PROTEIN MOLECULAR WEIGHT, REPLICATION, AND FALSE POSITIVE IDENTIFICATIONS

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

1Martha D. Stapels, 1Scott J. Geromanos, 1Craig A. Dorschel, 2James Langridge, 2Hans Vissers and 3Jan Claereboudt
1 Waters Corporation, Milford, MA, United States; 2 Waters Corporation, MS Technologies Centre, Manchester, United Kingdom; 3 Waters Corporation, Central Europe, HRMS Division, Zellik, Belgium

INTRODUCTION

A major challenge in proteomic studies is the elimination of false positive identifications. One method for minimizing false positives is searching a database containing reversed sequences[1]; however, this method appears to underestimate false positives[2]. The number of false positives in lists of identified proteins can alternatively be estimated by comparing the molecular weight (MW) distribution of identified proteins to the distribution from the protein database. The average MW of proteins in the human database is 52 kDa (median 40 kDa) and the distribution is very similar across species (Figure 1). Although large MW proteins generate more peptides, the identification of a protein and the number of peptides matched to it depend on both the MW and the concentration of that protein in a complex mixture (Figure 2). In this study, a real proteomic experiment is compared to a worst case scenario where peptides are randomly chosen from a database to show trends in good versus bad protein identifications.

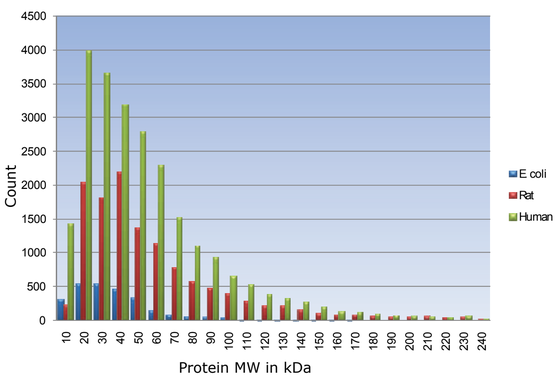

***Figure 1**. Similar distribution of protein MW for different species.*

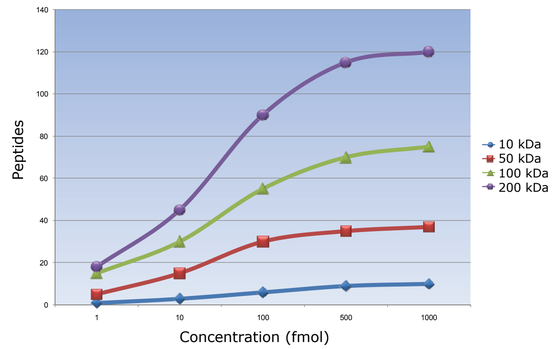

***Figure 2.** Number of identified peptides to a protein is proportional to the MW and concentration of the protein.*

METHODS

Real Proteomic Experiment

Proteins from rat brain lysate separated by 1D-PAGE
Four gel lanes cut into four fractions each
In-gel tryptic digestions performed
Waters Identity^E High Definition Proteomics system used to acquire and process data
Proteins had to be identified in at least 2 replicate injections

Random Proteomic Experiment

Rat database tryptically digested *in silico*
5,000 peptides chosen at random
Peptide identifications collapsed into identified proteins
Experiment repeated four times

RESULTS

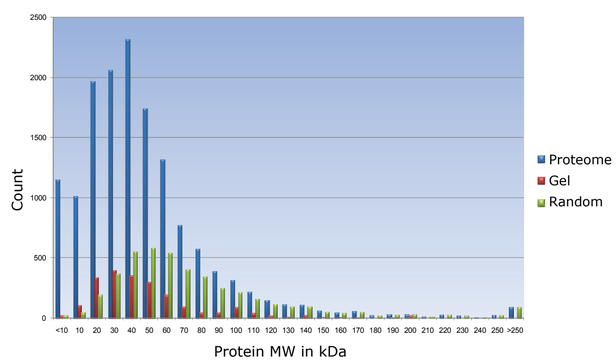

***Figure 3.** Distribution of protein MW for entire rat proteome, identified proteins from the gel, and randomly identified proteins.*

