Find Pairs: The Module for Protein Quantification of the PeakQuant Software Suite

Martin Eisenacher,¹ Michael Kohl,¹ Sebastian Wiese,² Romano Hebel,^{1,3}
Helmut E. Meyer,¹ Bettina Warscheid,^{2,4} and Christian Stephan¹

Abstract

Accurate quantification of proteins is one of the major tasks in current proteomics research. To address this issue, a wide range of stable isotope labeling techniques have been developed, allowing one to quantitatively study thousands of proteins by means of mass spectrometry. In this article, the FindPairs module of the PeakQuant software suite is detailed. It facilitates the automatic determination of protein abundance ratios based on the automated analysis of stable isotope-coded mass spectrometric data. Furthermore, it implements statistical methods to determine outliers due to biological as well as technical variance of proteome data obtained in replicate experiments. This provides an important means to evaluate the significance in obtained protein expression data. For demonstrating the high applicability of FindPairs, we focused on the quantitative analysis of proteome data acquired in $^{14}\text{N}/^{15}\text{N}$ labeling experiments. We further provide a comprehensive overview of the features of the FindPairs software, and compare these with existing quantification packages. The software presented here supports a wide range of proteomics applications, allowing one to quantitatively assess data derived from different stable isotope labeling approaches, such as $^{14}\text{N}/^{15}\text{N}$ labeling, SILAC, and iTRAQ. The software is publicly available at <http://www.medizinisches-proteom-center.de/software> and free for academic use.

Introduction

IN RECENT YEARS, DIFFERENT STABLE ISOTOPE labeling methods combined with mass spectrometry (MS) have been successfully applied to the relative quantification of proteins in complex biological systems (Kierszniowska et al., 2009; Munday et al., 2010; Skirycz et al., 2011; Soufi et al., 2010; Thorn and Orians, 2011; for review see Bantscheff et al., 2007; Ong and Mann, 2005). Two basic labeling strategies can be followed in such MS-based quantitative proteomics endeavors: Metabolic labeling approaches (Beynon and Pratt, 2005) rely on the *in vivo* incorporation of stable isotopes into proteins during their synthesis by using either “heavy” versions of certain amino acids (generally referred to as Stable Isotope Labeling with Amino Acids in Cell Culture, or SILAC; Ong et al., 2002), or of chemical elements (e.g., $^{14}\text{N}/^{15}\text{N}$ labeling). In contrast, *in vitro* techniques based on chemical labeling of intact proteins (e.g., ICAT; Gygi et al., 1999), or peptides (e.g., iTRAQ; Ross et al., 2004), involve covalent modification of amino acid side chains or peptide termini with stable isotope-coded reagents. Following differential labeling and mixing of samples, the “light” and “heavy” forms of one peptide (same

charge state, same sequence, and same modifications) are observed as mass peak pairs in the MS spectra measured. In isobaric tagging methods such as iTRAQ, specific reporter ions are released during peptide fragmentation and can be observed in MS/MS spectra at distinct mass-to-charge (m/z) values.

In chemical labeling approaches as well as SILAC, peptides show peak pairs with accurately defined additive mass difference (“additive shift”) in MS scans. A distinct feature of metabolic labeling using heavy elements (e.g., ^{15}N -salts) is that peak pairs of isotope-coded peptides show a variable mass difference dependent on the number of times the heavy element occurs in the respective peptide. For the subsequent accurate determination of protein abundance ratios, peaks of light and heavy peptide forms have to be correctly paired to be relatively quantified. Then peptide abundance ratios have to be summarized into protein ratios.

In $^{14}\text{N}/^{15}\text{N}$ labeling experiments, however, this process is usually complicated by two facts. First, *in vivo* incorporation of ^{15}N is usually not complete due to use of ^{15}N salts with 95–98% purity, resulting in more complex isotopic peak patterns of heavy-labeled peptides, and second, $^{14}\text{N}/^{15}\text{N}$ -labeled

¹Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany.

²Funktionelle Proteomik, Fakultät für Biologie and BIOS Centre for Biological Signalling Studies, Universität Freiburg, Freiburg, Germany.

³Brüker Daltonik GmbH, Bremen, Germany.

⁴Zentrum für Biosystemanalyse (ZBSA), Universität Freiburg, Freiburg, Germany.

peptides feature peak pairs with sequence-dependent, variable mass shifts. An automated quantification algorithm applicable to $^{14}\text{N}/^{15}\text{N}$ labeling must therefore account for both complex isotopic peak patterns (IPP) and variable mass shifts.

