

# Label-free quantitation, an extension to 2DB

Jens Allmer

Received: 7 January 2009 / Accepted: 17 June 2009 / Published online: 3 July 2009  
© Springer-Verlag 2009

**Abstract** Determining the differential expression of proteins under different conditions is of major importance in proteomics. Since mass spectrometry-based proteomics is often used to quantify proteins, several labelling strategies have been developed. While these are generally more precise than label-free quantitation approaches, they imply specifically designed experiments which also require knowledge about peptides that are expected to be measured and need to be modified. We recently designed the 2DB database which aids storage, analysis, and publication of data from mass spectrometric experiments to identify proteins. This database can aid identifying peptides which can be used for quantitation. Here an extension to the database application, named MSMAG, is presented which allows for more detailed analysis of the distribution of peptides and their associated proteins over the fractions of an experiment. Furthermore, given several biological samples in the database, label-free quantitation can be performed. Thus, interesting proteins, which may warrant further investigation, can be identified en passant while performing high-throughput proteomics studies.

**Keywords** Quantitation · Quantification · Label-free · Software · MS/MS · Spectral count

## Introduction

Proteomics aims to elucidate the protein complement of a genome taking into account the spatial and temporal

expression patterns of a protein. Post translational modifications (PTM) and differential expression levels are also vitally important for understanding biological function (Vissers et al. 2008). Relative quantitation of protein expression levels under different physiological states can aid in elucidating some of these areas and can further aid in drug discovery and can lead to improved diagnostic methods (Pan et al. 2008; Roddy et al. 2007). Mass spectrometry is the method of choice for investigating peptides and their aggregates, proteins (Aebersold and Mann 2003). Protein expression levels have a high dynamic range but not all respond to regulatory events or disease with large changes in abundance (Wang et al. 2008). Therefore, quantifying protein content is of importance in comprehensive proteome analysis (Blagoev et al. 2004; Cravatt et al. 2007). In quantitation, using MS and tandem-MS, two general strategies need mentioning.

One of these involves differential labelling of peptides in different samples. In this strategy, a label which changes the mass such that the mass to charge ratio difference between labelled and label-free peptides can be discerned by a mass spectrometer (Gygi et al. 1999; Krijgsveld et al. 2003; Wang et al. 2008). The label can be incorporated metabolically, after isolation of protein/peptide, or reference peptides can be spiked into the sample prior to mass spectrometric analysis (Bantscheff et al. 2007).

Unfortunately, not all peptides ‘fly well’ and may thus escape detection by mass spectrometric experiments (Aebersold and Mann 2003). Furthermore, in high-throughput experiments many peptides are missed due to a variety of reasons but they may be recovered by performing several mass spectrometric analyses of the same sample (Elias et al. 2005). For this reason, quantitative studies involving high-throughput experiments are comparably rare; more commonly, specific proteotypic peptides

---

J. Allmer (✉)  
Molecular Biology and Genetics, Izmir Institute of Technology,  
Urla, Izmir, Turkey  
e-mail: jens@allmer.de

(Duncan et al. 2005) are selected, differentially labelled, and measured by MS or MS/MS in separate experiments.

Labelling therefore implies an additional effort which involves additional cost and time as well as a potential increase in sample complexity (Yang et al. 2007) whereas label-free quantitation may be done without additional effort. Label-free quantitation may not be as precise; results should thus be carefully interpreted. However, as a bonus during high-throughput mass spectrometric experiments, it can help to identify proteins whose differential expression is worth closer evaluation.

Label-free quantitation determines the protein content without the need of adding a marker to the peptide. This idea is based on the notion that protein abundance and number of spectra and intensity of precursor ions are correlated (Higgs et al. 2005; Silva et al. 2005). One method is based on counting the occurrence of precursor peaks with the desired mass to charge ( $m/z$ ) ratio taking into account multiple charge states. Depending on the mass accuracy, this can lead to significant problems when analyzing complex mixtures because multiple peptides will lead to ions with the same  $m/z$  ratio for one or several of the  $m/z$  ratios for the peptide of interest. An improvement over this method is to include the elution time such that elution time and precursor mass form a pair which is clearly more discriminative (Silva et al. 2005, 2006). Unfortunately, these data pairs need to be established first which require additional MS analyses. Later samples that are compared need to be normalized with regard to differences in the elution profiles (America and Cordewener 2008).

Another method, spectral counting (Gao et al. 2003; Pang et al. 2002), is based on MS/MS analysis and the correct assignment of peptides to the measured spectra (Allmer et al. 2004; Bafna and Edwards 2001; Duncan et al. 2005; Eng et al. 1994; Geer et al. 2004; Mann and Wilm 1994; Perkins et al. 1999; Shevchenko et al. 1996; Tabb et al. 2003). Based on the same idea as above, the spectra that support a protein are counted and the sum is compared across experiments. In this method, information regarding peak abundance remains unused and the resolution as far as high-throughput experiments are concerned is limited to high-abundant peptides and proteins since only identified MS/MS spectra are used.

Another approach is viable which first quantifies the MS and MS/MS data, and then identifies those spectra that contribute most to the difference among samples (Hohenwarter et al. 2008). Through their program ProtMAX, these authors implemented this approach, which is dependent on high-resolution mass spectra and employs a reverse strategy as compared to the methods mentioned above.

Due to the large amount of data generated by high-throughput mass spectrometric experiments manual data

processing is not feasible. Automated methods are therefore mandatory. Many software tools for this purpose have been reviewed in Mueller et al. (2008). We recently developed 2DB, a database to hold, study, and publish proteomics data, as generated from MS experiments (Allmer et al. 2008). An automated quantitation facility has not been available in the 2DB software but is presented herein. Since it seemed natural to incorporate this functionality, two approaches to label-free quantitation have been incorporated into the 2DB application and are presented. The aim of the newly developed tool, named MSMAG, is the identification of peptides and proteins that significantly change in abundance between several experimental conditions. Utilization of multiple internal standards and a slightly changed spectral counting method are new features that will be shown in detail. MSMAG further enables the investigation of protein-expression profiles.

## Materials and methods

A combination of spectral count and total ion current is used to determine differential expression levels. The total ion current (TIC) for all supporting peptides of a protein is summed in each biological replicate after adjusting to standard peptides, if given. The differential expression levels are calculated by determining the ratio between the normalized TIC sum of the reference sample and all other samples. The samples used for quantitation were theoretically generated whereas the samples used to analyze protein expression were taken from a former study.

### Dataset generation

It has been pointed out that the availability of benchmark datasets is not sufficient in the field of peptide and protein quantitation via a traditional single stage of MS (Schulz-Trieglaff et al. 2008). This situation seems to be equally severe when using MS<sup>2</sup> data for quantitation; therefore no suitable publicly available dataset could be found to be used with MSMAG. The datasets described in Schulz-Trieglaff et al. (2008) were used to generate a suitable dataset mirroring a theoretical high-throughput experiment using the big-three acquisition method, which automatically measures an MS/MS spectrum for the three most abundant precursor ions following a full scan. The datasets, available at <http://www.ebi.ac.uk/pride/> with the accession numbers 8161–8168, are, to the best of my knowledge, generated such that the proteins are present in equal amounts. The underlying proteins and peptides were retrieved from the author and are made available through the 2DB database (see experiment descriptions at <http://www.biolnk.com/2db>). Thereafter, the datasets were

examined and for every fourth MS scan in the original dataset three MS/MS spectra were generated using the most abundant three peaks as precursor ions if their  $m/z$  ratio coincided with one of the underlying peptides. In cases where no peptide was able to explain the  $m/z$  of the precursor ion up to a charge of five, no MS/MS spectrum was generated. This could happen, for example, from the incorporation of post-translational modifications in the original datasets.

