Poster Note 64404

The Sliding Window Algorithm for the Analysis of LC/MS Intact Protein Data

Gazis, P. R., Horn, D. M.
Thermo Fisher Scientific, San Jose, CA, USA

Overview

Purpose: Identification of large molecules such as intact proteins in LC/MS is complicated by the difficulty of identifying the relevant peaks in the chromatography. In general, peaks associated with large molecules will have complicated profiles, ill-defined start and stop times, and often overlap with other components in complex protein samples. In this study, we discuss a novel "sliding window" approach that eliminates the need to identify chromatographic peaks and takes advantage of power of deconvolution algorithm to identify components directly.

Methods: The Sliding Window Algorithm averages spectra over a succession of windows in retention time, deconvolves each average spectrum, then merges similar masses from consecutive deconvolutions to identify components. This new algorithm has several advantages over conventional approaches: 1) it avoids the problems involved in trying to identify peaks associated with large molecules, 2) it can identify and characterize components that coelute at overlapping retention time ranges, 3) it produces a meaningful elution profile for each components it identifies, 4) it can reduce the rate of false positives, 5) in many cases it can also increase sensitivity.

Results: We apply the Sliding Window Algorithm to three representative data sets: a protein mixture, an antibody data set, and an ADC data set. The algorithm identifies components and their associated elution profiles.

Introduction

The Sliding Window Scheme is an alternative to the conventional approach to identifying components in LC/MS data. Rather than try to identify chromatographic peaks, then detect components associated with those peaks, it averages spectra over a succession of sliding windows in retention time, deconvolves each averaged spectrum, and merges similar masses to identify components. This algorithm is incorporated into Thermo Scientific™ Protein Deconvolution™ 4.0 Software.

Methods

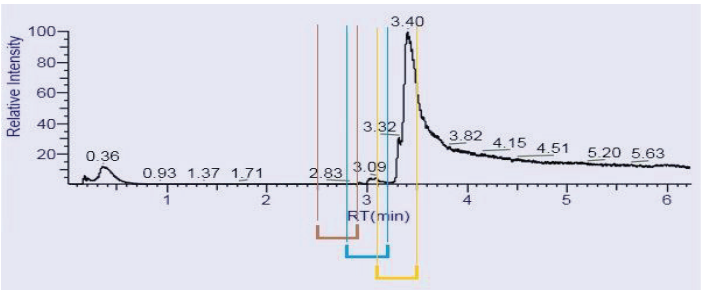
General Approach

The Sliding Window Algorithm involves two steps: the sliding window step and the mass merge step. These are described below.

The Sliding Window Step

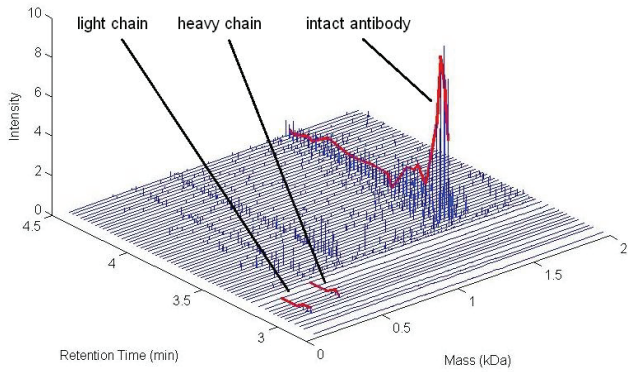
The sliding window step applies a sliding window in retention time to generate a succession of time-averaged spectra. The sequence of sliding windows is determined by four parameters: start time, stop time, width, and offset. This is illustrated in Figure 1, which shows a succession of three retention time windows starting at 2.5 min, with a width of 0.4 min and an offset of 75%, stopping at 3.5 min.

