

NEAPOLIS: web tool for the alignment of signals from mass spectra

Angelo Facchiano

Mass spectrometry is applied for investigations in many field of research, and in the last years it is largely involved in proteomics and derived studies. Instrumentations for mass spectrometry are commonly equipped with computers for the analysis of data, and basic and advanced software for data handling and analysis are distributed with the instruments, so that researchers have the opportunity to use it directly on the experimental output. This represents surely an advantage but, at the same time, it also reflects a lack of flexibility. The different experimental design, in particular for proteomics studies, needs of more opportunities of comparison and integration of data during the analysis of experimental data. Therefore, we developed original and freely available web tools for the alignment and comparison of mass spectra. NEAPOLIS is is freely available at the URL <http://bioinformatica.isa.cnr.it/NEAPOLIS/>

Introduction

Mass spectrometry is an analytical technique which gives a description of a sample in terms of a spectrum of the mass/charge (m/z) values of the molecular species present in the sample. During the last years, applications of mass spectrometry has largely extended due to its role in “omics” studies, and in particular for proteomics. This technique offers high-throughput capabilities and, as all the “omics” sciences, generates a large amount of data for which new analysis strategies are continuously required.

Instrumentations for mass spectrometry are interfaced with computers and software to analyze the spectra are often available in the labs, distributed by companies together with their instrumentation. However, these proprietary software often leave few flexibility to the researcher, which is guided to the analysis and can obtain interesting results but, on the other side, can not explore easily other kind of interpretation.

In particular, in our experience we find that it is often required to compare more spectra. It occurs in the simple case of the analysis of technical or experimental replicates, with the aim of verifying what signals are reproduced. In other cases, samples of two different classes are analyzed, as it can happen for patients affected by a given pathology and health individuals, and the analysis aims to find signals able to discriminate between the two populations. To this aim, we developed novel tools and used them in different studies, some of them already object of publications (1-3).

Tool description

NEAPOLIS is a tool to help the researchers in the comparison of many spectra, obtained as the result of the analysis of replicates as well as of different samples. The web interface is very simple and requires to set few options. It is possible, but not required, to give a name to the job, so that it will be reported in the result pages together with the date, to help the user for future searches in the own archive. Two parameters are required for the analysis. The first is the threshold value for the alignment of signals, and the second is the minimum number of signals to be aligned. These two settings will affect the result of the alignment, and they need a careful choice by the user. The choice of the threshold value should take into account the quality of the signals, which depends on the instrumentation features. The minimum number of signals to be aligned depends on the design of the experiment and on the needs of the user. In case of replicates of the same spectrum, it is expected that significant signals are present in almost all the spectra, so this setting will help to remove sporadic and not reproducible signals. On the contrary, when comparing spectra from different samples, it can be preferred to preserve all signals, so that is can be possible to verify if

they are observed in all or many samples, or just in few of them. It is useful to analyze the data more times by changing the settings of the two parameters, in order to verify how they affect the results, and to find the combination of settings most suitable for the aims of the analysis.

Further settings concerns the request to insert into the output some additional values, as mean values and standard errors of the aligned signals.

The main input is represented by the spectra data, which can be given by a cut-and-paste procedure. An example of the input format is reported on the web site, in summary each spectrum consists in two columns, the first with the m/z value and the second with the signal intensity; the spectrum has as a first line with a “#” symbol, which can be followed by a description of the spectrum. All spectra are listed in sequence and separated by a line, which starts with a double back slash.

NEAPOLIS performs a rapid alignment of all signals, and separates them in groups of values differing by less than the indicated threshold. All signals are sorted in numerical order, and starting from the first, for each value it is evaluated how many other signals are within the fixed threshold range, the mean value of these signals. These signals are considered to be “aligned”, so they will be reported on the same line. The tool applies a selection in case of inclusion of a value in two or more alignments, by preferring the case in which the highest number of signals are aligned together.

