

# A Computer Algorithm for the Identification of Protein Interactions from the Spectra of Masses (PRISM)

Joshua W. K. Ho, Bethny Morrissey, and Kevin M. Downard

School of Molecular and Microbial Biosciences, The University of Sydney, Sydney, Australia

A new algorithm is reported to assist with the identification of protein interaction domains by comparing pairs of MALDI mass spectra recorded for protein digests treated with a binding partner versus an untreated control. Known as PRISM, for protein interactions from the spectra of masses, the algorithm imports  $m/z$  versus peak area data directly from a pair of MALDI mass spectra recorded for the control and reaction sample. The algorithm is shown to be able to successfully identify antigenic determinants for protein antigens within mixed protein digests. The algorithm has general utility for the comparative analysis of differences within any two mass spectra of any type and is easily implemented using a simple, intuitive graphical user interface (GUI). (J Am Soc Mass Spectrom 2007, 18, 563–566) © 2007 American Society for Mass Spectrometry

We have previously reported on the use of mass spectrometry for the identification of protein interactions and interaction domains in which solution-specific protein complexes were preserved on [1, 2] and even ionized from [3] matrix-assisted laser desorption ionisation (MALDI) sample targets. Importantly, it has been shown [1–3] that this can be achieved without the affinity capture [4, 5], immobilisation [6], or absorption of either binding partner on such surfaces [7].

The method has been applied to identify the interaction interface of antigen-antibody complexes [1] and has been extended to the antigenic surveillance of the influenza virus, on both whole virus [2] and gel-resolved viral antigens [8, 9] as part of a proteomics-based strategy. Particular advantages of the approach include the ability to directly assess both the structure (by means of mass mapping) and antigenicity of the viral antigens in a single step. The approach offers the opportunity to rapidly identify multiple determinants simultaneously from a single sample or single MALDI target [8, 9] at low sample levels and can be extended to the analysis of any protein interactions that can withstand the deposition and analysis process.

Protein interaction interfaces are identified through a comparison of the MALDI mass spectra of the digested proteins within an unreacted control sample to the same mixture treated with a cross-reacting antibody or other binding partner [1, 2, 8, 9]. Ion signals for peptides that represent binding domains are shown to be selectively suppressed with a reduction in their relative intensity in the spectrum of the reaction mixture due to the preservation of protein complexes on the MALDI surface. Where the relative intensity of the binding peptide in the control sample is sufficient for it to be clearly discerned above noise, it is a simple matter to discern its interaction from a manual inspection of both mass spectra. However, where the peptide's relative intensity in the control spectrum is low or represents the most abundant or base peak, comparison of such spectra is more difficult by manual means. This is further exacerbated when multiple peptide segments occupy the binding domain.

To assist with the identification of the binding peptide(s) using this mass spectral approach, we have devised and implemented a new computer algorithm. The program, known as PRISM, is able to automatically identify protein interactions from the spectra of masses using  $m/z$  and peak area data directly imported for a pair of MALDI mass spectra recorded for the control and reaction sample. The program is operated through a simple, intuitive graphical user interface (GUI). As well as the application described above, the algorithm is shown to have general utility whenever differences between two mass spectra of any type need to be determined.

Published online December 11, 2006

The program is available to individuals from not-for-profit institutions under a license agreement. Interested persons should contact the corresponding author.

Address reprint requests to Dr. K. Downard, School of Molecular and Microbial Biosciences G-08, The University of Sydney, Sydney, NSW 2006, Australia. E-mail: kdownard@usyd.edu.au

## Methods

The PRISM algorithm is written in the Java programming language (v 1.5) and consists of three major components for data input and output, preprocessing, and analysis. The program has been compiled to run on the Windows 2000 operating system and operates within a simple graphical user interface (GUI). Spectral lists of  $m/z$  values and peak areas are imported in text file format. Mass or  $m/z$  values can be entered where ions can be “deisotoped” or their isotopic peaks merged before analysis as desired. Datasets are processed to remove  $m/z$  values of ions of low signal-to-noise relative to the most abundant base peak (default 1%) and which do not share a common  $m/z$  value in the control and reaction spectrum within a specified error (default  $100 \times 0.001 = 0.1\%$ ).