FIGURE 1. A succession of three windows in retention time, used to generate time-averaged spectra



Each average spectrum is then deconvoluted using the appropriate deconvolution scheme – Xtract for isotopically resolved and ReSpect™ for isotopically unresolved spectra – to generate a list of 'component peaks' for the successive retention time windows. The sliding window step produces a list of 'component peaks' for individual retention times. Figure 2 shows a plot of component intensities vs mass and retention time for a typical succession of sliding windows

FIGURE 2. component intensities vs mass and retention time for the example in Figure 1.



Valid components, such as the light chain, heavy chain, and intact antibody, occur at similar mass values, causing them to 'line-up' in the plot. Noise signals appear as randomly distributed peaks with no significant correlation in mass or retention time..

The Mass Merge Step

The mass merge step applies a sliding window in mass to the component peaks produced by the sliding window step and merges them to produce a list of 'merged components'. This sliding window in mass can be implemented as a window of constant width or as a minimum separation between components. The first approach will produce a list of components with a fixed width in mass, some of which could plausibly be merged. This approach is used for results from isotopically resolved spectra, for which masses are extremely well-determined The second approach produce a list of components that are separated by more than some minimum difference in mass. This will guarantee that related components peaks are merged, but can also incorporate unfortunately-placed noise peaks (see the discussion of Figure 6). This approach second is used for isotopically unresolved data.

The Sliding Window Algorithm returns a list of components and an 'abundance trace' for each component. Unlike XICs, which can incorporate unrelated parts of the original signal that might happen to share some m/z values with the primary component, the abundance trace is the actual elution profile associated with that component. This is illustrated in the following examples.

Thermo SCIENTIFIC

Results

The Sliding window algorithm was applied to three representative data sets – a protein mixture, the antibody data set shown in Figures 1 and 2, and ADC data -- to evaluate its effectiveness. The results are discussed below.

Protein Mixture Data

The protein mixture consisted of 9 proteins. Some of these involved isotopically resolved spectra and will be ignored for the purposes of this particular.. Seven component groups associated with isotopically unresolved spectra eluted between a retention time of 10 and 17 minutes. These are listed in Table 1.

Table 1. 7 Component groups from the 9 protein mixture that eluted between a retention time of 10 and 17 minutes

ID	Mass Range (Da)	Retention Time Range
1	14,300-14,300	10.3-10.6 min
2	18,300-18,700	13.1-13.4 min
3	19,900-20,100	12.5-12.7 min
4	28,900-29,100	14.6-14.9 min
5	36,100-36,200	13.9-14.2 min
6	66,400-66,600	11.7-12.1 min
7	79,200-80.200	10.9-11.2 min

The Sliding Window Algorithm was applied in conjunction with ReSpect™ to the data for this retention time range. Figure 3 shows the resulting deconvoluted spectrum. The algorithm identified all 7 component groups in Table 1. Each component group is associated with a well-defined cluster of peaks in Figure 3.

FIGURE 3. Deconvoluted spectrum generated by the Sliding Window Algorithm for a protein mixture.

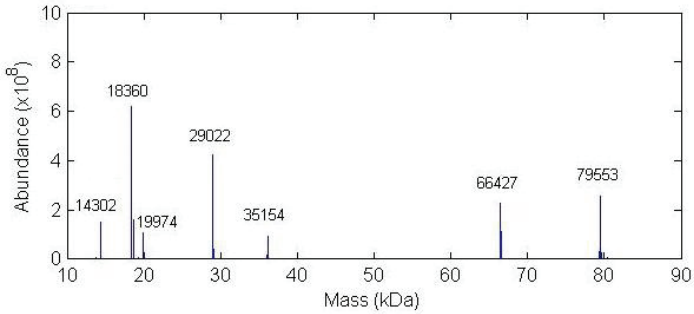
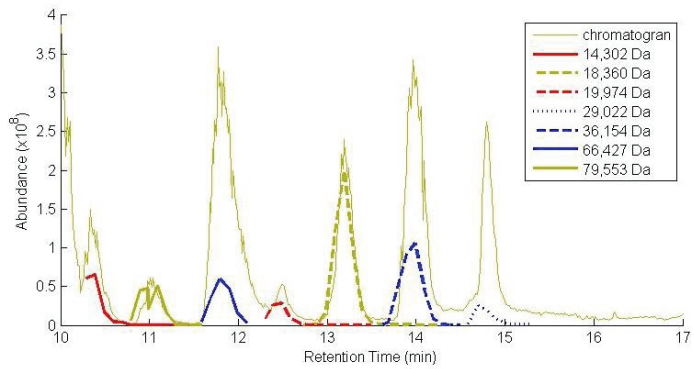