The output of the calculation presents the name of the job, if it has been given, the date, the number of spectra analyzed, and two tables. The first table reports only two columns, referred to the mean m/z values and the mean intensities of the aligned signals. The second table expands the first one by reporting also all the m/z values aligned, and all the intensities. For each line, it is reported the number of signals aligned. If required in the input form, the standard errors for m/z and intensities are also reported. Both tables can be easily copied-and-pasted in an Excel sheet for further manipulation or analyses of the data.

Examples of application

NEAPOLIS has been applied in studies object of publications (3-5). In addition, it has been included in a web tool that offer an integrated platform for manage spectra, from the storage to the filtering, averaging and aligning tasks (6).

Future development

Although we developed tools to give simple answers to requests of researchers in proteomics, and therefore we do not complicate the usage by adding further options and opportunity of analyzing the spectra, we are including the tools in more complex procedures upon request of research teams interested to develop an extended, personalized, automated procedure to analyze their data (6). We plan to make made available to the community these new procedures by web interfaces, but also to maintain NEAPOLIS as an independent tool, so that users can simply use it to align their spectra, and perform as separate procedures any further analysis. As an additional function of the tool, we plan to add the optional run as a background job, followed by email to deliver the results. In this way, the user is not required to wait online the output; this option can be very useful in case of large amount of data to be elaborated, although NEAPOLIS seems not dramatically affected for its performance by the amount of data. However, such option should require a registration policy and a restricted access in order to prevent the overcharge by automatic tools on the net.

References

1. Mazzeo MF, Sorrentino A, Gaita M, Cacace G, Di Stasio M, Facchiano A, Comi G, Malorni RA, Siciliano RA: Matrix-assisted laser desorption ionization-time of flight mass

- spectrometry for the discrimination of food-borne microorganisms. *Appl Environ Microbiol* 2006, 72(2):1180-1189. (PMID: 16461665; doi: 10.1128/AEM.72.2.1180-1189.2006)
2. D'Imperio M, Della Corte A, Facchiano A, Di Michele M, Ferrandina G, Donati MB, Rotilio D: Standardized sample preparation phases for a quantitative measurement of plasma peptidome profiling by MALDI-TOF. *J Proteomics* 2010, 73(7):1355-1367. (PMID: 20227539; doi: 10.1016/j.jprot.2010.03.001)
 3. Mangerini R, Romano P, Facchiano A, Damonte G, Muselli M, Rocco M, Boccardo F, Profumo A: The application of atmospheric pressure MALDI to the analysis of long-term cryopreserved serum peptidome. *Analytical Biochemistry* 2011, 417(2):174–181. (PMID: 21756868; doi:10.1016/j.ab.2011.06.021)
 4. Profumo A, Mangerini R, Rubagotti A, Romano P, Damonte G, Guglielmini P, Facchiano A, Ferri F, Ricci F, Rocco M, Boccardo F: Complement C3f serum levels may predict breast cancer risk in women with gross cystic disease of the breast. *Journal of Proteomics* 2013, 85:44–52. (PMID: 23639844; doi: 10.1016/j.jprot.2013.04.029)
 5. Siciliano RA, Mazzeo MF, Spada V, Facchiano A, d'Acierno A, Stocchero M, De Francis P, Colacurci N, Sannolo N, Miraglia N: Rapid peptidomic profiling of peritoneal fluid by MALDI-TOF mass spectrometry for the identification of biomarkers of endometriosis. *Gynecol Endocrinol* 2014, 30(12):872-876. (PMID: 25111755; doi: 10.3109/09513590.2014.943718)
 6. Romano P, Profumo A, Mangerini R, Ferri F, Rocco M, Boccardo F, Facchiano A: Geena, a tool for MS spectra filtering, averaging and aligning. *EMBnet.Journal* 2012, 18(A):125-126. (doi: 10.14806/ej.18.A.454)