BIOINFORMATICS APPLICATIONS NOTE

Vol. 21 no. 5 2005, pages 674–675
doi:10.1093/bioinformatics/bti077

Sequence analysis

A simple tool for drawing proteolytic peptide maps

Robert J. Beynon

Protein Function Group, Faculty of Veterinary Science, University of Liverpool, Crown Street,
Liverpool L69 7ZJ, UK

Received on July 6, 2004; revised and accepted on September 27, 2004
Advance Access publication November 11, 2004

ABSTRACT

Summary: I describe a simple standalone program that assists in the preparation of peptide digestion maps. These are useful for comparative studies and for locating peptides within a primary sequence. The program creates an output file as scalable vector graphics that can then be viewed in a web browser or imported into a graphics-editing program.

Availability: The program, as a standalone executable, is available upon request from the author.

Contact: r.beynon@liv.ac.uk

INTRODUCTION

Although the majority of proteomics applications are driven by the need for global comparisons of protein expression, there are other more specific applications where attention can be focused on a limited number of proteins. For example, peptide mass fingerprinting can be used to screen for and identify protein polymorphisms provided that there is adequate coverage of the protein sequence by the proteolytic fragments. The absence of N-terminal or C-terminal fragments can be used to define the approximate sites of proteolysis of a protein fragment. In such studies, it is sometimes necessary to conduct a detailed analysis of a protein and its fragments. Whilst this can be done by comparison of theoretical and observed peptide fragment masses, this can be tedious and error prone. There is scope for simple visualization tools that can aid interpretation and comparison of proteolytic maps. To meet this need, I have written a simple program that is capable of generating peptide maps from protein sequences (Fig. 1). To my knowledge, no such tool exists.

