

ICPL_ESIQuant – a Powerful Freeware Tool for Handling Proteomics LC-ESI-MS² Experiments

Achim Brunner*, Josef Kellermann and Friedrich Lottspeich

Max-Planck-Institute of Biochemistry, Martinsried, Germany

Abstract

Among the MS-based quantitative methods using stable isotope labelling, the Isotope-Coded Protein Label (ICPL) technique has emerged as a powerful tool to identify and relatively quantify thousands of proteins within complex protein mixtures. The ICPL_ESIQuant 3.0 software package is one of the key components of the ICPL-ESI workflow, covering data processing steps like LC-MS feature detection, ICPL doublet/triplet/quadruplet quantification as well as a merging step of LC-MS features and Mascot search results. As unique features, the software performs isotope pattern overlap corrections and utilizes additional chemical knowledge, e.g. the physico-chemical properties of the ICPL labels, to discard false positive isotope pattern, which significantly improves the quality of the final peptide and protein results.

ICPL_ESIQuant is the first freeware tool on the market, which supports both the shotgun proteomics strategy using Data Dependent Acquisition (DDA) and the directed proteomics strategy using mass inclusion lists for precursor ion selection.

ICPL_ESIQuant 3.0 (32 and 64 bit versions) can be downloaded from <https://sourceforge.net/projects/icplquant/files/>

Introduction

An important task of most MS-based proteomics experiments is to measure a difference in abundance of proteins between two or more biological conditions.

By using the Isotope-Coded Protein Label (ICPL) approach [1,2] up to four different proteomics states can be compared with regard to their relative protein amounts. Measuring the samples using Liquid Chromatography coupled to Mass Spectrometry (LC-MS) results in hundreds and thousands of MS signals, each comprising m/z, intensity and retention time information of a certain compound ion. In order to detect and quantify peptides and proteins, many different software tools have been developed in the recent years, most of them comprising very intelligent peak picking and deisotoping algorithms [3].

Unfortunately, there is no software solution on the market which satisfactorily analyzes MS data from ICPL experiments generated on LC-ESI platforms. This is due to (a) the lack of considering physico-chemical properties of the (ICPL) labels in the search algorithms (to decrease the detection rate of false positive peptide pattern), and (b) missing overlap correction of interfering isotopic pattern, which is indispensable if working with the ICPL triplex or quadruplex approach (ICPL4 and ICPL6 labels differ by only 2 Da) or if dealing with complex protein mixtures like plasma or tissue samples.

Moreover, many software pipelines are simply limited to the DDA strategy, which selects precursor ions from the MS¹ data based on their abundances for further fragmentation. Only identified peptides can be quantified by using this method, while not identified (but eventually regulated) analytes were neglected. The directed proteomics strategy is based on a user-defined peptide mass selection on MS¹ level using inclusion lists, which contain m/z and Retention Time (RT) information of peptides not identified yet in the first MS² analysis [4]. An inclusion list is subjected back to the machine for a subsequent MS² analysis. This application note describes ICPL_ESIQuant, developed to analyze LC-ESI-MS² peptide data generated from ICPL labelled protein samples.

Features

ICPL_ESIQuant reads LC-MS data in .mzXML raw file format [5]. LC-MS data processing is done in three main steps: Peak picking, Deisotoping and Quantification (Figure 1).

To detect peaks in the raw data, the local m/z maxima as well as the lower and upper m/z interval limits for each two-dimensional (2D) peak (comprising m/z and intensity information) are determined in each MS¹ scan. Each detected 2D peak is then represented by its centroid m/z value, calculated as the median of the five most intense data points of a 2D peak (Peak picking step) [6].

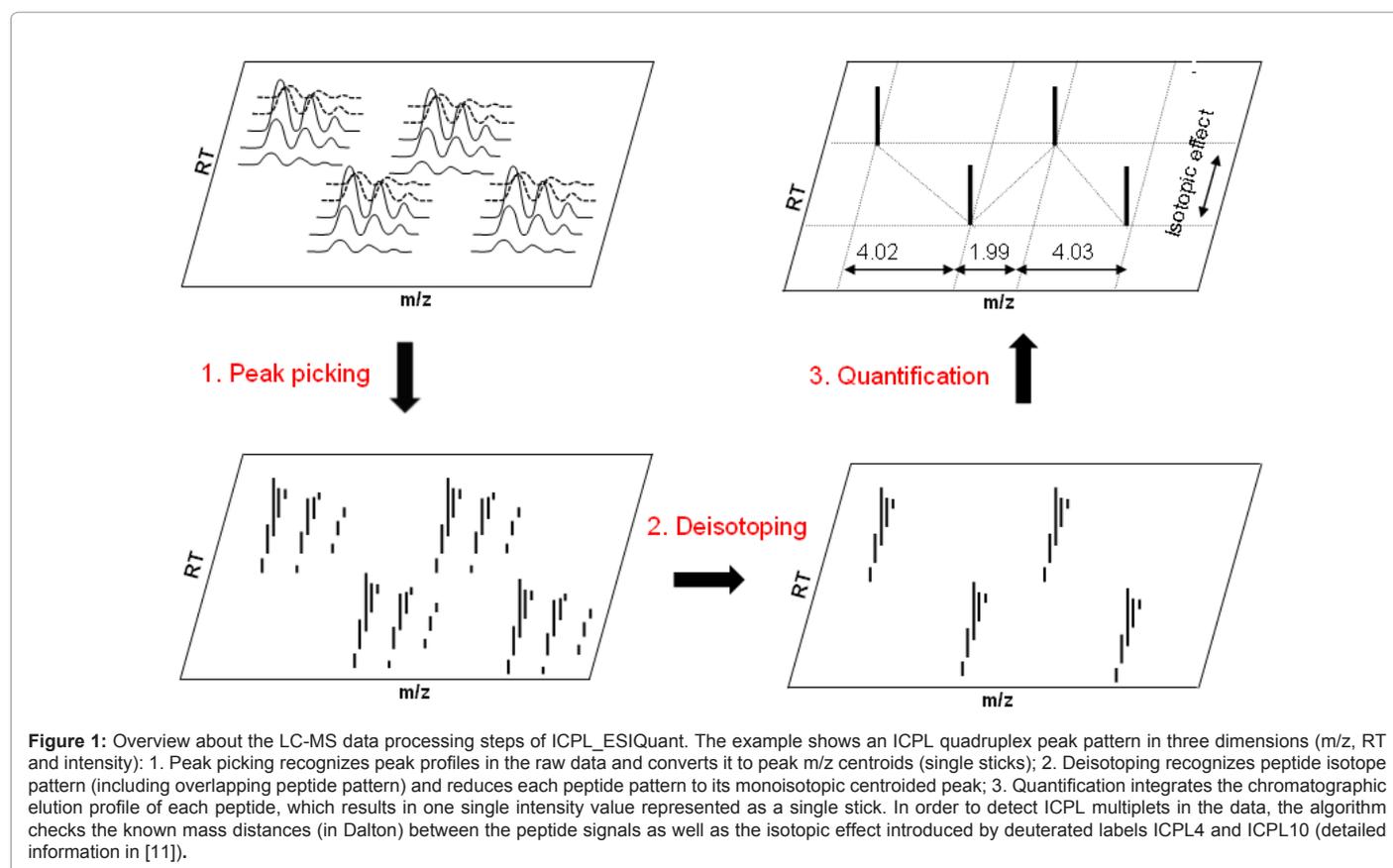
Since every peptide species generates a group of related 2D peaks (=isotope pattern) in a LC-MS dataset, the next algorithmic step is to recognize overlapping and non-overlapping peptide isotope pattern in each MS¹ scan (deisotoping step). This is accomplished by matching the observed peptide patterns with multiple calculated isotope patterns (natural isotope pattern and overlapping variants) generated by using a Poisson distribution [7,8]. All theoretical isotope patterns were estimated by using the formula of an average aminoacid C₁₀H₁₆N₃O₃, which was gained from averaging all aminoacids from the SWISS-PROT protein database. Based on a distance score (Hellinger distance) one can decide which of the multiple theoretical pattern fits best to the observed pattern [9,10] and whether the observed pattern is interfering

*Corresponding author: Achim Brunner, Am Klopferspitz 18, 82152 Martinsried, Germany, Tel: +49(0)89-8578-2465; Fax: +49-(0)89-8578-2802; E-mail: achimbrunner@gmx.de

Received December 03, 2012; Accepted December 20, 2012; Published December 22, 2012

Citation: Brunner A, Kellermann J, Lottspeich F (2012) ICPL_ESIQuant – a Powerful Freeware Tool for Handling Proteomics LC-ESI-MS² Experiments. J Proteomics Bioinform 5: 279-282. doi:10.4172/jpb.1000250

Copyright: © 2012 Brunner A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



with another pattern or not. In case of interference (e.g. in case of an ICPL4 and ICPL6 labelled peptide, differing by 2 Da), the isotopes were corrected by subtracting overlapping peak areas from each other (Supplementary Figure 1).

