Pep2Graph: A standalone tool to analyse proteolytic cleavages by proteases from gel-based mass spectrometry data

Sriram Gummadi¹ | Taeyoung Kang¹ | Pamali Fonseka¹ | Sai V. Chitti¹ | Ching-Seng Ang² | Suresh Mathivanan¹

¹Department of Biochemistry, La Trobe Institute for Molecular Sciences, La Trobe University, Melbourne, Victoria, Australia
²The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria, Australia

Correspondence
Suresh Mathivanan, Department of Biochemistry, La Trobe Institute for Molecular Sciences, La Trobe University, Bundoora, Victoria 3086, Australia.
Email: S.Mathivanan@latrobe.edu.au

Funding information
Australian Research Council Future Fellowship, Grant/Award Number: FT180100333

Abstract
Proteases are enzymes that regulate substrates via proteolytic activation and coordinate essential cellular functions including DNA replication, DNA transcription, cell proliferation, differentiation, migration and apoptosis. However, techniques to identify proteolytic events in a high-throughput manner is limited. PROtein TOpography and Migration Analysis Platform (PROTOMAP) is a technique that relies on mass spectrometry-based proteomics to globally identify the shifts in the in-gel migration of proteins and their corresponding fragments that are obtained by proteolysis. However, user-friendly software tool to analyse the proteomic data to identify proteolytic events is needed. Here, we report Pep2Graph, a user-friendly standalone tool that integrates peptide sequence information from in-gel proteomics and presents the data as two-dimensional peptographs with in-gel migration, sequence coverage and MS/MS spectra counts. Pep2Graph (http://www.mathivananlab.org/Pep2Graph) allows users to utilize in-gel proteomics data to study proteolytic events that may play a significant role in normal physiology and pathology.

KEYWORDS
peptides, peptograph, proteases, PROTOMAP

Proteolytic cleavage or proteolysis is the breakdown of proteins into shorter fragments and is catalysed by enzymes called proteases [1–4]. Hence, proteases regulate the stability, localization and activity of numerous proteins [5, 6]. Proteases are ubiquitous and are involved in multiple biological processes including growth, digestion, wound repair, protein catabolism and homeostasis [7, 8]. Additionally, proteases regulate essential cellular functions including proteolytic activation, DNA replication, DNA transcription, cell proliferation, differentiation, migration and apoptosis [9]. Considering the importance of proteases in both health and disease conditions, there has been significant efforts to understand the classes of proteases and the functional activity of the substrate peptides [6, 10]. For instance, in the context of cancer, proteases are implicated in invasion and metastasis, which are the major hallmarks of cancer [10]. Notably, proteases play important role in reshaping the tumour microenvironment [11]. Though high throughput mass spectrometry-based proteomics studies are performed routinely to identify protease substrates [12, 13], user-friendly software tool to analyse such proteomic data to identify proteolytic events are needed.

Currently, PROtein TOpography and Migration Analysis Platform (PROTOMAP) [12, 14], an existing software, is used by various research groups to plot peptographs [12, 13, 15]. This software scripts were written using perl programming language and has been designed to work only with output from DTAselct, a software package to group...
and to interpret the results of proteomics data. The major limitations with the existing software are: (A) the peptographs can only be plotted for experiments with one control and one treatment group, (B) software does not accept any other input format except DTASelect and (C) the tool is not user friendly for biologists to use easily. Hence, we report Pep2Graph, a generic application to the research community across the globe that provides the following advantages including: (A) ability to plot peptograph with multiple sample datasets and to analyse the proteolytic cleavages, (B) provide an easily accessible interface with generic excel template as an input format and (C) user friendly interface rendering the software usable without expert knowledge in proteomics.

Pep2Graph (http://www.mathivananlab.org/Pep2Graph) is a free to use stand-alone tool that is compatible with both Microsoft Windows and Macintosh and is developed using Tkinter, a graphical user interface toolkit, which runs on Python programming language (http://www.python.org) and is created on Spyder, a free and open-source integrated development environment [16]. As shown in Figure 1, Pep2Graph allows users to input data in Microsoft excel template format or in MaxQuant input format (Figure 2). Users can download the zip file containing the excel input template along with the application. Users need to populate the empty excel template with the data and can load the populated excel file into the application.