In this work, an advanced bioinformatics approach, called "FindPairs," is presented. FindPairs is designed as a generic algorithm aimed at automated quantitative analysis of isotope-coded mass spectra with high accuracy and reliability. It is also rather flexible, as most algorithm steps can be controlled by user-specifiable parameters. FindPairs is part of the PeakQuant software suite, which serves as an integrated platform for several proteomics tools and provides an easy-to-operate graphical user interface. Application of FindPairs is demonstrated here with particular attention to studies using $^{14}\text{N}/^{15}\text{N}$ labeling. However, the software is also applicable to the quantitative analysis of proteomics experiments employing SILAC or iTRAQ, for example. Finally, features of FindPairs are compared to those of similar software solutions aimed at accurate protein quantification.

Methods and Algorithms

In this section, the basic concepts behind the FindPairs algorithm are presented. These correspond to algorithm parameters grouped on the "Find Pairs" tab of the "Configuration / General" dialog or in the "FindPairs" dialog itself. The exact parameter names are mentioned in brackets.

In stable isotope-labeling experiments, two or more differently-labeled samples are mixed in an equal ratio for subsequent LC/MS analysis. So interpretation of spectra from stable isotope labeling experiments is mainly based on mass shift detection. The mass shift is the expected distance between the monoisotopic peaks of a light-labeled peptide and its heavy-labeled counterpart. Different labeling techniques introduce different mass differences:

- ICAT: Isotope-Coded Affinity Tags: e.g., 8 Da (Gygi et al., 1999)
- ALICE: Acid-Labile Isotope-Coded Extractants: 1 Da (Qiu et al., 2002)
- 2-Methoxy-4,5-dihydro-1 H-imidazole: 4 Da (Peters et al., 2001)
- MCAT: Mass-Coded Abundance Tagging: 42 Da (Cagney and Emili, 2002)
- GIST: Global Internal Standard Technology: 4 Da (Chakraborty and Regnier, 2002)
- $^{18}\text{O}/^{16}\text{O}$: 4 Da (Schnolzer et al., 1996; Yao et al., 2001)
- ICPL: Isotope Coded Protein Label: 6 Da (Schmidt et al., 2005)
- iTRAQ: isobaric Tag for Relative and Absolute protein Quantitation, reporter ions in MS/MS spectra: e.g., 114, 115, 116, and 117 Da (Ross et al., 2004)
- TMT: Tandem Mass Tags, reporter ions in MS/MS spectra (Thompson et al., 2003)
- SILAC: Stable Isotope Labeling with Amino Acids in Cell Culture, e.g., with deuterated leucine (Leu-d3) (Foster et al., 2003; Ong et al., 2002); or with heavy lysine and/or arginine (Andersen et al., 2005; Ibarrola et al., 2003); or with heavy methionine (Ong et al., 2004)
- $^{14}\text{N}/^{15}\text{N}$: variable additive mass shift; it can be estimated by a database-averaged multiplicative mass shift or mass shift factor (MSF): 1.012198 (Oda et al., 1999; Skirycz et al., 2011). MSF values can be used to detect

peak pairs without any knowledge about the protein sequence.

In principle, peak pairs can be detected without using sequence information. However, as the actual aim is usually the quantification of proteins, sequence information is used whenever available. It also offers the opportunity to detect only pairs belonging to "identified" or "high scoring" peptides, and thus lowers the number of false-positively detected peptide pairs. Therefore, FindPairs can be used either without sequence information or with sequence information from peptide or protein identifications of search engine runs.

In the following paragraphs of this section, first the basic idea of detecting peak pairs in spectral data with and without sequence information is shown. This detection is separately described for heavy amino acid labeling (i.e., SILAC, detection of an additive mass shift) and heavy element labeling by ^{15}N isotopes (i.e., detection of variable mass shifts and estimation by multiplicative MSF), as these approaches differ fundamentally. After that, an improved method for peak pair detection in $^{14}\text{N}/^{15}\text{N}$ labeling experiments is shown, making use of the complete IPP. Then the actual peptide quantification using the IPP is described. Preliminary data processing and data import for the FindPairs module is sketched. After that, the methods for the detection of false-positives and outliers are presented and the calculation of protein abundance ratios is described. Finally, PeakQuant's statistical possibilities for comparison of multiple analyses are briefly discussed.

