Journal of Chromatography B 1083 (2018) 258–270

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/jchromb

Manual method of visually identifying candidate signals for a targeted peptide

Aleksey Filimonov*, Arthur Kopylov, Andrey Lisitsa, Alexander Archakov

Orekhovich Institute of Biomedical Chemistry of Russian Academy of Sciences, Bld. 8, 10 Pogodinskaya Str., 119121 Moscow, Russian Federation

ARTICLE INFO

Keywords:
Targeted proteomics
Protein detection
Proteomic analysis
Signal identification
Peptide detection
Chromatogram analysis

ABSTRACT

The purpose of this study is to improve peptide signal identification in groups of extracted ion chromatograms (XICs) obtained with the liquid chromatography-selected reaction monitoring (LC-SRM) technique and a triple quadrupole mass spectrometer (QqQ) operating in one of the supported multiple reaction monitoring (MRM) modes. The imperfection of quadrupole mass analyzers causes ion interference, which impedes the identification of peptide signals as chromatographic peak groups in relevant retention time intervals. To investigate this problem in depth, the QqQ conversion of the eluate into XIC groups was considered as the consecutive transformations of the particles' abundances as the corresponding functions of retention time. In this study, the hypothesis that, during this conversion, the same chromatographic profile should be preserved as an implicit sign in each chromatographic peak of the signal was confirmed for peptides.

To examine chromatographic profiles, continuous transformations of XIC groups were derived and implemented in srm2prot Express software (s2pe, http://msr.ibmc.msk.ru/s2pe). Because of ion interference, several peptide-like signals may appear in one XIC group. Therefore, these signals must be considered candidates for a targeted peptide's signal and should be resolved after identification. The theoretical investigation of intensity functions as XICs that are not distorted by noise produced three rules for Identifying Candidate Signals for a targeted Peptide (ICSP, http://msr.ibmc.msk.ru/ICSP) that constitute the proposed manual visual method. We theoretically and experimentally compared this method with the conventional semiempirical intuitive technique and found that the former significantly streamlines peptide signal identification and avoids typical errors.

1. Introduction

The final stage of peptide detection involves the interpretation of data acquired via liquid chromatography-selected reacting monitoring (LC-SRM) accomplished on a triple quadrupole mass spectrometer (QqQ [1,2]) operating in one of the available multiple reaction monitoring (MRM) modes. This stage should give a list of the detected peptides (and their quantities when possible) among the targeted peptides for a SRM survey. Whichever way this list is obtained, it will be necessary to identify the peptide signal. However in LC-SRM, the peptide signal is represented by a XIC group in its retention time interval, i.e., a group of chromatographic peaks. Therefore, it will be necessary to establish that the same compound contributes to them.

Numerous computational methods [3,4] for data interpretation are proposed to analysts. And analysts are needed for the end ready-for-use tools such as Skyline [5], mProphet [6], PASSEL [7], and AuDIT [8] and the general-purpose software supplied with QqQ instruments, such as Agilent's MassHunter Qualitative Analysis and Thermo Scientific's Xcalibur. However, such programs do not provide most of those methods. The peculiarities of operating these methods are difficult to learn, verify and use. Moreover, the daily use of these methods requires their deep understanding for combination and specific settings, or numerous and non-trivial choices for each experiment or peptide signal.

For these reasons, unlike solving the peptide identification task by delegating it to such software tools as MASCOT [9], Sequest [10], or X!Tandem [11], the peptide-detection task continues to be generally solved by manually analyzing chromatograms on a screen and “by

Abbreviations: ICSP, Identifying Candidate Signals for a targeted Peptide; IT, Information Technology; LC-SRM, liquid chromatography-selected reaction monitoring; m2p, **ms2p**rot, our in-house software system; MRM, multiple reaction monitoring; OOA, object-oriented analysis; QqQ, triple quadrupole mass spectrometer; s2pe, **srm2p**rot **E**xpress software; SI, Supporting Information; XIC, extracted ion chromatograms

* Corresponding author.
E-mail address: filimonov.a.d@gmail.com (A. Filimonov).

https://doi.org/10.1016/j.jchromb.2018.01.017
Received 2 May 2017; Received in revised form 5 January 2018; Accepted 15 January 2018
Available online 31 January 2018
1570-0232/

eye". At first glance, it seems[1] that analysts know [12] how to visually locate chromatographic peak groups of peptides and how to compare them in terms of their elution time, mutual locations, shapes and profiles of relative intensities. However, the analyst affects chromatogram interpretation significantly, which is considered in detail further.

The objective of this work is to make peptide signals easily observable and verifiable in a XIC group. Such a method should allow checking the correctness of the programs implementing peptide detection algorithms. Therefore, it should not introduce any additional information from a priori assumptions or obtained from the analyst. This means that the method must be applied within one XIC group. Nevertheless, the method can apply transparent, continuous and preserving chromatogram transformations. Such transformations should not distort the information originally contained in acquired data. These requirements allow the analyst to be sure that the observed signal of a peptide is not an artifact resulting from the transformation of chromatograms.

