

# Universal Proteomics tools for Protein Quantitation and Data Management – Xome & Mass Navigator

**Koji Yanagisawa**<sup>1</sup>  
yanagi@hydra.mki.co.jp

**Yuji Miura**<sup>1</sup>  
yuji2@hydra.mki.co.jp

**Tsuyoshi Tabata**<sup>2</sup>  
t2-tabata@hhc.eisai.co.jp

**Yasuto Yokoi**<sup>1</sup>  
yokoi@hydra.mki.co.jp

**Yoshiya Oda**<sup>2</sup>  
y-oda@hhc.eisai.co.jp

**Takatoshi Kawai**<sup>2</sup>  
t-kawai@hhc.eisai.co.jp

**Ken Aoshima**<sup>1</sup>  
kaku@hydra.mki.co.jp

**Yasushi Ishihama**<sup>2</sup>  
y-ishihama@hhc.eisai.co.jp

**Takeshi Nagasu**<sup>2</sup>  
t-nagasu@hhc.eisai.co.jp

<sup>1</sup> R&D Dept. Bioscience Division, Mitsui Knowledge Industry Co., Ltd. Harmony Tower 21<sup>st</sup> floor, 1-32-2 Honcho, Nakano-ku, Tokyo 164-8721, Japan

<sup>2</sup> Laboratory of Seeds Finding Technology, Eisai Co., Ltd. 1-3, Tokodai 5-chome, Tsukuba-shi, Ibaraki 300-2635, Japan

**Keywords:** Xome, Mass Navigator, LC/MS/MS, Proteomics, Quantitation, BISCUIT

## 1 Introduction

In post-genomic era, functional genomics have performed to understand complex biological processes. Proteomics, as a tool for comprehensive analysis of proteins, has been used for protein expression, localization, post-translation modification and protein-protein interaction. Since protein expression is dynamically controlled, the quantitative information is important to address the function of proteins.

There are several approaches for quantitative proteomics. Seminal work on differential in vivo isotopic labeling proteins by Oda et al [1] has led to the development of tools such as isotope coded affinity tags (ICAT) [2]. In ICAT strategy, several drawbacks have been reported such as time-tedious labeling steps, less reproducibility, ambiguous ms/ms fragmentation and expensive reagents, and in vivo labeling method has faced the difficulty to label with isotopes in animal tissues. Eisai Co., Ltd. developed the new quantitation method for animal tissues, which were termed BISCUIT method [3]. This procedure should be applicable to variety of cell systems and tissues, so that BISCUIT cells might be the global standard for in vivo labeling quantitative proteomics.

One of the most powerful analytical approaches for quantitative proteomics is nano-scale liquid chromatography coupled with tandem mass spectrometry (nano LC/MS/MS). Because nano LC/MS/MS generates a huge amount of data, it needs an automated approach to extract the required information efficiently. Nowadays there are several software applications for protein identification from MS/MS data, however, the software tailored for protein quantitation are very limited. In this study, we have developed two software packages, named “Xome” and “Mass Navigator”, to extract the data and to perform quantitative proteomics using stable isotope-based approaches such as BISCUIT.

## 2 Architecture

### 2.1 Xome

“Xome” is an analysis and management software for LC/MS/MS data. Several methods are applicable on this software - using ICAT reagent method, in vivo isotope labeling method, and BISCUIT method, and most of the major formats of LC/MS/MS data are acceptable to this software.

This module consists of the following components;

- (1) Peak-picking module: This module has functions to pick correct peaks from MS/MS data and to create a peak-list, which sends to an identification engine. Our original algorithm has been developed to allow for filtering noises, normalizing baselines and recognizing the isotopic peaks during peak picking process.
- (2) Identification module: Currently, public and commercial engines such as Mascot (Matrix Science Ltd.), Sonar MS/MS (Genomic Solutions) and X-Tandem (The Global Proteome Machines Organization) are incorporated into the Xome

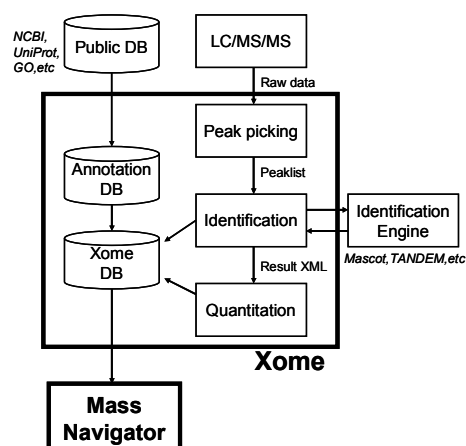


Figure 1: System architecture

identification module. This module is used not only to identify protein but also to summarize the result tables by pre-defined condition.