The analyzer is employed to compare the two mass spectra in terms of changes in the area under each of the ion signals relative to a “constant peak”. The constant peak is determined by measuring the ratio of areas of all adjacent peaks  $n + 1/n$  (in terms of  $m/z$ ) in both the spectrum of the control and the spectrum of the reaction mixture. The ion peak  $n$  for which the absolute difference is the smallest, is assigned as the constant peak, as these ions are the least likely to represent a binding peptide. The value for the relative area of the constant peak to the base peak in the control spectrum is then used to establish an imagi-

nary peak in the reaction spectrum so that the ratio of area of the constant peak to the imaginary peak is identical to this value (eq 1)

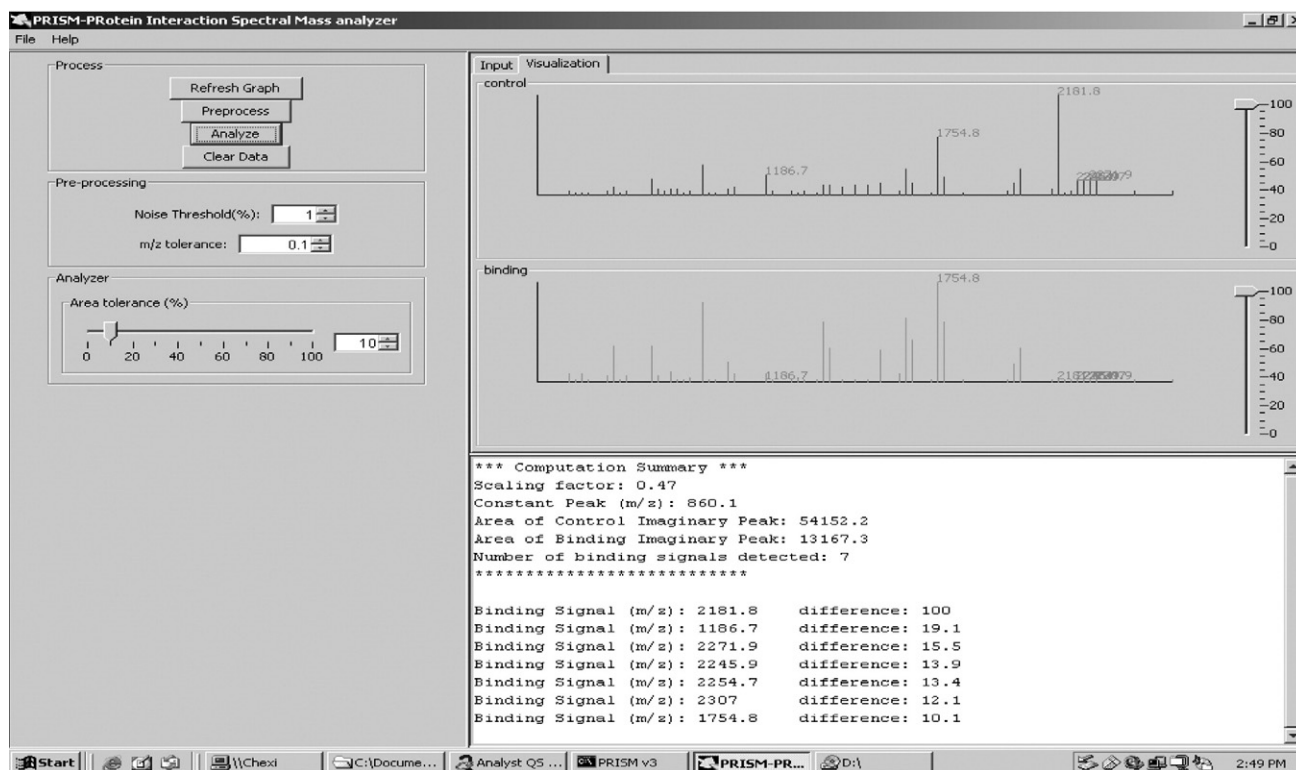
$$\frac{\text{control spectrum}}{\text{area (constant peak)}} = \frac{\text{reaction spectrum}}{\text{area (imaginary peak)}} \quad (1)$$

The algorithm then compares the absolute difference in area across both spectra relative to the base peak and imaginary peak of the control and reaction mixture, respectively, for ions common to both spectra. The data are output as a list of potential binding peptides, above a stipulated area difference tolerance, as the absolute difference in relative area calculated across the two datasets. The identity of the constant peak and its area and that of the imaginary peak are also reported in the output.

## Results and Discussion

### Overview of the PRISM Algorithm and Graphical User Interface

The PRISM algorithm consists of four major steps involving data input, preprocessing, output, and analysis. A screenshot of the graphical user interface is shown in Figure 1. It is comprised of a control panel to



**Figure 1.** View of the graphical user interface of the PRISM algorithm showing data and output for an antibody-treated tryptic digest of the hemagglutinin antigen of a type A H1N1 influenza strain (Beijing 262/95).

set the noise and  $m/z$  error and area difference thresholds, the input panel to display the lists of  $m/z$  data or spectra, and the output panel to display the results of running the algorithm with the input data.

Data are input as two lists of  $m/z$  values and peak areas obtained from the MALDI mass spectra for the control and reaction sample mixtures. The dataset is preprocessed to remove values for ions of low signal-to-noise where the default noise value is 1% of the most abundant base peak in each spectrum. Values for ions present, only the spectrum of the reaction mixture is also removed where their  $m/z$  values differ from those in the control by a user defined error (default 0.1%). The input data can be visualized on a  $m/z$  versus relative peak area plot by selecting the visualization tab. In this window, the user has manual control of the spectra by means of slides to the right hand side of each spectrum. This assists with a visual comparison of the relative peak areas of various ions in each dataset.

The analyzer is employed to compare the two mass spectra in terms of changes in the area under each of the ion signals relative to a “constant peak.” The constant peak can be considered to be an ion of a nonbinding component whose relative intensity remains unchanged (within experimental error defined by the area tolerance) in the two spectra. This peak may not be the base peak or most abundant of either the control or reaction spectrum since the base peak itself may represent ions of segments of the binding interface. In PRISM, the constant peak is determined first by measuring the ratio of areas of adjacent peaks common to both the spectra of the control and the reaction mixture. The smallest absolute difference in the ratio of areas for peaks  $(n + 1)/n$  between the two spectra establishes the constant peak ( $n$ ) as these ions are the least likely to represent a binding peptide. The area of the constant peak is then measured relative to the base peak in the control spectrum. This is then standardized in the spectrum of the reaction mixture by introducing an “imaginary peak” to the latter. Thus, if the constant peak assigned is 80% of the base peak in the control spectrum, an imaginary peak is added to the reaction mixture data such that the area of the constant peak is 80% of it in the reaction spectrum. All peak areas are measured relative to the base or imaginary peak in the control and reaction mixture spectra, respectively. The absolute relative peak area differences are subsequently calculated between the control and binding spectrum. The ion signals that undergo the largest absolute decrease in percentage area after reaction in terms of their relative area are identified as representing possible binding peptides. In this regard, some tolerance in the relative area change can be entered as a percentage (default 10%) by the user to take experimental fluctuations into account. Those ions whose relative areas vary in the control and reaction mixture spectra by a value greater than this tolerance are listed in terms of their  $m/z$  value and area difference in the lower right hand corner of the

GUI. Note that the percentage area difference for the constant peak will be zero.

An inspection of this dataset, and a manual adjustment of the spectral intensities within the visualization window to plot both spectra to a common base peak, enable the program output to be assessed. The user then inputs a higher or lower peak area difference tolerance to reduce or expand the number of identified binding peptides. The output appears in the output panel and lists the  $m/z$  for ions common to both mass spectra and the difference in relative peak area in both spectra. To assist the user with analysis, the identity of the constant peak is also shown in the header of the output data together with the noise,  $m/z$  error, and area difference tolerance. The final results of any run can be saved as a text file to save the user from having to repeat the process for the same datasets.

### *Evaluation of the Performance of the PRISM Algorithm for Various Datasets*

The performance of the PRISM algorithm has been tested for several experimental datasets. We have recently identified a determinant within the hemagglutinin to a type A H1N1 influenza strain (Beijing 262/95) partially separated from other viral proteins by gel electrophoresis before antibody treatment [8]. The MALDI mass spectra for control and antibody treated samples were input into the algorithm as a pair of lists of  $m/z$  values and peak areas calculated relative to the constant peak. Preprocessing of the spectral data removes peaks below a noise default threshold of 1% from further analysis. Ion peaks common in both spectra within a  $m/z$  error tolerance of 0.1% were retained.

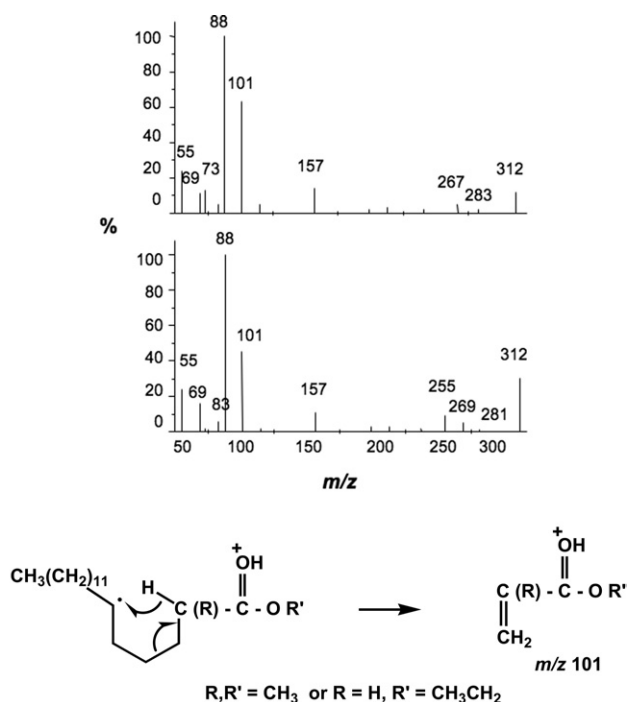
Relative peak areas that differ in absolute value by greater than 10% (area tolerance) were reported by the algorithm during analysis. Seven ions satisfied this criterion but only one ion signal at  $m/z$  2181.8 had an absolute change in relative area that exceeds 20%. This ion is completely absent in the spectrum of the antibody treated sample consistent with its complete binding to antibody. This peptide has been shown to comprise residues 206–224 of the hemagglutinin antigen [8]. The next ion signal at  $m/z$  1186.7, and the only other that

#### \*\*\* Computation Summary \*\*\*

Scaling factor: 0.6  
 Constant Peak ( $m/z$ ): 1219.5  
 Area of Constant Peak: 51016.8  
 Area of Imaginary Peak: 8152.8  
 Number of binding signals detected: 3  
 \*\*\*\*\*

Binding Signal ( $m/z$ ): 2071.8	difference: 50
Binding Signal ( $m/z$ ): 2464.1	difference: 45.5
Binding Signal ( $m/z$ ): 2835.2	difference: 18.5

**Figure 2.** Output of the PRISM algorithm for data obtained for an antibody-treated tryptic digest of the hemagglutinin antigen of a type A H3N2 influenza strain (Panama 2007/99).



**Figure 3.** Electron ionization mass spectra recorded for two isomeric long-chain fatty acids  $\alpha$ -methyl methyl stearate and ethyl stearate recorded with an electron beam energy of 70 eV [10].

undergoes an absolute change in relative area of greater than 15%, is identified as a region that flanks this determinant across residues 226–235 [8].

The above dataset poses little challenge to the algorithm since it represents a case where there is a single binding peptide with an abundant ion signal in the control spectrum. It is clear from the data showing in Figure 1 that this result could be discerned by manual evaluation of the spectral pair. In the case of several binding partners and/or those that contain ions that are less abundant in the control spectrum and that undergo a less pronounced change in relative area, such a manual evaluation is more challenging. This is where the algorithm provides particular advantages.