In addition, this approach increases the number of peptide peak pattern detected in an LC-MS experiment by discovering “hidden isotope pattern”, which arise if isotope pattern of low abundant peptides were (partially) overlaid by patterns of higher abundant peptides having similar physico-chemical LC-MS properties (Supplementary Figure 2).

Once having collected all isotope patterns in all MS¹ scans, ICPL_ESIQuant generates peptide features by combining all identical monoisotopic peptide masses eluting over a continuous and limited period of time. The assembly of features into ICPL multiplets followed by their quantification was done as described in the ICPLQuant manuscript (Quantification step) [11]. It is important to mention, that the slightly earlier LC elution time of peptides marked with deuterated ICPL labels (ICPL4 and ICPL10 isotopologues) is a positive predictor of true positive ICPL triplet or quadruplet peak pattern and therefore can be used for eliminating false positives. The algorithm therefore checks all isotopologues of a ICPL multiplet for their expected retention time behaviour in the MS space (Supplementary material).

ICPL_ESIQuant uses a small software tool called “MzXML2Search.exe” [12] to extract MS² peak lists in Mascot Generic file format (mgf). These files can be directly loaded into the Mascot search engine [13].

ICPL_ESIQuant imports Mascot result files (.dat files) and the intermediate results generated during the quantification step to

assemble peptide and protein result lists in .txt file format, which are easily editable in Microsoft Excel/ Access or Open Office applications. These files can also be loaded into visualization tools like Spotfire or into statistical programs like Matlab to do further bioinformatics calculations (Figure 2).

The user can decide whether to stop the analysis workflow at this point (DDA) or to use automatically generated mass inclusion lists of not yet identified (but quantified) ICPL multiplets for a further MS² identification. The last option can be used iteratively in order to maximize peptide and protein identification rate.

Conclusion

If working with the ICPL technology, ICPL_ESIQuant is the first choice to quantitatively analyse peptides and proteins measured on LC-ESI mass spectrometers for the following reasons:

- (1) There is no freeware on the market, that is able to analyze multiplexed MS data generated from ICPL experiments;
- (2) By using .mzXML as file input format, the software is compatible with the most common MS instruments;
- (3) ICPL_ESIQuant considers overlapping isotope pattern, which is essential for an accurate quantification of ICPL triplex and quadruplex labelled samples and the handling of complex MS data in general;
- (4) ICPL_ESIQuant takes advantage of the physico-chemical properties of the ICPL labels in its algorithm, significantly improving the true positive rate (and hence decreasing the false positive rate!) of detected ICPL multiplet pattern;
- (5) ICPL_ESIQuant supports both the shotgun proteomics strategy using DDA and the use of mass inclusion lists, enabling the user to preselect

