Development of software for automated LC/MS data processing, protein assignment, and inter-sample comparison of intact protein mixtures

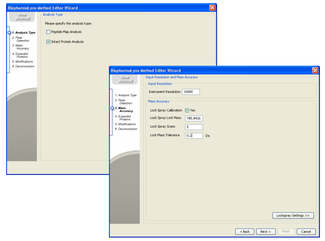
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Scott J. Berger, Asish B. Chakraborty, J. Daniel Ferrari and John C. Gebler
Biopharmaceutical Sciences Group, Waters Corporation, Milford, MA

OVERVIEW

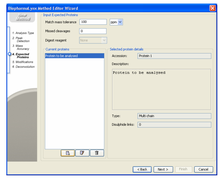
- Intact protein LC/MS analysis can provide a wealth of information about a biopharmaceutical, and the process by which it was created.
- Processing of data sets typically follows a generic workflow, making processing/annotation of large data sets a repetitive and often tedious process.
- Automated processing of intact protein LC/MS data has been accomplished (e.g. automated MaxEnt1 data processing under the open access program OpenLynx), but has not encompassed annotation of deconvoluted protein masses to their likely structures, nor facilitated direct comparisons between individual results.
- This poster describes ongoing efforts to integrate all such capabilities in to a single application geared towards automated analysis of protein infusion and LC/TOF-MS protein data.
- We will highlight the rationale used to process intact protein LC/MS data, assign masses to proteins and their variants, and illustrate various approaches for inter-sample comparisons.

DEFINING ACQUISITION PARAMETERS

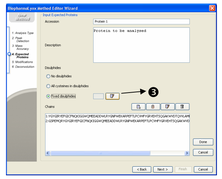


Initially, the software needs to be told it is analyzing intact protein data, and whether a lock mass data channel was acquired with the data. Instrument resolution is required for taking advantage of several automated processing settings related to protein deconvolution.

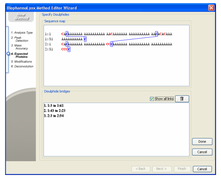
PROTEIN ASSIGNMENT PARAMETERS



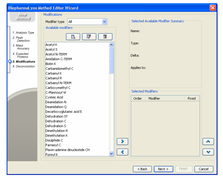
❶ Setting search parameters: What protein or proteins will be searched against, and at what initial mass search tolerance.



❷ Defining a protein: The basic unit of any protein is the protein chain. Multiple chains can associate noncovalently, or covalently through one or more disulfide linkages.

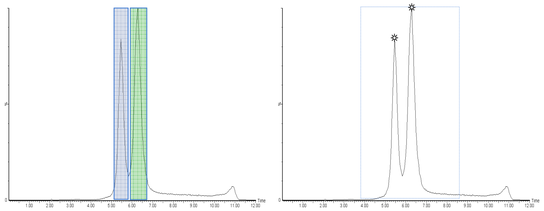


❸ Disulfide linkages within a chain or between chains can be specified and viewed.



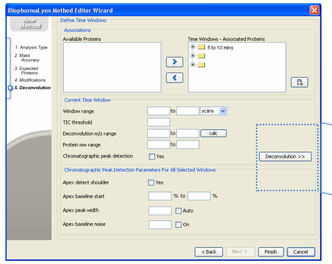
❹ Proteins can have fixed modifications and variable modifications. Modifications can be restricted to a given amino acid (e.g. Methionine Oxidation), to a specific location (N-terminal modifications), or at specific sequences (e.g. N-glycosylation).

RATIONALE FOR PROCESSING INTACT PROTEIN LC/MS DATA

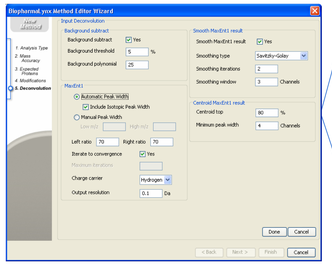


LC/MS runs can be divided into multiple processing segments where unique sets of processing parameters and protein matches can be defined. Within each zone, all spectra can be summed (LEFT), or individual peaks (RIGHT) can be recognized, processed, and assigned to unique apex retention times for inter-sample comparisons.

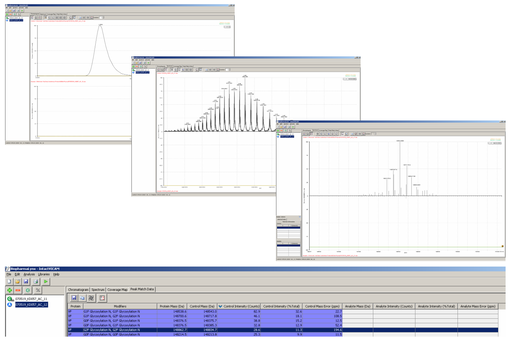
Even simple samples (e.g. reduced antibodies) can generate LC/MS data where peaks are optimally processed using distinct sets of deconvolution and protein assignment parameters. In this window the user assigns protein, peak detection, and deconvolution parameters to individual processing segments.



While the default deconvolution parameters should suffice for initial sample screening of LC/MS runs, the power to tune deconvolution parameters for optimal processing of each run segment is supported.

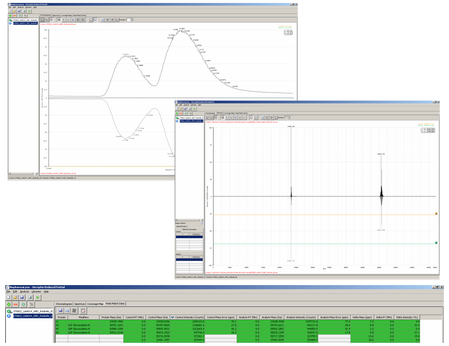


COMPARISON OF INTACT ANTIBODY LC/MS RUN WITH POST-RUN BLANK



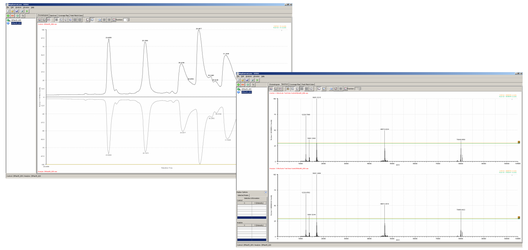
An analysis of an intact antibody by LC/MS and the subsequent blank injection run shows minimal carryover by stacked TIC (TOP), summed raw spectra plot (MIDDLE), and a mirror plot of the processed deconvoluted spectra (LOWER). The most abundant components in the antibody run corresponded to common glycosylation variants (TABLE AT BOTTOM).

COMPARISON OF REDUCED ANTIBODY LC/MS RUNS



An analysis of two preparations of a reduced antibody were analyzed by LC/MS. The stacked TIC (TOP), and a mirror plot of the processed deconvoluted spectra (LOWER) show consistency between the two preparations. The most abundant components in the antibody run corresponded to the light chain and a common series of heavy chain glycosylation variants (TABLE AT BOTTOM).

COMPARISON OF STANDARD PROTEIN LC/MS RUNS



Five standard proteins (Cytochrome C MW 12,230, Myoglobin 13,680, RNase A 16,950, Enolase 46,672, and apoTransferrin ~80 kD for average of 60 major glycoforms) were analyzed by LC/MS. Two repeat injections produced comparable results in a mirror TIC plot (TOP) and stacked deconvoluted spectra plot (BOTTOM). Individual processing segments were defined for optimal deconvolution of each protein peak.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS