BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 8 2006, pages 1002–1003
doi:10.1093/bioinformatics/btl052

Sequence analysis

Mass Analysis Peptide Sequence Prediction (MAPSP)

Martin Eisenacher†, Jürgen de Braaf† and Simone König*

Integrated Functional Genomics (IFG), Interdisciplinary Center for Clinical Research (IZKF), Westfalian Wilhelms-University of Muenster, Roentgenstr. 21, D-48149 Muenster, Germany

Received on September 14, 2005; revised on December 15, 2005; accepted on February 8, 2006
Advance Access publication February 24, 2006
Associate Editor: Christos Ouzounis

ABSTRACT

Summary: The software tool MAPSP allows the combinatorial prediction of novel short peptides such as hormones with common sequence features. In addition, it assists in *de novo* sequencing in general. The tool was designed for use in conjunction with the analytical identification method of mass spectrometry (MS) and it can considerably speed-up the analysis of unknowns.

Availability: The web interface is freely available at http://mapsp.ifg.uni-muenster.de/

Contact: koenigs@uni-muenster.de

1 INTRODUCTION

It has been observed that the common features of insect adipokinetic hormones (AKHs) allow a prediction of new sequences with respect to their molecular weights (MW), a fact that is extremely helpful in the identification of new peptides of that family (König 2005). AKH sequences exhibit a length of 8–10 amino acid residues, N-terminal pyroglutamic acid, an amidated C-terminus and tryptophane at position 8. A total of 25 different sequences can be currently found in the NCBI database (National Center for Biotechnology Information, USA; http://www.ncbi.nlm.nih.gov/) and their examination reveals that there is little variation in the amino acid residues present at certain positions in the peptides. Differences in the octapeptides are mainly determined by combinations within residues 2–7 as is shown in Table 1. Therefore, AKH sequences are partially accessible via combinatorial approaches based on known hormone MW, a fact that speeds up peptide identification. To that end, the web interface MAPSP was developed which allows the calculation of sequence combinations. Originally designed for octa- and decapeptides, the program was extended to 25 amino acid residues for additional functionality. In that way, combinations of amino acid residues for short stretches within longer peptides can be calculated when partial sequence and mass (MS/MS) information are available. At present, MAPSP is applicable to all unmodified peptides up to that length as well as most bioactive peptides since terminal pyroglutamic acid and amidation were taken into account. Further modifications can be easily added at user request.

2 IMPLEMENTATION AND DESCRIPTION

2.1 Web interface

The start page allows the user to specify the number of amino acids of the peptide. The program uses this information to bring up a box of choices consisting of the 20 naturally occurring amino acids in the columns and pyroglutamic acid or hydrogen and amide or the free acid for the N- and C-termini, respectively (Fig. 1). A mass search function limits the output to the sequences of interest. Depending on the chosen amino acids the minimal and maximal possible masses are interactively displayed in parentheses. A sorting function presents the sequences in the desired order. The output consisting of columns of MW, the mass of the ion MH^+ and the sequence can be obtained in html or in a text file (comma-separated values, csv) for download, the latter of which is considerably faster. To complete functionality, a button takes the user back to the first input page for a fresh start.

Example AKH prediction. For a novel AKH of measured MW of 915.4 and program input according to Table 1, the abbreviated output list is shown in Table 2. A search of 915.4 ± 1 Da limits the output to Pyr-VNFSPGW-NH_2 presenting a clear hypothesis for MS sequence identification.

Example peptide de novo sequencing. If a peptide has a mass of 1375.68 Da and MS/MS allows assignment of terminal amino acid residues such as QQDFVIxxIEGK then it is helpful to calculate possible residues for the missing stretch. AE, PC or VT mass pairs fill the gap.