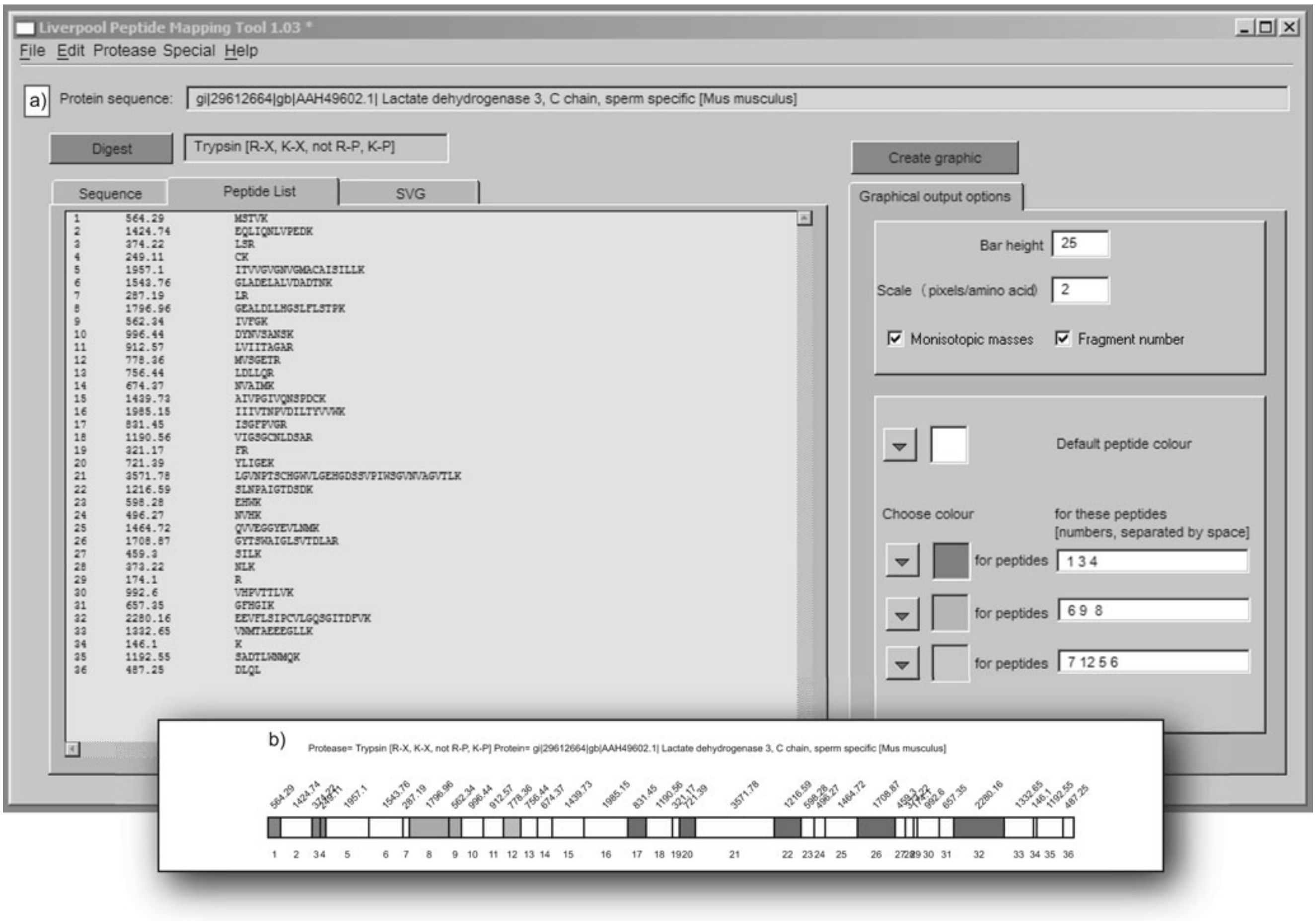
IMPLEMENTATION

The program was written in the software development environment provided by RunTime Revolution (http://www.runrev.com/). This environment is very similar to that provided by a number of other (some defunct) tools, including Hypercard (see, for example, Beynon, 1988), Supercard and Metacard, and is a powerful and rapid application development and prototyping tool. It provides a comprehensive user interface editor that makes laying out the overall programme window(s) a trivial task. Objects, such as pulldown menus, dialogue boxes, text fields and graphics are simple to create. Interaction with each object (for example, a menu, button or a text field) is then mediated through event handlers, defined as scripts associated with that object. The scripting language of Revolution, Transcript, is extremely simple to write and to read, but is sufficiently complex to allow, for example, direct POST operations to a website, use of regular expressions and use of functions and user-defined event handlers. Transcript code is easy to understand and almost reads like pseudocode. Revolution allows rapid testing of applications by interpretation of the code within the programming environment, but the final programme is compiled to include all required libraries to generate a standalone executable.

The peptide map drawing tool does not create a graphic directly. Rather, it generates XML output in the form of scalable vector graphics (SVG). SVG is a W3C recommended standard for vector graphics (http://www.w3.org/Graphics/SVG). As the name implies, SVG files specify the overall structure of the drawing, but not the scale. The image can be rendered on any output device, and at any scale, depending on the requirements of the user. Because SVG files are written in plain text, they can be straightforward to understand and can be manually edited. In its present iteration, the peptide mapping tool generates very simple SVG output. SVG graphics are immediately viewable in several web browsers, provided that a plug-in viewer is installed (http://www.adobe.com/svg/main.html). However, this will generate a bitmap image that cannot be edited. Alternatively, SVG files can be readily imported into programs such as recent iterations of Adobe Illustrator—at present it is not clear that the Microsoft Office applications are able to import SVG although a growing list of conversion tools are becoming available.

The protein sequence may be imported into the program via the clipboard or by a direct download from NCBI, EBI or Expasy. The Protease menu offers the user a choice of proteolytic enzymes (presently trypsin, endopeptidase LysC, endopeptidase ArgC and endopeptidase GluC, but other fragmentations can be added). The program then ‘digests’ the protein sequence and presents the fragments, in sequence order, together with their monoisotopic masses, in the lower left pane of the application window. Once the protein sequence has been introduced, and the appropriate protease chosen, the sequence and peptide list is created through the ‘Digest’ button. At this stage, the user can elect to generate an SVG map without highlighting specific peptides, or can choose to fill the rectangles corresponding to specific peptides with a limited set of colours—these can subsequently be edited in a more complex editor. Further options include the ability to annotate peptides with their monoisotopic masses and/or fragment number.