It was assumed that an MS/MS scan would approximately take the same amount of time as an MS scan. The MS spectra were not represented in the resulting dataset when generated in the ams format, suitable for import into 2DB, and thus the amount of generated MS/MS spectra is 25% less than the amount of MS spectra in the original dataset. For each MS/MS spectrum, a, b, c, x, y, and z ions were generated for  $m/z$  values between 200 and 2,000 with their respective water and ammonium losses at an arbitrary maximum charge five or less having been determined from the precursor ion. The abundance of the precursor ion was equally distributed over all generated fragment ions. Usually, less than the possible 1,800 MS/MS spectra were generated from the original datasets comprised of 2,400 MS spectra due to the fact that some precursor peaks could not be explained by the available peptides up to a charge of five. An average of about  $1,588 \pm 38$  MS/MS spectra was generated from the eight datasets.

The resulting eight datasets thus contain data that can substitute for high-throughput experiments with similar inherent flaws such as many missing data points due to scan speed restrictions, peptides masked due to their low abundance and so forth. The MS/MS spectra are however highly theoretical and do not closely model experimental MS/MS spectra. Since the identification was done using a small peptide set (588 distinct peptides) generated from 100 proteins, and the mapping could be achieved using the peaks in the MS spectrum, the theoretical nature of the MS/MS spectra is of low importance to the current study.

Another dataset was generated from the original datasets. This time a more targeted analysis, with a higher degree of separation of the proteins applied to the MS/MS experiment, was assumed. A similar experiment as above was conducted with only two protein, IPI00108270.1 and IPI00108277.5, out of the 100 proteins available in the dataset.

The *in silico* tryptic digest of these proteins resulted in significantly less distinct peptides, more than five amino acids in length, than for the first experimental dataset. For each full scan in the eight original MS datasets, the three most abundant peaks corresponding to one of the peptides in the set were selected and MS/MS spectra were generated as described above. An average of about  $144 \pm 32$  MS/MS spectra was generated from the eight original datasets.

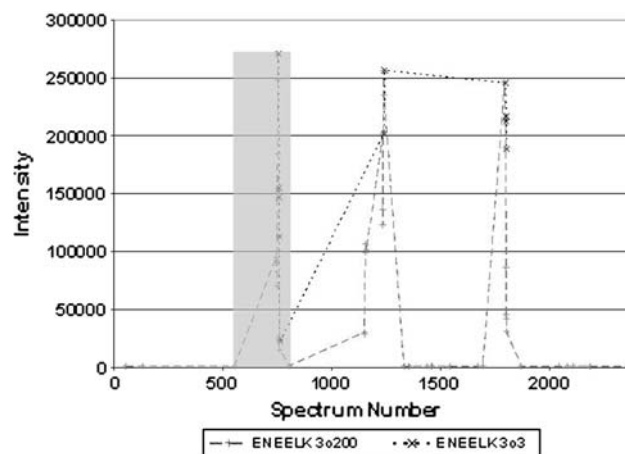
Although better, the results were not significantly different from the results using 100 proteins. Therefore, another dataset was generated where out of the 200 most abundant peaks in an MS spectrum for the first-three matching peptides (see above) an MS/MS spectrum was generated. It was assumed that the LC peak could be approximated better using this method than the big-three acquisition method as above. Unfortunately, a different error was introduced in this case (Fig. 1).

While the number of supporting spectra increased as compared to the big-three acquisition method, and the LC peaks could be re-established better for the second of the two methods described above the number of false positive identification sharply increased.

All generated datasets are available for review at <http://www.biolnk.com/2db>. All datasets are named using their acquisition number. The three different data generation methods have been named 100p3o3, 2p3o3, and 2p3o200.

## Implementation

The quantitation functionality was implemented in the JAVA™ (<http://www.java.com/>) programming language in order to shift more of the computation from the server to the client computer and thus allows more interactive data analysis. MSMAG, developed in this study, depends on an installation of the 2DB database application (Allmer et al. 2008), either locally or on a web server. The 2DB software is available at no charge with source code openly available. Data necessary for analysis are retrieved from the database



**Fig. 1** Distribution of a representative peptide from the experiment with the accession number 8161 using the big-three acquisition method for theoretically generating a high-throughput MS/MS dataset (ENEELK 3o3) and using the best 3 out of the first 200 precursor ions (ENEELK 3o200). The area which most likely corresponds to the elution time is highlighted

in a three tier fashion. The request is sent to a PHP (<http://www.php.net>) script which then queries the database and returns the result set. The application is launched using JAVA™ Web Start and therefore no local installation of the program is necessary. The Web Start feature also insures that always the latest version of MSMAG is used without the need of user intervention. Depending on the experiments stored in the database, large amounts of data may need to be transferred which calls for fast connections between clients and server. The tool presented here is also available at no charge and is currently being integrated into the 2DB package but its source is not yet available since additional extensive work is expected (see outlook).

### 2DB dependency

In addition to the dependencies mentioned in the “Implementation” section above, the software presented here is directly dependent on 2DB, a database application for storage, analysis, and presentation of data from MS experiments. 2DB is open source and MSMAG depends only on two scripts for data transfer, lending itself to facile implementation in different software environments. While 2DB allows the storage of identifications without spectral data, such data cannot be quantified with MSMAG and will be ignored by the program. In future, the dependence of MSMAG on actual fragmentation spectra may increase when other types of quantitation methods are incorporated into this software. Spectral counting, not depending on the presence of a complete MS/MS spectrum, has recently been enabled for 2DB. It is, however, limited to the comparison of two experiments at a time. Such a limitation does not exist for MSMAG.

### Data pooling

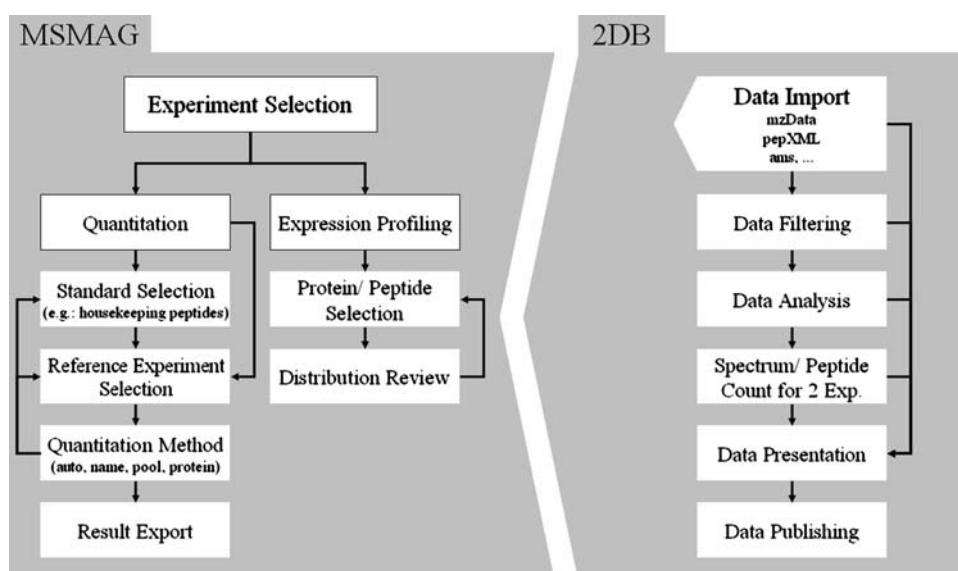
Sometimes several liquid chromatography (LC) fractions or several bands/spots from gels need to be combined since proteins of interest may smear over several bands/spots on a gel or may be present in subsequent LC fractions. To avoid manual computation in these cases, results can be pooled based on their fraction using 2DB. In case more fractions than those containing the protein are pooled, no effect on the result is expected unless cleavage products of the protein or non-proteotypic peptides are present in these fractions. In most comparison modes, data pooling has to be set specifically but for absolute protein quantification over two or more experiments all peptides found anywhere in the experimental context are pooled. Data pooling is achieved using a tool provided with 2DB which requires ownership of the experiments or administrative rights to the instance of 2DB used for the analysis.