2.2 Implementation

The web interface was realized in the script language Perl (version 5.8.5) running on an Apache web server. A hash containing the masses is initialized with the appropriate values. Depending on user input the column structure is dynamically created via the CGI module included in the standard Perl library. The column structure

Table 1. Sequence variations in positions 2–7 (König 2005)

2	3	4	5	6	7
I/L	N	F	T	P	N
V	T	Y	S	S	S
Y				T	G
F				A	D
					W

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

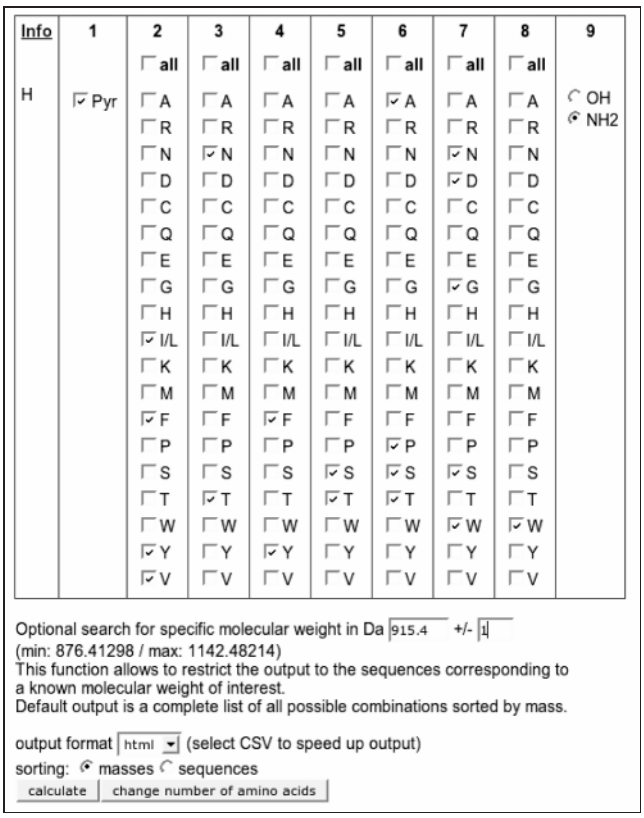

Fig. 1. Input page for choice of possible amino acids in an octapeptide.

is then visualized as a HTML table with check boxes. After clicking the calculate button, the script is called again with the selected parameters. In principle all possible combinations of peptides of the specified format are then created in a complete enumeration. If the user specifies a search limit, the algorithm uses the lower and upper bound in a heuristic approach in order to prune the search tree. This basic branch-and-bound method speeds up the algorithm and can be extended in further developments. Depending on the specified range of masses the algorithm restricts the number of selectable symbols and the maximal number of sequences (at present: 1 000 000 combinations) in order to ensure a reasonable response time. When the output is displayed as a csv file it is stored temporally with a randomized name on the web server to enable multiple concurrent queries. To deal with potentially high access rates the core algorithm may be parallelized by distributing sub-tasks over a local grid network.

Table 2. Output around mass 915.4 following input according to Table 1 showing MW, MH^+ and sequence

MW	MH^+	Sequence
906.4235	907.4314	Pyr-I/LTYSAGW-NH_2
908.4028	909.4106	Pyr-VTYSSGW-NH_2
915.4239	916.4317	Pyr-VNFSPGW-NH_2
916.4443	917.4521	Pyr-VTFTPGW-NH_2
916.4443	917.4521	Pyr-I/LTFSPGW-NH_2
917.4395	918.4473	Pyr-I/LNFTAGW-NH_2

ACKNOWLEDGEMENTS

The authors thank G. Gäde (University of Cape Town, Republic of South Africa) for insights and discussions concerning AKHs. Technical support by Karl Große Vogelsang is gratefully acknowledged.

Conflict of Interest: none declared.

REFERENCES

König,S. (2005) Prediction of insect adipokinetic hormone sequences assists in *de novo* structure elucidation. *Rapid Commun. Mass Sp.*, **19**, 2103–2104.