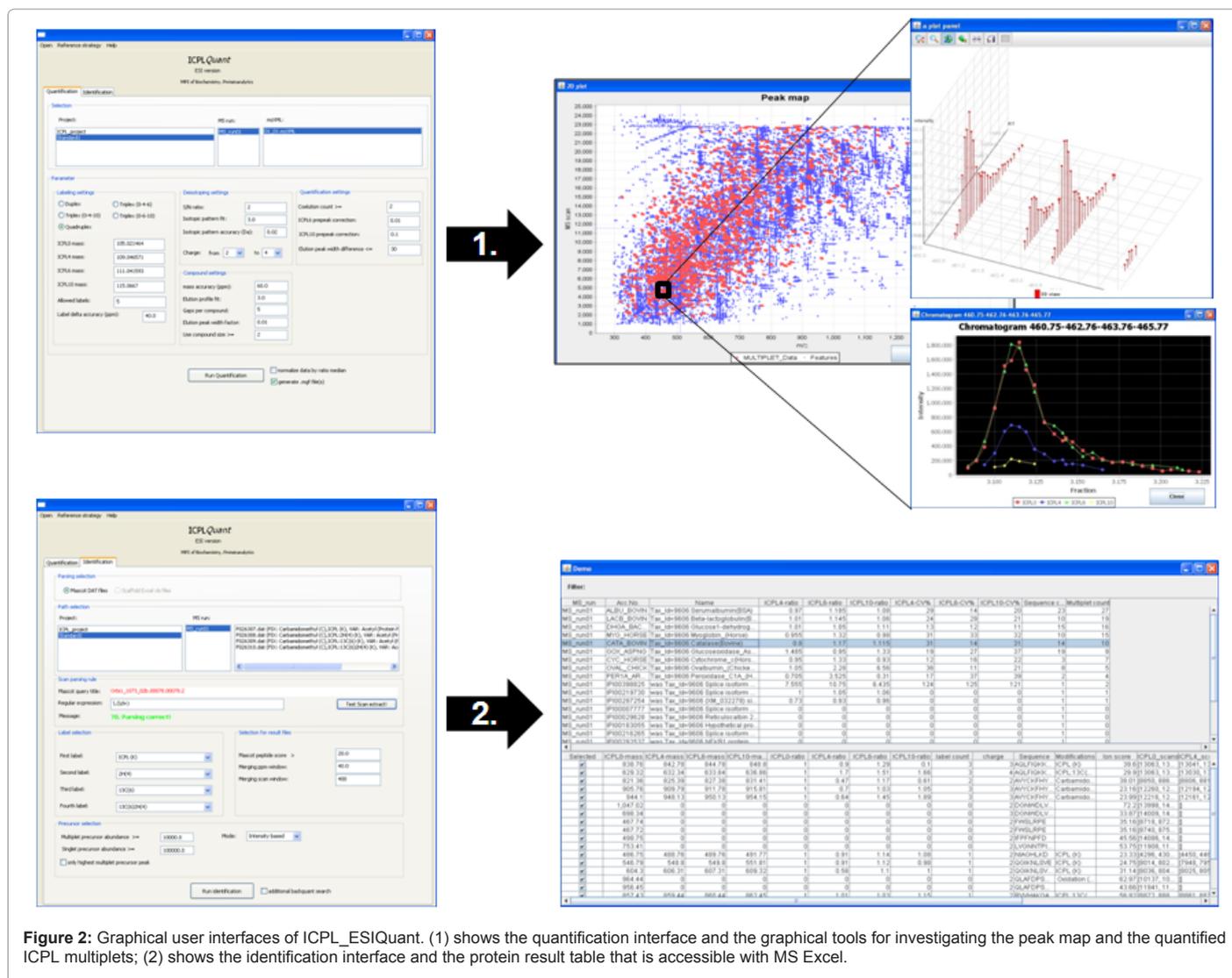


Figure 2: Graphical user interfaces of ICPL_ESIQuant. (1) shows the quantification interface and the graphical tools for investigating the peak map and the quantified ICPL multiplets; (2) shows the identification interface and the protein result table that is accessible with MS Excel.

biological interesting peptides for further MS² identifications.

ICPL_ESIQuant has already been successfully applied to proteomics experiments in the field of cancer biomarker discovery and anti-diabetic hemorphins [3,14,15].

Acknowledgements

This project was funded by the BMBF (German Federal Ministry for Education and Research) Grants FKZ 031U101 and FKZ 031U201.

The authors have declared no conflict of interest.

References

- Kellermann J, Lottspeich F (2012) Isotope-Coded Protein Label. *Methods Mol Biol* 893: 143-153.
- Schmidt A, Kellermann J, Lottspeich F (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5: 4-15.
- Lemeur S, Hahne H, Pachel F, Kuster B (2012) Software tools for MS-based quantitative proteomics: a brief overview. *Methods Mol Biol* 893: 489-499.
- Domon B, Aebersold R (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* 28: 710-721.
- Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, et al. (2004) A

common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 22: 1459-1466.

- Lehmann WD (1995) *Mass Spectrometry in Biochemistry*. ISBN 3-8602-5094-9.
- Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, et al. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* 22: 1902-1909.
- Valkenburg D, Assam P, Thomas G, Krols L, Kas K, et al. (2007) Using a Poisson approximation to predict the isotopic distribution of sulphur-containing peptides in a peptide-centric proteomic approach. *Rapid Commun Mass Spectrom* 21: 3387-3391.
- Eslinger PW, Woodward WA (1991) Minimum Hellinger distance estimation for normal models. *Journal of Statistical Computation and Simulation* 39: 95-113.
- Karlis D, Xekalaki E (2003) Choosing initial values for the EM algorithm for finite Poisson mixtures. *Computational Statistics & Data Analysis* 41: 577-590.
- Brunner A, Keidel EM, Dosch D, Kellermann J, Lottspeich F (2010) ICPLQuant - A software for non-isobaric isotopic labeling proteomics. *Proteomics* 10: 315-326.
- Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* 1: 2005.0017.

13. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
14. Barkhudaryan NH, Sarukhanyan FP, Kellermann J, Lottspeich F (2011) Calcineurin and μ -opioid receptors are involved in the molecular mechanisms of anti-diabetic effect of LVVYPW. *Medical science in Armenia* 51: 33-42.
15. Lottspeich F, Kellermann J, Keidel EM (2010) Molecular biology tools: Proteomics techniques in biomarker discovery. *Scand J Clin Lab Invest Suppl* 242: 19-22.