### Working modes

Two modes of operation are currently supported in MSMAG. On one hand, expression profiling can be performed in order to examine the expression level of peptides and proteins in the fractions of an experiment. On the other hand, label-free differential protein expression can be investigated among several experiments in the database. Figure 2 displays the general flowchart for importing MS/MS data to 2DB and the major steps used in MSMAG to achieve expression profiling and quantitation.

The general flowchart in Fig. 2 indicates that data are stored in 2DB and fed forward to MSMAG. In 2DB, when data have been sufficiently reviewed it can be published, for 2DB which means that the access constraints to the data

**Fig. 2** Flowchart for achieving an expression profile and quantitation results with MSMAG. *On the right*, the processing steps in 2DB are indicated. MSMAG can be used following the import step; all other steps are optional but do not interfere with the usage of MSMAG. In general, data are fed from 2DB to MSMAG which is indicated by the lock and key shape of the two sections. *At left*, the major steps in MSMAG are shown



can be adjusted such that it can be visible for a target group (e.g., the public). The effect on MSMAG is the same as for all other data in 2DB which means that the analysis and results can be reproduced by anyone given the proper access rights.

In the following, expression profiling and quantitation will be described in more detail. Figure 2 may serve as a guide in this context.

### Expression profiling

The simplest analysis which can be done with MSMAG is locating the percentage-wise distribution of identified proteins over all the fractions in an experiment. This includes all identifications of a protein not only the most significant one. By analogy, for a one-dimensional gel separation, a protein with high molecular mass will be found at the top of the gel. If it is processed by endogenous digestion, the proteolytic fragments will be found in later bands. All instances and their distribution over the experiment are visualised and such processing can thus be uncovered. Each protein is usually identified by a set of peptides whose composition may also vary between fractions. The percentage distribution of peptides constituting a protein is also shown for each peptide.

The motivation for performing such an analysis is at least threefold. One is to investigate proteolytic control of proteins under different experimental conditions. Furthermore, it aids in finding samples to pool for later differential analysis of the overall protein content for different experimental conditions. Finally, this analysis can help in detecting proteotypic peptides for further investigations.

### Label-free quantitation

Selecting two or more experiments, the problem is that they need to be brought into correlation. This can be done automatically if corresponding fractions have been named consistently. If this is not the case, similar fractions can be correlated by the pooling tool of 2DB described elsewhere (e.g., see 2DB online help). Pooling is both valid within an experiment, and among experiments but the semantics for the label-free quantitation is slightly different. While fractions pooled from different experiments will be quantified, results within pooled fractions of one experiment will actually be summed and subsequently quantified, i.e., compared to the other experiments. The last mode comes into play if corresponding fractions between two experiments are not named equally and pooling has not been applied. In this case, all fractions of the experiments are pooled and then the overall protein amount is quantified among different experiments.

Often some proteins can serve as references for the relative quantitation of other proteins. Housekeeping proteins may be expected not to change under different physiological conditions and can therefore be used as standards. If some proteins can serve as such markers, their peptides can be used to adjust the relative differential protein expression among several biological samples. All selected reference peptides are first quantified among the selected experiments, if they occur in all of them, by the third mode of quantitation (see above). The relative difference is then averaged over all selected references. This averaged relative difference is then used to adjust the results obtained from quantitation. Since each experiment may incorporate errors at different stages of processing, such standards can help in normalizing the data and thus lead to more sound results. The remaining proteins in the experiments are then quantified. If all samples do not contain the protein in the same fraction, those samples without the protein are set to zero to indicate this. All other samples are relatively quantified with regard to the selected sample. Thus by selecting different samples, the relative quantitation is re-calculated to show the relative quantitation with respect to the currently selected sample.

### Calculations

For expression profiling, it is assumed that all measurements are done subsequently and that a total ion current (TIC), measured for a peptide at the beginning of an MS run and at the end of an MS run, would stay constant if the amount of peptide entering the mass spectrometer was the same.

For expression profiling, the TICs of all supporting spectra of a peptide are summed to yield an overall TIC. The resulting peptide TICs are in the same fashion aggregated to form a protein TIC. For each identification of a protein or a peptide, in separate fractions, the percentage of the TIC in that fraction from the overall TIC for that protein/peptide is calculated. The percentage from the overall TIC is recorded for each protein/peptide and each fraction of the experiment.

Calculations for label-free quantitation are done using the total ion current for all peptide identifications. This means that the relative quantitation is different from a spectral count since spectra go into the calculation with different weights according to their TIC. This also means that providing reference peptides for the calculation is essential with respect to the fact that between different runs, even on the same mass spectrometer, the total ion current can vary significantly. Therefore, adjusting the results by a factor, intrinsic to each MS run, can help to account for this error.

Proteins are aggregated from all their supporting peptides as measured in the selected samples. The TICs for these peptides are normalized using reference peptides, if such were defined, and then summed to determine an overall TIC for the protein. No attempt is made to restrict the peptide set to peptides present in all samples simultaneously. Pooling will increase the summed total ion current if the pooled fractions contain further supporting peptides for the proteins that are quantified. Given normalized TICs for each of the proteins in each of the samples, their relative abundance can be determined by simply selecting the protein in one sample as a reference and dividing the TICs of the protein in the other samples unless the TIC is zero. In that case, the TICs are not divided but reported as calculated.

In both modes, total ion current is used which is similar as calculating protein abundance by the area under the chromatographic peak. This is, however, only possible for studies specifically designed with quantitation in mind. Otherwise, the number of spectra for a peptide or protein will not be enough to establish a good quality elution profile thus the area under the peak cannot be calculated with high precision. The reasoning for incorporating the TIC in the calculation is that the chance for measuring spectra at their highest abundance in the LC profile is greatest under normal acquisition settings such as big-three. Thus, different samples will in general be compared

by spectra from the area of their highest abundance in the corresponding LC profile. Therefore, differences in LC peak height can be captured using the TIC whereas differences in overall abundance or LC peak width are represented by the number of spectra for a peptide/protein.

### Example

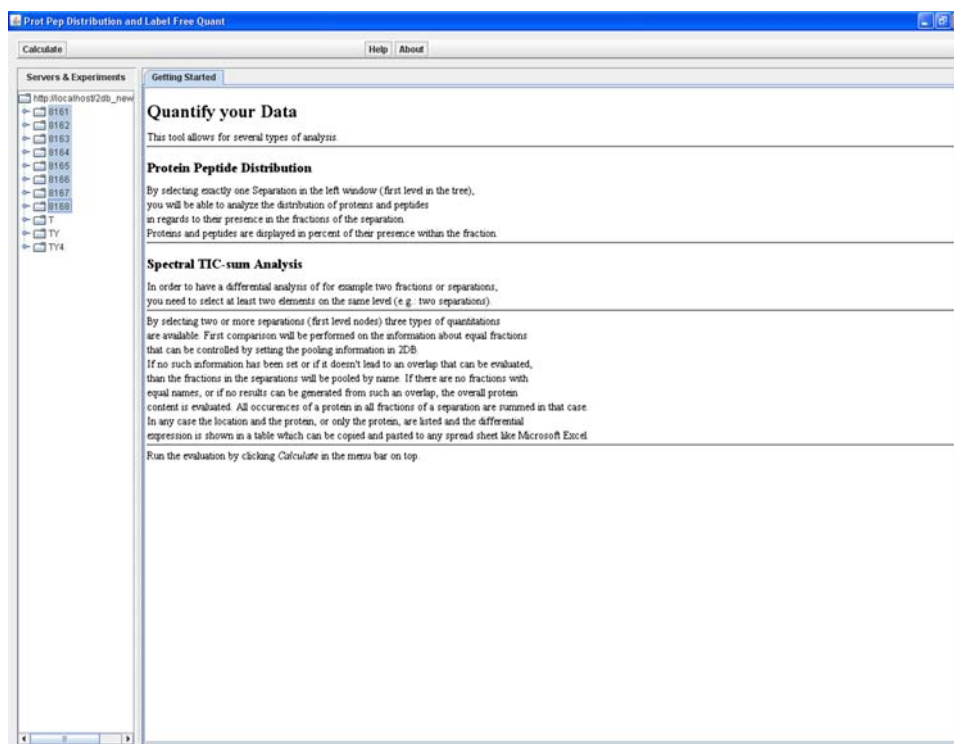
A brief example, which may be reviewed at <http://www.biolink.com/2db>, containing data of an earlier study (Allmer et al. 2006), shall clarify the importance of this extension to 2DB for expression profiling. The theoretical dataset generated in this study is intended to highlight the usage of quantitation with MSMAG. Initially, a view of available experiments within the database is displayed (Fig. 3).