2. Materials and methods

In this study, we used our SRM data series **ds0** [13] deposited in PASSEL [7]. The data series ds0, which consists of the datasets **bio0** and **QC0**, was acquired with an Agilent Triple Quad 6410 QqQ for 25 targeted peptides of 21 proteins encoded by human chromosome 18. The dataset bio0 was acquired from blood plasma that had been depleted of major proteins (MARS Hu-14 system), and QC0 was acquired from a mixture of synthetic peptide analogs, i.e., external standards, in a matrix of 8 non-human proteins with a deliberate overloading of the column capacity. In our previous study [13], a qualitative analysis of ds0 (this process is referred to as **"MH + ms2prot"**) was performed[2] by the analyst using the Agilent MassHunter Qualitative Analysis software and was validated (see the chromatograms and peptide-detection tables [14]) with **ms2p**rot, our in-house software system[3] (m2p, derived from mQuest/mProphet v1.0.3.5 software [6]), and PASSEL [7], which is based on the original version of mQuest/mProphet software [6]. The m2p reports [14] include a catalog of the signals of detected peptides (referred to as a set of signals "**j**") and the chromatograms that were interpreted in "MH + ms2prot".

3. Theory

3.1. Applying the stream concept to describe crucial objects analyzed via LC-SRM

From an informatics perspective[4], LC maps each eluted molecule onto the retention time axis. These molecules can then be logically grouped by the respective compound identifiers into series arranged on the compound axis. To consistently consider and understand such series and describe and analyze such objects[5], which properties are measured using the LC-SRM technique, we applied the Information Technology (IT) concept of stream, as shown in Fig. 1.

Other types of particles can also be logically grouped into series. Thus, the stream concept can be applied to particles such as eluted peptides or other compounds, their precursor or fragment ions, and electrons. This concept is used because it implies the following informational aspects, unlike the concepts of beam, current, flow, flux and fluid.

1. **Structure**: The **stream** is a collection of discrete ordered items of concrete types. Such items can be, e.g., molecules of a peptide, ions as charged molecules, bytes or video frames. For an eluted peptide, the order of items is defined by the retention times of its eluted molecules. The stream concept does not limit how items can be distinguished. Thus, the stream is just a structuring, ordering and grouping concept. Hence, by definition, the stream does not depend on the aggregate states of the particles. This inference is important for understanding the ion stream as an ion beam inside of the MS.
2. **Definite borders**: One item is at the beginning of the stream, and another is at its end. Thus, the length and lifetime of the stream of an eluted peptide correspond to the duration of its elution time.
3. **Continuity**: In the stream, one item follows another and preserves its order and interval, which may change locally in some special cases during movement or transfer.
4. **Density**: Density is the number of items per unit distance between items (i.e., the focus is on the stream itself). The density may vary from the beginning of the stream to its end.
5. **Intensity**: Intensity is the number of items in a fixed time interval. For eluted molecules, the distances between these molecules map into corresponding time intervals in the retention time axis depending on the flow rate. Because of the changing density, the intensity also varies, forming shapes. Thus, the stream's **intensity function** can be characterized as the quantities of items in equal-length time segments.
6. **Profile**: The profile is a distribution function of the number of items along the length of the stream from its beginning to its end. By definition, this function has unit area. Thus, the *intensity function* can be represented as the number of items of the *stream* multiplied by its profile. For an eluted peptide, the **chromatographic profile** is the profile of that eluted peptide's stream.
7. **Parallelism**: This concept does not limit how many *streams* can be present simultaneously in one space. Parallelism is shown in Fig. 1: The *stream* of one peptide can be partially or fully eluted concurrently with another one and move jointly in the same capillary.

These characteristics lead to the following **consequences resulting from the *stream* concept**.

1. **Properties of substreams.** When the item of the *stream* can be logically divided into several parts, i.e., when that item has **subitems**, the *stream* has **substreams** of each *subitem*. Hence, each *substream* has the same properties as the *stream* itself because the number of instances of each *subitem* and the intervals between them are the same as those of the whole items.
2. **Properties of the derived stream.** When the item of a *stream* can be practically split into several parts, the *subitem* can become a **derived item**. The *stream* of a *derived item* is referred to as a **derived stream**. However, in practice, items can be split into *derived items* in several ways. If more than one such way acts jointly but sequentially, *derived streams* become selections from the corresponding *substreams*. Hence, *derived streams* preserve the *profile* of the source *stream*, but their *intensity functions* are relatively smaller than the *intensity functions* of the corresponding *substreams*, i.e., the source *stream*.
3. **Sampling.** It is practically impossible to register each instance of the

[1] For the analysis of the software suitability for the peptide-detection task, see Section S-1[18].

[2] See protocols, datasets, reports and details on the measurement structure and statistics [14].

[3] For details on the methods and features of the ms2prot system compared to mProphet v1.0.3.5 [6] and, hence, SPECTRODIVE, see Section S-1.3.

[4] Chromatogram interpretation consists of the extraction and transformation of information via signal processing[6] and the examination and interpretation of this information, which is conducted by the analyst. Therefore, informatics and its disciplines, particularly computer science and IT, as fields that involve the study, design and implementation of such processes, are sources of appropriate methods and techniques for analyzing and representing information phenomena, aspects, objects, relationships and operations for solving the problem of peptide detection in chromatogram interpretation. This approach takes the MS domain "as is" and, preserving every original sense, produces and emphasizes all of the aspects, features and properties needed for further understanding and development without changing the well-established background.

[5] IUPAC does not define any special term for such objects [2,16,19].