- (3) Quantitation module: This module searches a pair of peptides of the same sequence with and without labeling by using user-defined stable isotopes, and calculates these quantitative values. In this process, as well as  $m/z$  and peak intensities, we consider the elution time from LC and extract the mass chromatograms of peptides to obtain peak area of peptides. We calculate only the area of target peptide by dividing overlapped parts if two peaks are the same  $m/z$  and are accidentally eluted at the same time.
- (4) Xome Database: The results of identification and quantitation are stored in RDBMS.
- (5) Annotation Database: Public protein function database is downloaded automatically, and linked with identified proteins in the Xome Database.

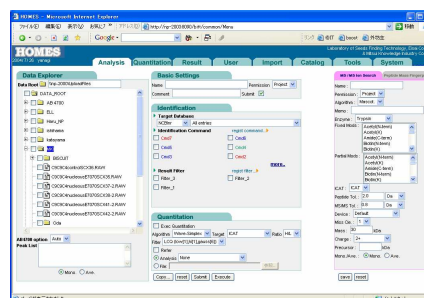


Figure 2: Xome screen shot

“Xome” is written in Java servlet architecture, is accessed via WWW browser (e.g. Internet Explorer) from client computers. The peak-picking and quantitation module is written in C++ and those outputs are stored the Oracle database. Since the results stored in database is able to output as XML or CSV format, the knowledge is shared with any other systems.

## 2.2 Mass Navigator

“Mass Navigator” is a mass spectrum data management suite. This is written in C++ and works on Windows 2000/XP. The chromatogram viewer is one of its components. In this view, it can display the area of peptide quantitated by Xome server. Furthermore, manual quantitation and correction are available since this software has the same quantitation algorithm as Xome. So, this viewer is helpful for valid the data of LC-MS/MS.

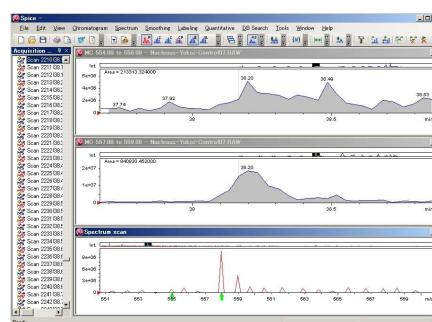


Figure 3: Mass Navigator screen shot

## 3 Discussions

In quantitative proteomics using stable isotopes, the key points are to select the pair of peaks, labeled and unlabeled, correctly, and measure their peak intensities.

Peak-picking algorithm of Xome and Mass Navigator focuses on improving accuracy by subtracting the noise of spectrum, detecting isotope clusters of the target peak and normalizing baseline of chromatogram. Therefore, the false positive of protein identification is decreased, the search time is reduced, and the calculated peak area for quantitative data is more accurate.

Xome Database is also helpful for increasing the accuracy of identification and quantitation of proteins. The result tables filtered by user-defined threshold are helpful for creating higher reliable database and eliminating redundant proteins, which have been already registered. Additionally, by storing unassigned spectra in the database, they can be identified after protein database updated, and then quantified.

## References

- [1] Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chait, B.T., Accurate quantitation of protein expression and site-specific phosphorylation, *Proc. Natl. Acad. Sci.*, 96: 6591-6596 (1999).
- [2] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.*, 17, 994-999, 1999
- [3] Oda, Y., Ishihama, Y., Sagane, K., Miyamoto, N., Sato, T., Nagasu, T., Quantitative Tissue Proteomics by Bridging Intensities using internal Standard as Cultured Isotope Tags (BISCUIT) , 51<sup>st</sup> ASMS 2004.