This view of the accessible experiments in the database currently allows for the decision to either perform an expression profiling of one experiment or compare the differential protein expression in several experiments.

Selecting several secondary nodes (here: 8161–8168), brings up another dialog which will aid in analyzing differential protein expression. Selecting one experiment (for example T), displays the protein content per fraction for the experiment which is useful for expression profiling.

The experiment names used here may appear cryptic but more detailed descriptions are available in the online database and in the respective publications.

**Fig. 3** Initial view of MSMAG, the quantitation and distribution extension for 2DB. *On the left* the available experiments are displayed in a tree structure with the all experiments generated in this study marked. *On the right* some quick help is displayed which explains the next steps. Clicking on Calculate in the toolbar with the current selection will for instance bring up the quantitation results for the eight selected experiments



## Expression profiling

As detailed earlier, it is important to know which fractions actually contain an identifiable amount of a protein or peptide for example to measure the associated sample with different parameters. In order to, for example, enrich a protein, its percent distribution over all fractions of an experiment, may be of interest.

Lastly, the information may aid in identifying proteins which may be proteolytically processed (Naumann et al. 2005). Figure 4 shows how MSMAG displays this information.

A list of all proteins and identified peptides, not associated with proteins in the database, is presented in a tree structure on the left of the graphical user interface. If a protein is selected (here the protein with the JGI transcript ID 184490), its distribution over all fractions of the experiment is displayed. Below that view, the distribution of the first of the supporting peptides, ordered alphabetically, is presented unless a different peptide has been chosen from the list of supporting peptides for the protein. A similar percent distribution indicates that the peptide is always supporting the protein identification. Expanding the protein node displays all supporting peptides found in the sample. Selecting one of the peptides displays its percent-wise distribution over the fractions of the sample. This can aid in identifying endogenous proteolytic processing. In some of the fractions a larger amount of peptides supports the occurrence of the protein but in subsequent,

chromatographically distant fractions, subsets of these peptides identify the same protein possible cleavage of the protein into two or more distinct fragments. From the missing peptides it may also be speculated where, approximately, the protein has been cleaved.

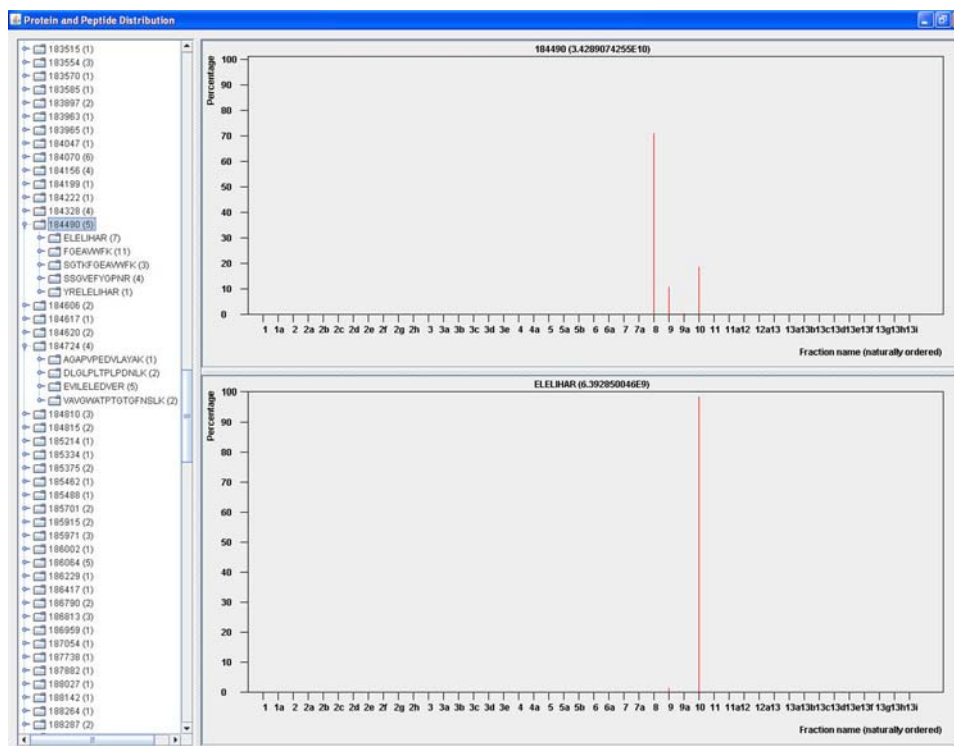
For a more detailed analysis of the results, it is often important to know where on a gel or within which fractions of an LC run a protein is located. This information can shed light on post-translational modifications of a protein if information about, for example, the isoelectric point has been determined during the experiment.

## Label-free quantitation

Selecting at least two experiments in the initial dialog allows for the analysis of differential protein expression in the selected experiments. The dialog shows the selected experiments, all peptides shared among all the experiments, and selected standards (Fig. 5).

If there are peptides shared across all experiments, they can be dedicated as internal standards given that their expression levels are not expected to vary under among experimental conditions. These internal standards can compensate for differences in processing and loading of biological samples. Several peptides can be designated as internal standards. Their average will be subsequently used to normalize the protein expression ratios. On the right of Fig. 5, the results of the quantitation for the experiments 8161–8168 are shown. These results can be copied and

**Fig. 4** An example of a protein distribution for the experiment named T on <http://www.biolnk.com/2db>. The information for the protein with the JGI transcript ID 184490 is expanded and the four supporting peptides, with the third one selected, are visible. On the top right side the protein distribution shows that this protein was found in fractions 7 to 9 with the majority of the protein found in fraction 8. The selected peptide is almost equally distributed over the same range which can be seen in the lower panel on the right



**Fig. 5** The eight theoretical datasets (8161–8168), which can be found on <http://www.biolnk.com/2db>, are used for quantitation by name. Since 8161 is selected, it serves as the reference; all others are shown with respect to their relative difference to 8161. A number of peptides have been selected as internal standards (bottom box on the left). On the right the fractions that have been pooled are spelled out in the first column with all experiments and the applicable fractions as well as the protein identifier are listed. The following columns display the differential protein expression for the selected experiments

pasted to any spread sheet application such as OpenOffice Calc for further analysis.

It is impossible to automatically determine how fractions from different experiments are to be combined. Therefore, four different processing modes are available. Fractions are quantitated, if they are from different experiments and pooling has previously been defined for them, using 2DB. Since this may amount to a large manual workload, it is easier to give the same name to corresponding fractions of different experiments. This is the second mode which will look for fractions with the same name in different experiments. Here pooling will be applied internally if pooling has been set for several fractions of one experiment. These results will be pooled each time when any of the fractions are quantitated with any fraction of another experiment. The third mode of quantitation assumes that the fractions cannot be defined well and therefore simply quantifies the complete protein/peptide content of the experiment with the respective protein/peptide content of the remaining experiments regardless of the membership to a certain fraction. The last mode, *auto*, in MSMAG first tries to find pooled fractions to perform quantitation, if that fails it tries to find fractions that are named equally. If that fails as well, it looks for proteins which are shared among the experiments and quantifies by whole protein/peptide content. The results for the quantitation of experiments 8161–8168 without the selection of an internal standard are available in Table 1.