Similarly, output files from MaxQuant can be fed as the input to the application with no further analysis or changes. Of all the MaxQuant output files, only protein groups (group), peptides (group) and evidence files (separate) are to be fed as the input to the application. After upload, the data is sorted and can be viewed in the data viewer tab and can be downloaded in either excel or csv file format (Figure 3). Before the files are loaded, the users must enter information such as number of experiments (total of control and treatments) and number of gel bands. There are two check boxes available for users to trim data (Figure 4). Show only control checkbox allows users to only consider peptides that are present in control or present in both control and treatment. Show only distinct allows users to consider peptides that are identified in at least one experiment but not in all. This feature allows for detecting peptides that are identified in certain samples alone and hence could be potentially significant proteolytic products. Users can check both checkboxes for data present in control and distinct. Once the data is loaded into the application, there are two options to select the proteins to plot peptograph (Figure 5): (A) to select from input protein list, (B) to manually upload a protein txt file. Selecting from input proteins list would display a selectable drop-down box with a scroll bar with all the unique proteins identified in the input data. Filter option is associated to the drop-down boxes for both options, which would filter out proteins when the user starts typing the protein name in the search box.

The peptograph is plotted using matplotlib library using python programming language. The peptides cleaved during the proteolysis are represented using rectangle patches ranging 0 to the length of the respective peptide. All rectangles in the experimental groups have constant width but with varying length depending on the cleaved peptide length. The rectangles are colour coded depicting their respective experiment which the users can customize. The hover functionality will
FIGURE 2  Snapshot of the Pep2Graph home page when peptide data is submitted as MaxQuant txt file.

FIGURE 3  Snapshot of data viewer tab once the input data is loaded into the application. The filter option is available to select the data for a specific gene of interest and the save option to download data as a excel or a csv file on to the local computer.
FIGURE 4 Pep analysis dashboard with various filters, tables, options to analyse and examine the data. The Plot quick peptograph option can be used to plot a peptograph with filtered data associated with the checkboxes and filters.

FIGURE 5 Schematic representation of how to plot peptographs. Radio buttons can be used to select genes from input or from custom uploaded txt file. Checkboxes to plot peptograph with selected checkbox associated filtered data. Select colours list box to add colours to each experiment. Various options are available to customise the plot like title, quality, y-axis, etc.

help the users to visualise the peptides and their start and end position. In case of multiple peptides overlapping, the hover functionality will show all the peptides involved as a list. The y-axis is shared between the peptograph plot, spectral count plot with different x-axis and the user has the option to update the y-axis with the defined molecular weights and gel bands. The users have multiple options to edit the fonts, colour, colour intensity and the quality of image prior to saving an image compatible for publications.

Overall, Peptograph plots in-gel migration in the vertical dimension (high-to-low molecular weight, top to bottom) and sequence coverage in the horizontal dimension (N to C terminus, left to right). Peptograph visualises peptides from proteolytic cleavage for different experiment
FIGURE 6 Peptographs plotted from the example dataset for different proteins for four experiments (four different colours representing each).

groups, having assigned a unique colour for each experiment group (Figure 6). This helps users to analyse and understand the behaviour of peptides under each experiment group. When the user hovers the mouse on to the rectangular patch (peptide), it shows the sequence of the peptide and the start and end position of the peptide and on the right hand. The spectral count plot shows the magnitude of the proteolytic cleavage for that protein at that specific band in that respective experimental group. The pep analysis tab has three data tables visualised as a dashboard filter section on the top and a button to plot a peptograph for the filtered data. This helps users to analyse proteins undergoing various alterations like any changes in the molecular weight of the protein after proteolytic cleavage in each experimental group.

AUTHOR CONTRIBUTIONS
Suresh Mathivanan conceived and directed the study; Sriram Gummadi developed the application; Sriram Gummadi, Taeyoung Kang, Ching-Seng Ang, Sai V. Chitti, Pamali Fonseka and Suresh Mathivanan drafted and finalized the manuscript with inputs from other authors; Sriram Gummadi, Taeyoung Kang and Suresh Mathivanan prepared the figures.