I have recently argued that proteomics does not just require sophisticated gel image comparison, mass spectral interpretation and protein database searching tools, but that there is a need for simple, end-user tools that perform limited functions (Beynon, 2004). Although



Protein sequence: gi|29612664|gb|AAH49602.1| Lactate dehydrogenase 3, C chain, sperm specific [Mus musculus]

Trypsin [R-X, K-X, not R-P, K-P]

```
1     564.29     MSTVK
2     1424.74    EQLIQNLVPEDK
3     374.22     LSR
4     249.11     CK
5     1957.1     ITVVGVGNVGMACAISILLK
6     1543.76    GLADELALVDADTNK
7     287.19     LR
8     1796.96    GEALDLLHGSLFLSTPK
9     562.34     IVFGK
10    996.44     DYNVSANSK
11    912.57     LVIITAGAR
12    778.36     MVSGETR
13    756.44     LDLLQR
14    674.37     NVAIMK
15    1439.73    AIVPGIVQNSPDCK
16    1985.15    IIIVTNPVDILTYVVWK
17    831.45     ISGFPVGR
18    1190.56    VIGSGCNLDSAR
19    321.17     FR
20    721.39     YLIGEK
21    3571.78    LGVNPTSCHGWVLGEHGDSSVPIWSGVNVAGVTLK
22    1216.59    SLNPAIGTDSDK
23    598.28     EHWK
24    496.27     NVHK
25    1464.72    QVVEGGYEVLNMK
26    1708.87    GYTSWAIGLSVTDLAR
27    459.3      SILK
28    373.22     NLK
29    174.1      R
30    992.6      VHPVTTLVK
31    657.35     GFHGIK
32    2280.16    EEVFLSIPCVLGQSGITDFVK
33    1332.65    VNMTAEEEGLLK
34    146.1      K
35    1192.55    SADTLWNMQK
36    487.25     DLQL
```

Bar height: 25

Scale (pixels/amino acid): 2

☑ Monisotopic masses ☑ Fragment number

Default peptide colour

Choose colour — for these peptides [numbers, separated by space]

▼ ▨ for peptides: 1 3 4

▼ ▨ for peptides: 6 9 8

▼ ▨ for peptides: 7 12 5 6

Protease= Trypsin [R-X, K-X, not R-P, K-P] Protein= gi|29612664|gb|AAH49602.1| Lactate dehydrogenase 3, C chain, sperm specific [Mus musculus]

Fig. 1. The peptide mapping tool and example output. The actual tool (panel a) is in colour, and is a standalone executable. No external libraries are required, other than capabilities embedded in the operating system. To generate the graphic in panel b, a sequence was downloaded from an online database and 'digested' with trypsin. Selected peptides were highlighted (purely for the purpose of illustration), and a peptide map was exported.

this peptide mapping tool is targeted at data presentation, it is representative of the tools that are low key, but useful. There are many other such tools that are needed, and which could, when bound into a common graphical or scripted user interface, create a powerful proteomics workbench in which users would be able to assemble primitive functions into complex, user-specific tools for data analysis. Although this programme does not meet many of those requirements, it might still be of value to anyone who wishes to present peptide mapping data. [The programme, which can be compiled to an executable for Macintosh (OSX and previous OS versions), Windows and Unix distributions (BSD, HP-UX, Iris, Linux, RS/6000, SPARC/Solaris) will be made freely available upon receipt of an email specifying which version is required.]

ACKNOWLEDGEMENTS

Research projects that raised the desirability of this and related tools are supported by BBSRC and NERC.

REFERENCES

Beynon,R.J. (1988) A Macintosh Hypercard stack for calculation of thermo-dynamically-corrected buffer recipes. *Comput. Appl. Biosci.*, **4**, 487–490.

Beynon,R.J. (2004) Enabling proteomics: the need for an extendable 'workbench' for user-configurable solutions. *Comp. Funct. Genomics*, **5**, 52–55.