An internal standard cannot be set due to the theoretical nature of the data which did not take such standards into account. However, in contrast to the expectation, the data significantly benefited from normalization. Therefore, the peptides FILTFNYGNHNYGLNR, GLCALLLLELPPR, LPFALATSSSETVTFQTK, MAVILSLEQGNR, LQVLDDQ ELEHK, LMTLQHDLEMATMEAR, ISFLEGGENNELQ SR, and LMENNLIK were chosen as standards. A closer analysis might have revealed a better suitable set of peptides but the current selection already underlines the usefulness of using standards.

From the 100 different proteins in the dataset, 83 were identified and could be relatively quantified. This is significantly more than that could be expected from an experimental dataset where the number of proteins would be higher and the number of identifications lower. In an experimental dataset, the samples would therefore usually be split into smaller fractions thus reducing the complexity per fraction. This was not necessary in this dataset but should greatly enhance the resolution and thus the result of the experiment. However, the aim was to have a comparable complexity to experimental studies and therefore no further fractionation was performed.

The assumption for the original dataset is that all proteins are represented in equal amounts in all experiments. The model produced by Schulz-Trieglaff et al. is highly accurate (Schulz-Trieglaff et al. 2008) but not perfect which can account for a small fraction of the error that can be deduced from Table 1. However, in the study, random



**Table 1** Results of the quantitation of a simulation of eight high-throughput experiments (accession numbers presented as column headers) using big-three acquisition settings for 100 proteins

| Protein accession/MS-dataset accession | 8161 | 8162   | 8163   | 8164    | 8165     | 8166    | 8167   | 8168   | Average | Deviation |
|--|------|--------|--------|---------|----------|---------|--------|--------|---------|-----------|
| 117655.00                              | 1.00 | 0.50   | 0.00   | 0.28    | 0.00     | 1.10    | 0.00   | 0.00   | 0.36    | 0.46      |
| IPI00107908.1                          | 1.00 | 0.78   | 2.39   | 1.98    | 0.88     | 3.81    | 0.29   | 2.39   | 1.69    | 1.16      |
| IPI00107940.1                          | 1.00 | 1.10   | 0.81   | 2.76    | 1.57     | 2.00    | 2.40   | 0.81   | 1.56    | 0.75      |
| IPI00107941.1                          | 1.00 | 0.20   | 0.18   | 0.00    | 0.43     | 0.46    | 0.20   | 0.18   | 0.33    | 0.31      |
| IPI00107952.2                          | 1.00 | 0.00   | 3.92   | 4.32    | 0.58     | 3.99    | 1.37   | 3.92   | 2.39    | 1.81      |
| IPI00107954.3                          | 1.00 | 270.28 | 0.00   | 442.44  | 31.64    | 367.82  | 68.25  | 0.00   | 147.68  | 183.34    |
| IPI00107962.1                          | 1.00 | 1.34   | 1.83   | 0.53    | 0.48     | 0.88    | 0.48   | 1.83   | 1.05    | 0.57      |
| IPI00107965.1                          | 1.00 | 0.00   | 0.00   | 0.00    | 0.00     | 1611.57 | 0.10   | 0.00   | 201.58  | 569.72    |
| IPI00107975.9                          | 1.00 | 0.24   | 0.89   | 0.69    | 0.42     | 1.02    | 0.40   | 0.89   | 0.69    | 0.30      |
| IPI00107989.1                          | 1.00 | 58.80  | 367.56 | 92.31   | 49.66    | 521.63  | 117.60 | 367.56 | 197.02  | 192.75    |
| IPI00107992.1                          | 1.00 | 1.15   | 0.06   | 0.31    | 0.78     | 1.89    | 0.27   | 0.06   | 0.69    | 0.64      |
| IPI00108004.1                          | 1.00 | 146.11 | 0.02   | 1494.89 | 1,946.53 | 479.16  | 570.71 | 0.02   | 579.81  | 747.47    |
| IPI00108011.1                          | 1.00 | 3.93   | 0.00   | 1.61    | 3.45     | 2.25    | 2.93   | 0.00   | 1.90    | 1.50      |
| IPI00108041.1                          | 1.00 | 0.49   | 2.91   | 2.39    | 0.46     | 2.77    | 0.74   | 2.91   | 1.71    | 1.13      |
| IPI00108058.3                          | 1.00 | 0.66   | 0.19   | 0.86    | 1.49     | 0.22    | 0.07   | 0.19   | 0.59    | 0.50      |
| IPI00108061.3                          | 1.00 | 0.00   | 0.00   | 0.00    | 0.00     | 0.00    | 0.00   | 0.00   | 0.13    | 0.35      |
| IPI00108063.1                          | 1.00 | 11.30  | 13.13  | 11.55   | 11.94    | 4.59    | 4.71   | 13.13  | 8.92    | 4.73      |
| IPI00108066.2                          | 1.00 | 0.51   | 2.77   | 2.03    | 2.14     | 4.10    | 1.54   | 2.77   | 2.11    | 1.13      |
| IPI00108067.3                          | 1.00 | 0.71   | 0.52   | 0.30    | 0.51     | 0.50    | 0.26   | 0.52   | 0.54    | 0.23      |
| IPI00108079.1                          | 1.00 | 10.44  | 2.77   | 7.35    | 5.11     | 4.91    | 1.82   | 2.77   | 4.52    | 3.15      |
| IPI00108083.1                          | 1.00 | 0.00   | 5.17   | 3.21    | 2.41     | 0.95    | 0.64   | 5.17   | 2.32    | 2.03      |
| IPI00108098.1                          | 1.00 | 3.11   | 0.42   | 3.18    | 0.00     | 1.67    | 0.64   | 0.42   | 1.31    | 1.24      |
| IPI00108110.1                          | 1.00 | 0.89   | 0.00   | 0.35    | 0.00     | 0.54    | 0.85   | 0.00   | 0.45    | 0.43      |
| IPI00108111.1                          | 1.00 | 2.60   | 2.08   | 1.83    | 3.37     | 10.13   | 7.86   | 2.08   | 3.87    | 3.29      |
| IPI00108115.3                          | 1.00 | 0.39   | 0.78   | 0.96    | 0.42     | 0.42    | 0.59   | 0.78   | 0.67    | 0.25      |
| IPI00108122.1                          | 1.00 | 4.20   | 23.59  | 17.89   | 2.96     | 37.22   | 24.55  | 23.59  | 16.88   | 12.93     |
| IPI00108135.1                          | 1.00 | 1.58   | 4.10   | 5.07    | 0.93     | 0.00    | 1.62   | 4.10   | 2.30    | 1.85      |
| IPI00108140.1                          | 1.00 | 0.63   | 0.28   | 1.46    | 0.84     | 0.00    | 0.00   | 0.28   | 0.56    | 0.52      |
| IPI00108150.1                          | 1.00 | 0.43   | 0.61   | 0.75    | 0.46     | 0.34    | 0.37   | 0.61   | 0.57    | 0.22      |
| IPI00108151.1                          | 1.00 | 0.55   | 5.67   | 4.10    | 0.43     | 7.11    | 0.66   | 5.67   | 3.15    | 2.78      |
| IPI00108152.2                          | 1.00 | 0.61   | 0.46   | 1.35    | 0.20     | 0.50    | 0.40   | 0.46   | 0.62    | 0.37      |
| IPI00108170.1                          | 1.00 | 0.00   | 0.00   | 0.00    | 0.00     | 0.00    | 0.00   | 0.00   | 0.13    | 0.35      |
| IPI00108171.3                          | 1.00 | 0.50   | 0.64   | 0.45    | 0.31     | 1.07    | 0.59   | 0.64   | 0.65    | 0.26      |
| IPI00108174.2                          | 1.00 | 0.51   | 0.00   | 0.04    | 0.19     | 0.49    | 0.23   | 0.00   | 0.31    | 0.35      |
| IPI00108178.1                          | 1.00 | 2.02   | 1.10   | 1.70    | 2.58     | 0.40    | 1.01   | 1.10   | 1.36    | 0.69      |
| IPI00108184.1                          | 1.00 | 1.49   | 2.35   | 0.89    | 1.54     | 3.69    | 0.85   | 2.35   | 1.77    | 0.98      |
| IPI00108187.1                          | 1.00 | 0.34   | 0.53   | 1.07    | 0.63     | 0.67    | 0.47   | 0.53   | 0.66    | 0.26      |
| IPI00108189.4                          | 1.00 | 0.33   | 0.17   | 1.17    | 0.43     | 1.44    | 0.00   | 0.17   | 0.59    | 0.54      |
| IPI00108190.3                          | 1.00 | 0.79   | 0.90   | 0.98    | 0.81     | 1.30    | 0.53   | 0.90   | 0.90    | 0.22      |
| IPI00108200.2                          | 1.00 | 3.50   | 1.16   | 2.58    | 0.33     | 4.24    | 0.49   | 1.16   | 1.81    | 1.45      |
| IPI00108204.2                          | 1.00 | 0.25   | 0.11   | 0.57    | 0.36     | 1.50    | 0.33   | 0.11   | 0.53    | 0.49      |
| IPI00108206.1                          | 1.00 | 1.44   | 2.03   | 1.77    | 1.46     | 1.98    | 1.55   | 2.03   | 1.66    | 0.36      |
| IPI00108220.4                          | 1.00 | 0.23   | 0.83   | 0.11    | 0.26     | 0.72    | 0.06   | 0.83   | 0.51    | 0.38      |
| IPI00108260.1                          | 1.00 | 0.68   | 0.84   | 2.87    | 1.20     | 1.31    | 0.83   | 0.84   | 1.20    | 0.71      |
| IPI00108270.1                          | 1.00 | 1.99   | 0.00   | 0.71    | 2.04     | 2.54    | 0.28   | 0.00   | 1.07    | 1.00      |
| IPI00108277.5                          | 1.00 | 0.68   | 0.25   | 1.36    | 0.42     | 0.23    | 1.22   | 0.25   | 0.68    | 0.46      |
| IPI00108279.3                          | 1.00 | 0.80   | 5.18   | 1.66    | 1.34     | 2.67    | 1.80   | 5.18   | 2.45    | 1.78      |

**Table 1** continued

| Protein accession/MS-dataset accession | 8161 | 8162   | 8163     | 8164   | 8165   | 8166  | 8167   | 8168     | Average | Deviation |
|--|------|--------|----------|--------|--------|-------|--------|----------|---------|-----------|
| IPI00108287.1                          | 1.00 | 0.00   | 1.59     | 0.47   | 1.39   | 2.39  | 2.50   | 1.59     | 1.37    | 0.87      |
| IPI00108328.3                          | 1.00 | 0.72   | 0.68     | 0.48   | 0.55   | 1.67  | 0.74   | 0.68     | 0.82    | 0.38      |
| IPI00108337.5                          | 1.00 | 0.57   | 0.45     | 0.54   | 0.38   | 0.74  | 0.15   | 0.45     | 0.54    | 0.25      |
| IPI00108367.2                          | 1.00 | 3.13   | 5.14     | 5.78   | 6.46   | 3.54  | 0.25   | 5.14     | 3.81    | 2.25      |
| IPI00108376.1                          | 1.00 | 0.00   | 0.00     | 0.00   | 0.00   | 0.00  | 0.00   | 0.00     | 0.13    | 0.35      |
| IPI00108378.1                          | 1.00 | 0.74   | 1.02     | 0.85   | 0.52   | 1.19  | 0.59   | 1.02     | 0.87    | 0.23      |
| IPI00108389.5                          | 1.00 | 0.37   | 1.47     | 0.78   | 0.67   | 1.01  | 0.62   | 1.47     | 0.92    | 0.40      |
| IPI00108390.6                          | 1.00 | 0.36   | 0.00     | 1.81   | 4.81   | 1.98  | 0.54   | 0.00     | 1.31    | 1.60      |
| IPI00108418.1                          | 1.00 | 2.80   | 7.54     | 2.34   | 7.54   | 5.87  | 7.00   | 7.54     | 5.20    | 2.72      |
| IPI00108426.2                          | 1.00 | 0.56   | 0.73     | 1.67   | 0.85   | 0.92  | 0.65   | 0.73     | 0.89    | 0.35      |
| IPI00108427.5                          | 1.00 | 1.57   | 5.42     | 4.21   | 3.22   | 13.70 | 3.96   | 5.42     | 4.81    | 3.94      |
| IPI00108439.1                          | 1.00 | 1.96   | 2.19     | 3.32   | 2.17   | 3.10  | 3.52   | 2.19     | 2.43    | 0.83      |
| IPI00108444.1                          | 1.00 | 0.50   | 0.08     | 1.26   | 0.28   | 0.30  | 0.60   | 0.08     | 0.51    | 0.43      |
| IPI00108454.2                          | 1.00 | 0.40   | 3.31     | 1.98   | 1.05   | 0.00  | 0.49   | 3.31     | 1.44    | 1.29      |
| IPI00108457.1                          | 1.00 | 3.71   | 1.35     | 4.80   | 1.64   | 7.68  | 1.23   | 1.35     | 2.85    | 2.39      |
| IPI00108475.3                          | 1.00 | 0.72   | 2.10     | 4.15   | 0.26   | 7.24  | 1.83   | 2.10     | 2.43    | 2.28      |
| IPI00108478.1                          | 1.00 | 0.96   | 1.41     | 2.41   | 0.76   | 1.73  | 0.79   | 1.41     | 1.31    | 0.56      |
| IPI00108481.5                          | 1.00 | 285.88 | 1,782.06 | 323.93 | 738.41 | 0.13  | 684.69 | 1,782.06 | 699.77  | 720.59    |
| IPI00108482.1                          | 1.00 | 0.79   | 1.01     | 0.00   | 0.92   | 0.50  | 0.54   | 1.01     | 0.72    | 0.36      |
| IPI00108484.5                          | 1.00 | 0.60   | 0.53     | 1.12   | 0.97   | 0.42  | 0.62   | 0.53     | 0.72    | 0.26      |
| IPI00108492.1                          | 1.00 | 0.69   | 0.29     | 0.00   | 1.49   | 1.99  | 1.08   | 0.29     | 0.85    | 0.67      |
| IPI00108508.4                          | 1.00 | 0.21   | 0.25     | 0.33   | 0.33   | 1.15  | 0.35   | 0.25     | 0.48    | 0.37      |
| IPI00108549.2                          | 1.00 | 0.33   | 0.00     | 0.24   | 0.33   | 0.75  | 0.25   | 0.00     | 0.36    | 0.35      |
| IPI00108565.1                          | 1.00 | 0.74   | 3.28     | 3.44   | 1.29   | 2.89  | 2.66   | 3.28     | 2.32    | 1.12      |
| IPI00108569.1                          | 1.00 | 0.55   | 1.65     | 1.99   | 1.23   | 1.93  | 1.07   | 1.65     | 1.38    | 0.50      |
| IPI00108570.5                          | 1.00 | 6.54   | 0.77     | 1.40   | 1.72   | 1.60  | 4.49   | 0.77     | 2.29    | 2.10      |
| IPI00108579.3                          | 1.00 | 1.47   | 4.30     | 0.70   | 1.94   | 0.00  | 1.28   | 4.30     | 1.87    | 1.60      |
| IPI00108584.2                          | 1.00 | 0.60   | 2.32     | 1.58   | 0.64   | 2.46  | 0.81   | 2.32     | 1.47    | 0.81      |
| IPI00108596.1                          | 1.00 | 0.00   | 0.00     | 1.33   | 2.51   | 3.65  | 0.00   | 0.00     | 1.06    | 1.38      |
| IPI00108597.1                          | 1.00 | 0.00   | 0.31     | 0.00   | 0.00   | 0.00  | 0.00   | 0.31     | 0.20    | 0.35      |
| IPI00108606.1                          | 1.00 | 0.19   | 0.00     | 0.00   | 0.35   | 0.00  | 0.66   | 0.00     | 0.28    | 0.38      |
| IPI00108609.1                          | 1.00 | 1.40   | 2.50     | 1.67   | 2.48   | 2.41  | 0.62   | 2.50     | 1.82    | 0.76      |
| IPI00108610.1                          | 1.00 | 0.00   | 2.89     | 5.74   | 0.00   | 0.00  | 3.26   | 2.89     | 1.97    | 2.08      |
| IPI00108627.1                          | 1.00 | 1.94   | 0.00     | 2.45   | 1.08   | 1.39  | 0.70   | 0.00     | 1.07    | 0.86      |
| IPI00108657.1                          | 1.00 | 0.22   | 1.31     | 0.98   | 1.09   | 1.80  | 0.00   | 1.31     | 0.96    | 0.59      |
| IPI00108663.2                          | 1.00 | 10.41  | 1.32     | 1.52   | 5.53   | 4.03  | 2.07   | 1.32     | 3.40    | 3.25      |

The identified proteins (accession numbers provided as row headers) are given with their relative quantity, normalized to the dataset with the accession number 8161

noise was deliberately added to the ions such that their abundance differs among the datasets. This can account for some of the spread of the data. In this study, additional inherent errors come into effect. Peptides measured in an experiment may not be measured in another experiment. Only 120 out of the 365 possible peptides were present in all datasets which may mostly be due to masking of peptides. This means that the peptide was never among the most abundant three peaks of a full scan for that

experiment or it was, but at a time where MS/MS spectra were measured instead of full scans. Proteins with a large number of detectable peptides such as IPI00108270.1 have a higher chance of being measured and are therefore generally closer to the expected relative difference.

Out of the 83 quantified proteins about 58% are within an average relative difference between 0.5 and 2 for all experiments where the expectation would have been 1. In such a case it would be beneficial to pool the results of

**Table 2** Results of the quantitation of a simulation of eight high-throughput experiments (presented with their accession numbers as column headers) using big-three acquisition settings (3o3) as well as the best 3 out of the first 200 (3o200) for two different proteins

| Protein accession/MS-dataset accession | 8161 | 8162 | 8163 | 8164 | 8165 | 8166 | 8167 | 8168 | Average | Deviation |
|--|------|------|------|------|------|------|------|------|---------|-----------|
| IPI00108270.1 (3o3)                    | 1.00 | 0.37 | 0.84 | 0.88 | 0.50 | 1.02 | 0.53 | 0.84 | 0.75    | 0.25      |
| IPI00108277.5 (3o3)                    | 1.00 | 0.82 | 0.54 | 0.82 | 0.40 | 1.03 | 0.52 | 0.54 | 0.71    | 0.24      |
| IPI00108270.1 (3o200)                  | 1.00 | 0.61 | 1.16 | 1.10 | 0.92 | 1.42 | 0.89 | 1.16 | 1.03    | 0.24      |
| IPI00108277.5 (3o200)                  | 1.00 | 0.94 | 1.11 | 1.46 | 0.70 | 1.89 | 0.99 | 1.11 | 1.15    | 0.37      |

The accession numbers of the identified proteins are given with their relative quantity (presented as the row headers), normalized to the dataset with the accession number 8161

multiple biological replicates to maximize the number of identified peptides and proteins per dataset, and thus increase the resolution and accuracy of the results which was not done for this dataset since no relationship among the datasets was expected.

Note that this method guarantees that data actually support the peptides which cannot be assured for quantification with full scans from high-throughput analyses where each peak in the MS spectrum may be explained by multiple possible peptides from the proteome of the organism further complicated by multiple charge states and possible post-translational modifications as well as different types of noise.

In order to examine the effect of less complex samples, the second experiment which involved just two proteins was analyzed in the same fashion as the experiment detailed above. Both proteins were present in equal amounts in all experiments with the same limitations as pointed out above. The resolution should increase since the mixture is less complex (Table 2).

The deviation and the average of the measurements are significantly better than the average presented in Table 1. This is a little surprising since the re-established LC profile did not meet the expectations. It was most likely not re-established very precisely due to the generation method which introduced a lot of false positive identifications for both generation methods. This can also be seen in Fig. 1 where both re-established LC profiles have multiple peaks which is unexpected. Using the big-three acquisition method this was less pronounced, however, the resolution suffered like the one seen at the left of the highlighted area where the 3o200 method picks up more peptides, completely re-establishing the first LC peak. Results like these with a high number of false positives (all identifications before scan number 700 and those following scan number 800) are not to be expected with an experimental dataset where MS/MS spectra are first mapped to a proteome by software such as Mascot and Sequest, and then filtered and imported into 2DB and finally analyzed by MSMAG.

## Discussion

Often label-free quantitation methods are targeted to a highly specific environment and are not broadly applicable. The quantitation software, Serac, by Old et al. (2005) for instance directly depends on a commercial software package. Other studies directly use commercial software packages for quantitation (Mintz et al. 2008). Census (Park et al. 2008) is a software for quantitation, currently more powerful than the one presented in this study, but with the drawback of working directly from measured data and therefore not as flexible in its application. ProteinQuant (Mann et al. 2008) employs the same strategy. Further, software tools employing the same idea as Serac are reviewed in Mueller et al. (2008). We recently developed 2DB, a database to hold, study, and publish proteomics data based on MS experiments (Allmer et al. 2008). Therefore, the limitation, mentioned above, is not found in MSMAG; the tool presented here can draw from the ability of 2DB to abstract from instrumentation details and its ability to combine results from multiple experiments as well as its ability to integrate findings from multiple MS/MS identification tools. 2DB can hold data generated from high-throughput experiments allowing identification of proteotypic peptides. These peptides identify only one protein from the entirety of proteins of an organism which is essential for precise quantitation. Furthermore, their detectability via MS/MS has been confirmed due to their presence in the database. Other approaches aim to predict such peptides (Sanders et al. 2007), which underlines the importance of knowledge about proteotypic peptides.

The automated quantitation facility, which so far was not available in the 2DB software, has been presented herein. Since it seemed natural to incorporate this functionality, two approaches to label-free quantitation have been incorporated into the 2DB application. The aim of these tools is the identification of peptides and proteins that significantly change in abundance between two conditions. Their results cannot replace more targeted follow-up experiments also including differential labelling strategies

but they greatly simplify identification of potentially interesting proteins and their proteotypic peptides as a bonus in high-throughput studies.

It has been shown that the spectral counting in MS driven proteomics has a higher linearity and the number of quantifiable proteins is increased as compared to peak integration (Wienkoop et al. 2006). Spectral counting, however, makes no use of abundance information present in either the precursor ion or the MS/MS spectrum. This is why spectral counts are used in this study but are further amended with total ion current data in order to exploit abundance information. Another limitation to current techniques is that they usually do not employ internal standards as done explicitly in a study by Tabata et al. (2007) and implicitly by Naumann et al. (2007). Internal standards, if known, can be utilized in the tool developed in this study. Tabata et al. (2007) used labelled standards in their study. MSMAG aims to keep the complete process label-free and does not employ labelling for the standards; however, it enables the possibility of designating several peptides as standards as proposed in other areas where quantification is important for example during a real time polymerase chain reaction experiment (Vandesompele et al. 2002). MSMAG, presented here, enables investigation of protein expression profiles, and presents a new method for label-free quantitation.

## Conclusion

An extension to 2DB is presented which allows for thorough investigation of protein and peptide occurrence within an experiment to investigate endogenous proteolytic processing and to pool fractions with similar protein content as well as to determine proteotypic peptides. The software presented is furthermore suitable for label-free quantitation for any number of experiments thus also allowing analysis of time series which is important for the investigation of protein expression dynamics (Mintz et al. 2008). The robustness of the results can be increased by providing peptides that can serve as standards. A new quantitation method is introduced which consists of spectral counting amended with abundance information derived from the total ion current of the supporting spectra.

## Outlook

Currently, only label-free quantitation is possible with the newly developed extension to 2DB. The database is, however, able to hold information about post-translational modifications and is therefore well-equipped to represent modified or labelled peptides. In future, another mode of operation will involve quantitation for any given

modification to a peptide, either a label or a PTM. The relative abundance of PTMs per protein and per fraction as well as differential expression of PTM patterns is another analysis that will be available in the near future which will aid in biomarker discovery (Wei and Li 2009).

No statistical analysis of this method has been performed since data were believed to be of low reproducibility and since no suitable benchmark dataset was available. In light of evidence pointing towards good reproducibility of label-free quantitation (Griffiths et al. 2001; Stevenson et al. 2008; Zhang et al. 2006), such analyses are indicated and statistics will be presented along with suitable benchmark quantitation data in the future. Especially, the new label-free quantitation method, using spectral counts and TIC, will be examined more closely with regard to its statistical confidence in the future.

**Acknowledgments** I would like to thank Ritchie Eanes for proof-reading and additional valuable comments on the paper. I am thankful to Ole Schulz-Trieglaff for providing the underlying datasets from their paper and additional clarifying information.

## References

- Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207
- Allmer J, Markert C, Stauber EJ et al (2004) A new approach that allows identification of intron-split peptides from mass spectrometric data in genomic databases. *FEBS Lett* 562:202–206
- Allmer J, Naumann B, Markert C et al (2006) Mass spectrometric genomic data mining: novel insights into bioenergetic pathways in *Chlamydomonas reinhardtii*. *Proteomics* 6:6207–6220
- Allmer J, Kuhlert S, Hippler M (2008) 2DB: a Proteomics database for storage, analysis, presentation, and retrieval of information from mass spectrometric experiments. *BMC Bioinformatics* 9:302
- America AH, Cordewener JH (2008) Comparative LC-MS: a landscape of peaks and valleys. *Proteomics* 8:731–749
- Bafna V, Edwards N (2001) SCOPE: a probabilistic model for scoring tandem mass spectra against a peptide database. *Bioinformatics* 17(Suppl 1):S13–S21
- Bantscheff M, Schirle M, Sweetman G et al (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389:1017–1031
- Blagoev B, Ong SE, Kratchmarova I et al (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22:1139–1145
- Cravatt BF, Simon GM, Yates JR 3rd (2007) The biological impact of mass-spectrometry-based proteomics. *Nature* 450:991–1000
- Duncan DT, Craig R, Link AJ (2005) Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. *J Proteome Res* 4:1842–1847
- Elias JE, Haas W, Faherty BK et al (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2:667–675
- Eng J, McCormack AL, Yates JR III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989
- Gao J, Opitck GJ, Friedrichs MS et al (2003) Changes in the protein expression of yeast as a function of carbon source. *J Proteome Res* 2:643–649

- Geer LY, Markey SP, Kowalak JA et al (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3:958–964
- Griffiths WJ, Jonsson AP, Liu S et al (2001) Electrospray and tandem mass spectrometry in biochemistry. *Biochem J* 355:545–561
- Gygi SP, Rist B, Gerber S et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17:994–999
- Higgs RE, Knierman MD, Gelfanova V et al (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. *J Proteome Res* 4:1442–1450
- Hoehenwarter W, van Dongen JT, Wienkoop S et al (2008) A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. *Proteomics* 8:4214–4225
- Krijgsveld J, Ketting RF, Mahmoudi T et al (2003) Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* 21:927–931
- Mann M, Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66:4390–4399
- Mann B, Madera M, Sheng Q et al (2008) ProteinQuant Suite: a bundle of automated software tools for label-free quantitative proteomics. *Rapid Commun Mass Spectrom* 22:3823–3834
- Mintz M, Vanderver A, Brown KJ et al (2008) Time series proteome profiling to study endoplasmic reticulum stress response. *J Proteome Res* 7:2435–2444
- Mueller LN, Brusniak MY, Mani DR et al (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* 7:51–61
- Naumann B, Stauber EJ, Busch A et al (2005) N-terminal processing of Lhca3 is a key step in remodeling of the photosystem I-light-harvesting complex under iron deficiency in *Chlamydomonas reinhardtii*. *J Biol Chem* 280:20431–20441
- Naumann B, Busch A, Allmer J et al (2007) Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. *Proteomics* 7:3964–3979
- Old WM, Meyer-Arendt K, Aveline-Wolf L et al (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4:1487–1502
- Pan J, Chen HQ, Sun YH et al (2008) Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology. *Lung* 186:255–261
- Pang JX, Ginanni N, Dongre AR et al (2002) Biomarker discovery in urine by proteomics. *J Proteome Res* 1:161–169
- Park SK, Venable JD, Xu T et al (2008) A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 5:319–322
- Perkins DN, Pappin DJ, Creasy DM et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567
- Roddy TP, Horvath CR, Stout SJ et al (2007) Mass spectrometric techniques for label-free high-throughput screening in drug discovery. *Anal Chem* 79:8207–8213
- Sanders WS, Bridges SM, McCarthy FM et al (2007) Prediction of peptides observable by mass spectrometry applied at the experimental set level. *BMC Bioinformatics* 8(Suppl 7):23
- Schulz-Trieglaff O, Pfeifer N, Gropl C et al (2008) LC-MSSim—a simulation software for liquid chromatography mass spectrometry data. *BMC Bioinformatics* 9:423
- Shevchenko A, Jensen ON, Podtelejnikov AV et al (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci USA* 93:14440–14445
- Silva JC, Denny R, Dorschel CA et al (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* 77:2187–2200
- Silva JC, Denny R, Dorschel C et al (2006) Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* 5:589–607
- Stevenson SE, Chu Y, Ozias-Akins P et al (2008) Validation of gel-free, label-free quantitative proteomics approaches: applications for seed allergen profiling. *J Proteomics* 72(3):555–566
- Tabata T, Sato T, Kuromitsu J et al (2007) Pseudo internal standard approach for label-free quantitative proteomics. *Anal Chem* 79:8440–8445
- Tabb DL, Saraf A, Yates JR 3rd (2003) GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 75:6415–6421
- Vandesompele J, De Preter K, Pattyn F et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034
- Vissers JP, Pons S, Hulin A et al (2008) The use of proteome similarity for the qualitative and quantitative profiling of perfused myocardium. *J Chromatogr B Analyt Technol Biomed Life Sci* 877(13):1317–1326
- Wang F, Ye M, Dong J et al (2008) Improvement of performance in label-free quantitative proteome analysis with monolithic electrospray ionization emitter. *J Sep Sci* 31:2589–2597
- Wei X, Li L (2009) Mass spectrometry-based proteomics and peptidomics for biomarker discovery in neurodegenerative diseases. *Int J Clin Exp Pathol* 2:132–148
- Wienkoop S, Larrainzar E, Niemann M et al (2006) Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics. *J Sep Sci* 29:2793–2801
- Yang F, Jaitly N, Jayachandran H et al (2007) Applying a targeted label-free approach using LC-MS AMT tags to evaluate changes in protein phosphorylation following phosphatase inhibition. *J Proteome Res* 6:4489–4497
- Zhang B, Verberkmoes NC, Langston MA et al (2006) Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* 5:2909–2918