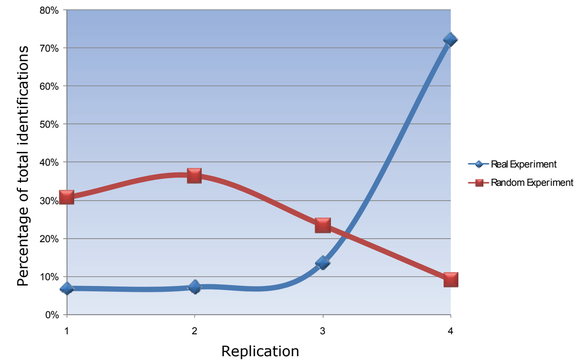

***Figure 4.** Replication rate in four experiments.*

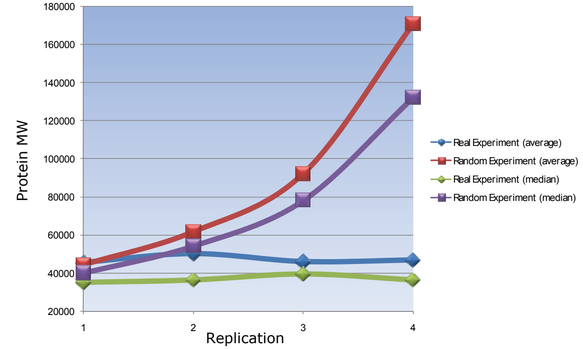

***Figure 5.** Regardless of replication rate, the average or median MW of identified proteins mimics that of the proteome. With randomly identified proteins, the average MW increases as a function of replication.*

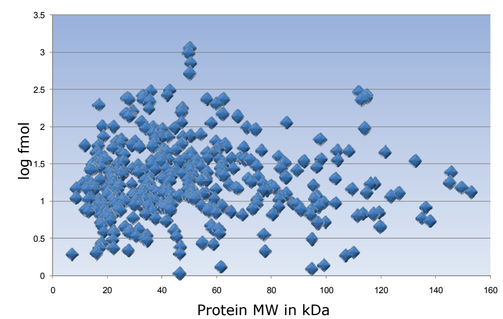

***Figure 6.** Plot illustrates the MW and measured molar amounts of proteins in rat brains. The majority of proteins are low in MW. Very few abundant high MW proteins are present.*

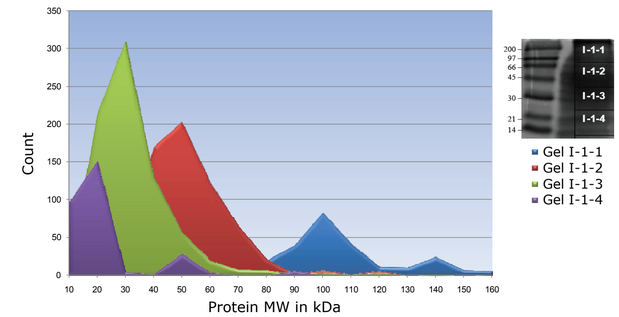

***Figure 7.** Proteins identified from 1D-PAGE mimic the MW distribution in the gel.*

CONCLUSION

- **MW distribution of correctly identified proteins mimic that of the proteome**
- **Random identifications shift the median MW towards higher MW**
- **In good proteomic data, the majority of identified proteins replicate**
- **Replication rate should increase as an experiment is repeated**
- **Proteins separated by MW show resulting MW distributions**

References

1. Elias, *et. al*, *Nat. Biotechnol*, **2004**, *22*, 214-219.
2. Yu, Li-Rong, et. al, *J. Proteome Res,* **2007**, *6*, 4150-4162.

ACKNOWLEDGEMENTS

The authors wish to thank An Zhou, Chelsea Piper, and Roger Simon at Legacy Research for sample preparation.

50 YEAR LEGACY
VISION ■ CONTRIBUTION ■ ADVANCEMENT

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

720002501EN