# Progenesis LC-MS v2.0 – Validating a unique approach for label-free LC-MS data analysis

M. O'Gorman[§], A. R. Ivanov[¤], P. Lavery[§], M. Bennett[§*],

[§]Nonlinear Dynamics, Newcastle-upon-Tyne, UK. [¤]Department of Genetics and Complex Diseases Proteomics, HSPH Proteomics Resource, Harvard School of Public Health, Boston, USA,
*all correspondence to mark.bennett@nonlinear.com

## Introduction

**Progenesis LC-MS** (Nonlinear Dynamics) applies a unique approach to label-free analysis based on high quality alignment of LC-MS data. Detection and quantification of all peptide ions is followed by expression analysis and identification of the proteins they originate from. Progenesis LC-MS v2.0 has a Protein View for protein level analysis, which includes quantification by combining peptide ion measurements with qualitative protein identification data generated from LC-MS/MS runs and database searching. The software can perform differential expression analysis of digested complex protein mixtures and determine the proteins of interest within them. It can also be used for protein characterisation based on peptide ions within a single sample. The present study validates the analysis approach by quantifying and identifying known proteins in the 2-1000 fmol range and demonstrates that the linearity of measured response is consistently high (>0.9) across this range.

## Experimental Details

**Sample -** A mixture of known purified protein standards (Sigma) listed in Table 1 was digested with trypsin (Promega) and three replicate injection volumes of 2, 5 and 10uL of the mixture were subjected to LC-MS/MS.

| Protein | Conc. fmol/uL | 2uL | 5uL | 10uL |
|---|---|---|---|---|
| | | Protein Digest Amount, Fmol | | |
| Enolase 1 (2-phosphoglycerate dehydratase, Eno1) [Saccharomyces cerevisiae] | 100 | 200 | 500 | 1000 |
| Glutamate dehydrogenase (GDH) | 50 | 100 | 250 | 500 |
| Bovine serum albumin (BSA) [Bos taurus] | 25 | 50 | 125 | 250 |
| Serotransferrin precursor (Transferrin) [Bos taurus] | 25 | 50 | 125 | 250 |
| Glycolytic enzyme phosphoglucose isomerase; Pgi1p [Saccharomyces cerevisiae] | 10 | 20 | 50 | 100 |
| Immunoglobulin gamma-chain (IgG) [Sus scrofa] | 10 | 20 | 50 | 100 |
| Lysozyme C precursor (Gal d IV) [Gallus gallus] | 10 | 20 | 50 | 100 |
| Lactoperoxidase precursor (LPO) [Bos taurus] | 5 | 10 | 25 | 50 |
| Alcohol dehydrogenase I (ADH1, YADH-1) [Saccharomyces cerevisiae] | 1 | 2 | 5 | 10 |

***Table 1:*** *Nine proteins and their concentrations within a mixture that was digested by trypsin and analysed by LC-MS/MS*

**LC conditions -** 100 um x 16 cm column packed with Magic C18 AQ, 3 um (Michrom Bioresources) with effective gradient 2%B to 40%B in 30 min. Solvent A was 0.1% formic acid 2% acetonitrile in water, solvent B was 0.1% formic acid, 5% isopropanol, 10% water in acetonitrile.
**Mass Spectrometry -** Thermo LTQ Orbitrap (Thermo Fisher Scientific) ESI with Xcalibur v2.2.
**LC-MS data analysis –** Progenesis LC-MS v2.0 (Nonlinear Dynamics).

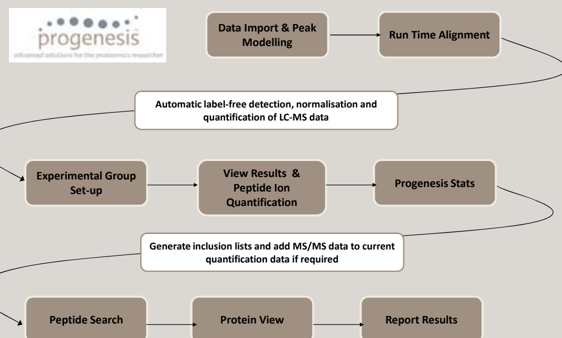


***Figure1:*** *A schematic illustrating the label-free LC-MS/MS data analysis workflow within Progenesis LC-MS v2.0, for more information visit www.nonlinear.com/LC-MS .*

## Results

Reliable quantification of peptide ions was assessed by generating linear fits of mean raw abundance from three technical replicates vs. injection volume for every peptide ion detected within the experiment (a total of 1569). By plotting the $R^2$ values vs. peptide ions 83% of features gave linear fits with $R^2$ values of >0.8 and 75% of features gave linear fits with $R^2$ values of >0.9.