Figure 4 shows the original chromatogram and elution profiles generated by the Sliding Window algorithm for the most abundant modification for the components at 14,302, 18,360, 19,974, 29,022 , 36,154, 66,427, and 79,553 Da.

FIGURE 4. Chromatogram and elution profiles of 7 components identified by the Sliding Window Algorithm for a protein mixture



The algorithm identified retention times and generated elution profiles even for time periods such as the one between of 10.7 and 11 min when two components coeluted.

Antibody data

The antibody data set consisted of an intact antibody and associated modifications in the vicinity of 151 kDa along with component groups associated with a light chain, a heavy chain, and a combination of light and heavy chains. These are listed in Table 2

Table 2. 4 Component groups in the antibody sample

ID	Mass Range (Da)	Retention Time Range	Comments
1	23,400-25,200	2.9-3.2 min	Light chain
2	47,100-47,400	3.0-3.1 min	Heavy chain
3	135,000-135,400	3.2-3.3 min	Light + 2 Heavy
4	150,100-152,800	3.2-6.2 min	Intact antibody

Figure 5 shows the deconvoluted spectrum generated by the Sliding Window Algorithm in conjunction with ReSpect™. Due to the more challenging nature of this data set, it is not as clean as Figure 3, and it includes some false positives that the current version of the algorithm was unable to exclude, but it shows well-defined clusters of peaks corresponding to all four of the component groups listed in Table 2.

FIGURE 5. Deconvoluted spectrum generated by the Sliding Window Algorithm for an antibody sample.

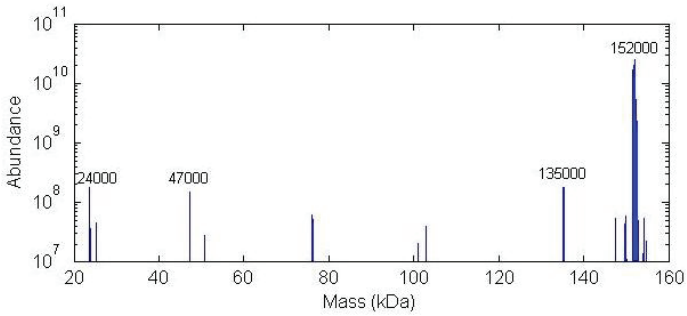
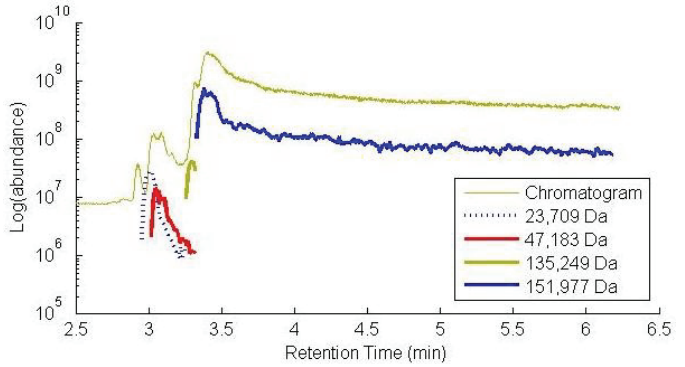


Figure 6 shows the original chromatogram and elution profiles generated by the Sliding Window for the most abundant modification of each of the 4 component groups listed in Table 2: at 27,709, 47,183, 135,249, and 151,977 Da. Like the deconvoluted spectrum, these profiles are not as clean as the results for the protein mixture in Figure 4, but they successfully resolve the two coeluting components at 23,709 and 47,183. They also resolve the component associated with the shoulder in the chromatogram at 3.2 min that can be difficult to detect using conventional techniques

FIGURE 6. Chromatogram and elution profiles of 4 components identified by the Sliding Window Algorithm for IgG_source_cid



ADC Data

The ADC data consisted of the antibody plus 8 components groups associated with ADCs. These are listed in Table 3.

Table 3. 9 Component groups in the ADC sample

ID	Mass (Da)	Retention Time Range
0	145,168	8.45-8.60 min
1	146,125	8.45-8.65 min
2	147.082	8.55-8.75 min
3	148,038	8.55-8.80 min
4	148,996	8.55-8.80 min
5	149,953	8.55-8.60 min
6	150,911	8.55-8.80 min
7	151,865	8.65-8.85 min
8	152,826	8.70-8.90 min

The top panel of Figure 7 shows the deconvoluted spectrum generated by the Sliding Window Algorithm for this data set. The bottom panel shows results from a conventional deconvolution for the same time range. The Sliding Window Algorithm performed significantly better, identifying ADCs at 151,866 and 151,827 Da that the conventional failed to detect.

FIGURE 7. Comparison between deconvoluted spectra generated by the Sliding Window Algorithm and conventional methods for an antibody sample.

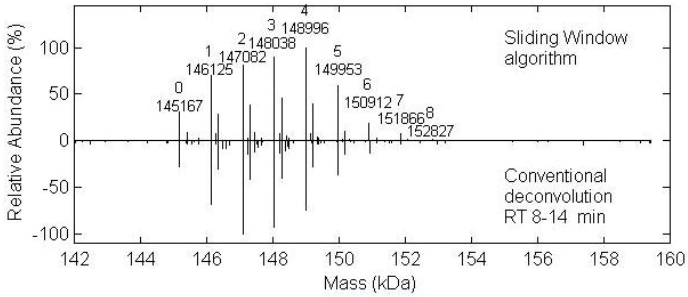
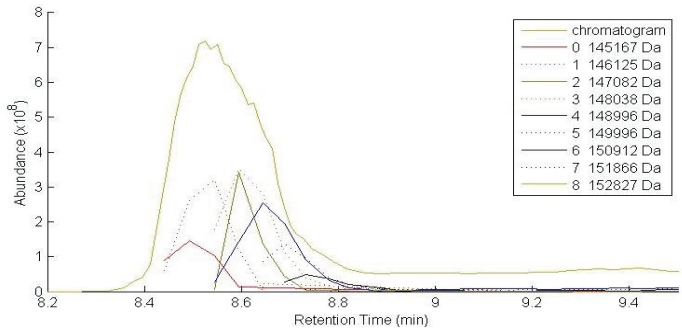


Figure 8 shows the original chromatogram and elution profiles generated by the Sliding Window Algorithm. The algorithm is able to distinguish between the 9 different components and generate unique profiles in a way that would be difficult using conventional techniques.

FIGURE 8. Chromatogram and elution profiles of 9 components identified by the Sliding Window Algorithm for an ADC sample



Conclusions

The Sliding Window Algorithm is a novel approach to identifying components in LC/MS data. This new algorithm has several advantages over the conventional approach:

- it avoids all the problems involved in trying to identify the complicated and ill-defined chromatographic peaks associated with large molecules,
- it can identify and characterize components that coelute at overlapping retention time ranges,
- it produces a meaningful elution profile for each components it identifies,
- it can dramatically reduce the rate of false positives.
- In many cases, it can increase sensitivity.

The Sliding Window Algorithm has been incorporated into Protein Deconvolution 4.0 software. It can be used to significantly improve the results of a charge state deconvolution. In the future, we will explore enhancements to this algorithm, such as improved noise peak rejection and the addition of a fitness measure to reduce incidence of false positives.

Certified System ISO 13485
Thermo Fisher Scientific, San Jose, CA USA is ISO 13485 Certified.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

PN64404-EN 0615S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand