a given sequence, the pI algorithm was as follows. In the formula, N_{aa} represents the number of special amino acids in a special protein, K_{aa} the values of pK_R of special amino acids [8].

1	$R = CR_p - CR_n$	→	(1)
2	$CR_p = \frac{N_{Lys}[H^+]}{[H^+] + K_{Lys}} + \frac{N_{Arg}[H^+]}{[H^+] + K_{Arg}} + \frac{N_{His}[H^+]}{[H^+] + K_{His}} + \frac{[H^+]}{[H^+] + K_{NT}}$	→	(2)
3	$CR_n = \frac{N_{Asp}K_{Asp}}{[H^+] + K_{Asp}} + \frac{N_{Glu}K_{Glu}}{[H^+] + K_{Glu}} + \frac{N_{Cys}K_{Cys}}{[H^+] + K_{Cys}} + \frac{N_{Tyr}K_{Tyr}}{[H^+] + K_{Tyr}} + \frac{K_{CT}}{[H^+] + K_{CT}}$	→	(3)