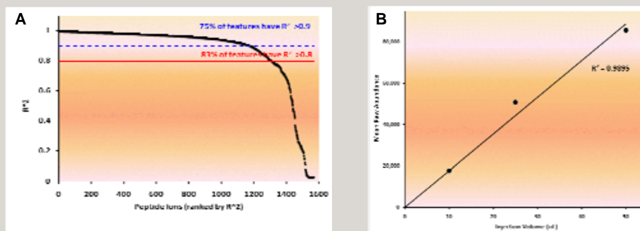


***Figure 2:*** ***(A)*** *$R^2$ plotted for every peptide ion in the experiment showing % of features with $R^2$ >0.8 with 83% of features and >0.9.* ***(B)*** *One example of mean raw abundance plotted against injection volume with a linear fit showing how $R^2$ values were generated for every peptide ion. In this example the protein concentration range was 5 - 500fmol.*

| Protein | Conc. fmol/ul | $R^2$ | No. Peptides | Score |
|---|---|---|---|---|
| Serotransferrin precursor (Transferrin) [Bos taurus] | 25 | 0.988 | 48 | 177.77 |
| Bovine serum albumin (BSA) [Bos taurus] | 25 | 0.981 | 36 | 141.117 |
| Glutamate dehydrogenase (GDH) | 50 | 0.982 | 20 | 76.013 |
| Enolase 1 (2-phosphoglycerate dehydratase, Eno1) [Saccharomyces cerevisiae] | 100 | 0.979 | 20 | 72.253 |
| Lactoperoxidase precursor (LPO) [Bos taurus] | 5 | 0.993 | 18 | 60.021 |
| Lysozyme C precursor (Gal d IV) [Gallus gallus] | 10 | 0.989 | 13 | 58.674 |
| Alcohol dehydrogenase I (ADH1, YADH-1) [Saccharomyces cerevisiae] | 1 | 0.976 | 14 | 45.666 |
| Trypsinogen | - | 0.997 | 6 | 25.115 |
| Glutamate dehydrogenase 1, mitochondrial precursor (GDH) | (50) | 0.924 | 4 | 19.625 |
| Catalase | - | 0.948 | 6 | 14.9 |
| Glycolytic enzyme phosphoglucose isomerase; Pgi1p [Saccharomyces cerevisiae] | 10 | 0.968 | 4 | 11.734 |
| Immunoglobulin gamma-chain (IgG) [Sus scrofa] | 10 | 0.961 | 2 | 6.377 |

***Table 2:*** *All nine known proteins in the mixture, as well as several contaminants and an additional entry for GDH, were quantified and correctly identified by label-free LC-MS data analysis using Progenesis LC-MS.*

All the known proteins within the mixture were correctly detected. The list of proteins generated by Progenesis LC-MS ranked by score (Table 2) has a different order than the list of known proteins ranked by concentration (Table 1). This is expected since the number of tryptic peptides, and therefore the protein score calculated by the sum of peptide scores, is dependent on protein structure and abundance not on protein concentration.

The linearity of response for measuring increasing amounts of all nine known proteins was assessed based on plots of calculated protein abundance, generated by Progenesis LC-MS, vs. injection volume. Figure 3 shows three examples of such plots used to generate the $R^2$ values reported above (Table 2). The value of $R^2$ is not significantly affected by either the number of peptide ions or the protein concentration in the injected sample.

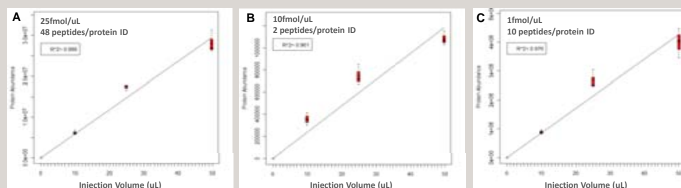


***Figure 3:*** *Best linear fit measured from plots of protein abundance vs. injection volume (10, 25 and 50uL) including box plots showing replicate variation for* ***(A) Serotransferrin precursor. (B) Immunoglobulin gamma-chain, (C) Alcohol dehydrogenase I.***

## Workflow performance

**Speed**
**Complete quantification of 9 LC-MS/MS runs, average file size of 62MB**, imported as .mzXML files **took 15 minutes** excluding database searches. Fast analysis enables you to run more replicates and get reliable results in less time.

**Objectivity**
Analysis was performed using automatic settings at every step so **identical results were produced from three independent analyses** (data not shown). This facilitates the ability to reproduce results between experiments and across-labs.

**Statistics**
Peptide ion quantification results and **the final list of proteins were selected based on objective measures within the software** e.g. Protein score, number of peptides. The alignment based analysis approach provides a complete data set with no missing values for valid and robust statistical analysis including ANOVA (p-value), false discovery rate control, PCA, correlation analysis and power analysis.

## Conclusions & Discussion

- The unique approach for label-free data analysis including high quality alignment of LC-MS data, detection and quantification of peptide ions, expression analysis and protein identification has been validated.
- Peptide ion expression changes in the 2-1000fmol range were reliably detected with 83% of features producing $R^2$ >0.8 from plotting mean raw abundance vs. injection volume (Figure 2). There was no noticeable difference in the number of features showing $R^2$ >0.8 as abundance increased over the 2-1000fmol range.
- All the known proteins within the mixture were correctly detected. Every one showed a linear response based on calculated protein abundance vs. injection volume with $R^2$ values >0.9 (Figure 3).
- The appearance of trypsinogen, catalase and an additional GDH entry within the final protein list (Table 2) is due to insufficient purity of the standards, sample contamination during sample preparation and addition of enzymes. The protein view allows you to investigate these anomalies and accept or reject a protein identity as valid for experimental results.

## Future Work

- Develop tools that allow users to filter results at the "protein search" step using prior knowledge to reduce the complexity of data and increase the quality of results displayed at the protein level.
- Investigate reproducibility of analysis using a complex sample and Progenesis LC-MS.
- Investigate reliability and sensitivity of quantitative LC-MS analysis using stable isotope labelling by Progenesis LC-MS. **Initial results have been submitted to the sPRG study being presented at ASMS (Special Poster ABRF 2)** and further results are in progress for publication.

**Visit us at Booth #45 today or at www.nonlinear.com/ASMS**