ACKNOWLEDGEMENTS
Suresh Mathivanan is supported by Australian Research Council Future Fellowship (FT180100333). Pamali Fonseka is supported by the Jack Brockhoff Foundation Early Career Medical Research Grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The raw data that is used in the software are available from the corresponding author upon reasonable request. Pep2Graph software is freely available to download from http://www.mathivananlab.org/Pep2Graph.

ORCID
Suresh Mathivanan https://orcid.org/0000-0002-7290-5795

REFERENCES
1. Duffy, M. J. (1987). Do proteases play a role in cancer invasion and metastasis? European Journal of Cancer & Clinical Oncology, 23(5), 583–589. https://doi.org/10.1016/0277-5379(87)90326-9
2. Lyapina, I., Ivanov, V., & Fesenko, I. (2021). Peptidome: Chaos or inevitability. International Journal of Molecular Sciences, 13128(23), 13128. https://doi.org/10.3390/ijms222313128
3. Martinez-Vendrell, X., & Kikkert, M. (2021). Proteomics approaches for the identification of protease substrates during virus infection. Advances in Virus Research, 109, 135–161. https://doi.org/10.1016/bs.avir.2021.03.003
4. Rogers, L. D., & Overall, C. M. (2013). Proteolytic post-translational modification of proteins: proteomic tools and methodology. Molecu-
5. Canbay, V., & Auf Dem Keller, U. (2021). New strategies to identify protease substrates. Current Opinion in Chemical Biology, 60, 89–96. https://doi.org/10.1016/j.cbpa.2020.09.009

6. López-Otín, C., & Bond, J. S. (2008). Proteases: multifunctional enzymes in life and disease. Journal of Biological Chemistry, 283(45), 30433–30437. https://doi.org/10.1074/jbc.R800035200

7. Liz, M. A., & Sousa, M. M. (2005). Deciphering cryptic proteases. Cellular and Molecular Life Sciences, 62(9), 989–1002. https://doi.org/10.1007/s00018-005-4544-2

8. Turk, B. (2006). Targeting proteases: successes, failures and future prospects. Nature Reviews Drug Discovery, 5(9), 785–799. https://doi.org/10.1038/nrd2092

9. Klingler, D., & Hardt, M. (2012). Profiling protease activities by dynamic proteomics workflows. Proteomics, 12(4–5), 587–596. https://doi.org/10.1002/pmic.201100399

10. Koblinski, J. E., Ahram, M., & Sloane, B. F. (2000). Unraveling the role of proteases in cancer. Clinica Chimica Acta, 291(2), 113–135. https://doi.org/10.1016/s0009-8981(99)00224-7

11. Srinivasan, S., Kryza, T., Batra, J., & Clements, J. (2022). Remodelling of the tumour microenvironment by the kallikrein-related peptidases. Nature Reviews Cancer, 22(4), 223–238. https://doi.org/10.1038/s41568-021-00436-z

12. Dix, M. M., Simon, G. M., & Cravatt, B. F. (2008). Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell, 134(4), 679–691. https://doi.org/10.1016/j.cell.2008.06.038

13. Simon, G. M., Dix, M. M., & Cravatt, B. F. (2009). Comparative assessment of large-scale proteomic studies of apoptotic proteolysis. ACS Chemical Biology, 4(6), 401–408. https://doi.org/10.1021/cb900082q

14. Fuhrman-Luck, R. A., Silva, L. M., Hastie, M. L., Gorman, J. J., & Clements, J. A. (2017). Determining protease substrates within a complex protein background using the PROtein TOpography and Migration Analysis Platform (PROTOMAP). Methods in Molecular Biology, 1574, 145–170. https://doi.org/10.1007/978-1-4939-6850-3_11

15. Niessen, S., Hoover, H., & Gale, A. J. (2011). Proteomic analysis of the coagulation reaction in plasma and whole blood using PROTOMAP. Proteomics, 11(12), 2377–2388. https://doi.org/10.1002/pmic.201000674

16. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95. http://doi.org/10.1109/MCSE.2007.55

How to cite this article: Gummadi, S., Kang, T., Fonseka, P., Chitti, S. V., Ang, C.-S., & Mathivanan, S. (2022). Pep2Graph: A standalone tool to analyse proteolytic cleavages by proteases from gel-based mass spectrometry data. Proteomics, 22, e2200147. https://doi.org/10.1002/pmic.202200147