A number of determinants have been identified in a single mass spectrum for data obtained for several H3N2 subtypes of the influenza virus [9]. An output for the dataset for a Panama strain of the virus is shown in Figure 2. Three ions are detected by the algorithm above the nominated area tolerance threshold of 15%. The ion signal with the greatest absolute difference in relative area (50%) appears with a  $m/z$  of 2071.8. These peptide ions have been confirmed by mass mapping and tandem mass spectrometry to comprise residues 109–125 (AYSNCYPYDVPDYASLR) of the hemagglutinin HA1 subunit. Ions at  $m/z$  2464.1 which, under an absolute relative area decrease of 45.5%, reside within a second determinant across residues 159–178 of the HA2 subunit, while the larger peptide with ions at  $m/z$  2835.2 encompasses the same region over residues 160–183.

The PRISM algorithm has utility for the comparison of any two mass spectra more generally. Two electron ionization mass spectra for the isomeric long-chain fatty acids  $\alpha$ -methyl methyl stearate and ethyl stearate are shown in Figure 3 [10]. Only one ion signal at  $m/z$  101 is found to undergo an absolute change in relative area that exceeds 15% when the spectral data are analyzed by the PRISM algorithm. This ion is associated with a fragment ion formed by a McLafferty type  $\gamma$ -hydrogen rearrangement [11] in the vicinity of the ester group, resulting in cleavage of the aliphatic chain (Figure 3) and the formation of ions that have the elemental composition  $C_5H_9O_2^+$ . This ion has the structure  $CH_2=CH(R)-C(=OH^+)OR'$ . The formation of this ion is favored in the case of ethyl stearate, since the  $\alpha$ -carbon has two hydrogen atoms available for the rearrangement over a single hydrogen atom at the same position in  $\alpha$ -methyl methyl stearate (Figure 3).

## Conclusions

The PRISM algorithm has been shown to be able to successfully identify protein-binding domains from pairs of MALDI mass spectra that reflect a decrease in the relative peak area of their ion signals in the spectrum of a treated sample versus an untreated control. The algorithm will advance the automation of the MALDI approach for identifying protein interactions and should prove useful in any mass spectrometric studies in which the identification of differences across pairs of mass spectra is required.

## Acknowledgments

The algorithm was written by JWKH as part of his activities under the talented student program (TSP) at the University of Sydney.

## References

- Kiselar, J. G.; Downard, K. M. Direct Identification of Protein Epitopes by Mass Spectrometry Without Immobilization of Antibody and Isolation of Antibody-Peptide Complexes. *Anal. Chem.* **1999**, *71*, 1792–1801.
- Kiselar, J. G.; Downard, K. M. Antigenic Surveillance of the Influenza Virus by Mass Spectrometry. *Biochemistry* **1999**, *38*, 14185–14191.
- Kiselar, J. G.; Downard, K. M. Preservation and Detection of Specific Antibody-Peptide Complexes by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 746–750.
- Hutchens, T. W.; Yip, T.-T. New Desorption Strategies for the Mass Spectrometric Analysis of Macromolecules. *Rapid. Commun. Mass Spectrom.* **1993**, *7*, 576–580.
- Yip, T.-T.; Hutchens, T. W. Affinity Mass Spectrometry. Probes with Surfaces Enhanced for Affinity Capture (SEAC) of Lactoferrin. *Exptl. Biol. Med.* **1997**, *28*, 39–58.
- Nelson, R. W.; Krone, J. R.; Bieber, A. L.; Williams, P. Mass Spectrometric Immunoassay. *Anal. Chem.* **1995**, *67*, 1153–1158.
- Schuerenberg, M.; Luebbert, C.; Eickhoff, H.; Kalkum, M.; Lehrach, H.; Nordhoff, E. Prestructured MALDI-MS Sample Supports. *Anal. Chem.* **2000**, *72*, 3436–3442.
- Morrissey, B.; Downard, K. M. A Proteomics Approach to Survey the Antigenicity of the Influenza Virus by Mass Spectrometry. *Proteomics* **2006**, *6*, 2034–2041.
- Morrissey, B.; Streamer, M.; Downard, K. M. Antigenic Characterization of the H3N2 Subtypes of the Influenza Virus by Mass Spectrometry, submitted for publication.
- Watson, J. T. *Introduction to Mass Spectrometry*; Lippincott-Raven Publishers: Philadelphia, 1997; p 189.
- McLafferty, F. W. Mass Spectrometric Analysis. Molecular Rearrangements. *Anal. Chem.* **1959**, *31*, 82–87.