Peak pair detection in stable isotope labeling experiments

In order to determine relative abundances of proteins from differentially-labeled samples, at first isotope-coded peptide peak pairs have to be uniformly detected in the acquired MS spectra. Then abundance ratios based on the peak intensities can be calculated for each peak pair. An apparent way for detecting the peaks of light peptides together with their heavy counterparts is to search all possible peak pair combinations in MS survey spectra and to extract those fulfilling the mass shift criterion: The mass shift between light and heavy peaks is either a constant (additive) mass difference, or a variable mass shift, that can be estimated by a (multiplicative) MSF. Alternatively, amino acid sequence information can be used for accurately detecting peak pairs.

FindPairs without sequence information

Peak pairs in experiments incorporating heavy amino acids show a constant mass difference. In all stable isotope-labeling experiments that feature such a constant mass shift, the molecular mass of the heavy peptide can be determined by

$$m^{(\text{heavy pep})} = m^{(\text{light pep})} + \Delta M, \quad (1)$$

where $m^{(\text{heavy pep})}$ and $m^{(\text{light pep})}$ are the masses of peptides comprising the heavy and light label, respectively, and ΔM is a constant mass difference, depending on the labeling technique used.

In contrast, when proteins are labeled *in vivo* by incorporating heavy elements such as ^{15}N , the molecular mass of the heavy peptide can be calculated by

$$m^{(\text{heavy pep})} = m^{(\text{light pep})} * \text{MSF}_{\text{pep}}, \quad (2)$$

where MSF_{pep} is the MSF of peptide pep . MSF_{pep} depends on the number of isotopes ($=: n_{ei}$) that were exchanged within pep . Without sequence information for proteins / peptides, n_{ei} and thus MSF_{pep} are unknown. In this case, FindPairs uses an estimated value MSF_{DB} , or $MSF_{species}$, determined in advance as the average MSF for a distinct protein sequence database DB, or the subset of sequences for a specific species.

For each MS level 1 line spectrum (also known as “centroid spectrum,” “stick spectrum,” or “peak list”), the FindPairs algorithm starts at the peak l with the lowest m/z . This peak is considered as a potential light peak. Next, the theoretical mass of the corresponding heavy peak is then computed by applying [Eq. 1] or [Eq. 2], depending on the stable-isotope labeling strategy applied. In the latter case proton masses have to be subtracted before and re-added after the calculation (as MSF_{DB} or $MSF_{species}$ have been calculated using molecular masses of peptides, without charge). If a matching heavy peak can be found at the calculated theoretical mass (within a certain mass tolerance, parameter “mass tolerance for mono-isotopic peak”), both the light and heavy peaks are considered as a matching peak pair. The algorithm then continues with the peak next to peak l (in the direction of increasing m/z).

FindPairs with sequence information

If amino acid sequence information is at hand, n_{ei} is known, and thus MSF_{pep} can be exactly calculated even in cases of heavy element labeling. Applying the correct MSF_{pep} values generally reduces the number of false-positively matched peptide pairs. Consequently it lowers the error in relative peptide / protein quantification. To this end, FindPairs performs an *in silico* digestion of the provided protein sequences (parameters in the “Configuration” and “Modifications” dialogs, and on the “Digest/Mod” tab; especially select the “ ^{15}N labeling” check box!). For each peptide resulting from the simulated digest, theoretical light and heavy peak masses are calculated and used for the detection of peak pairs as described above.

Estimation of the MSF for $^{14}N/^{15}N$ labeling experiments

Metabolic labeling of proteins by incorporation of ^{15}N leads to a variable mass shift between peak pairs. The actual mass shift for a given peptide depends on the number of nitrogen atoms comprised in its amino acid sequence. The number of nitrogen atoms in each proteinogenic amino acid and the resulting mass shifts are shown in Table 1. If no sequence information is available, the MSF_{pep} value for each peptide pep can be estimated by an average value for a complete sequence database (MSF_{DB}), or a species-specific sub-database $MSF_{species}$. MSF_{DB} or $MSF_{species}$ provides a better estimate for MSF_{pep} ; the more similar the amino acid distribution of pep is to that of the complete or species-specific sub-database, the better. This implies that the estimation is asymptotically better for longer peptides. A specific tool (“ ^{15}N shift calculation”) integrated in the PeakQuant suite performs calculation of MSF_{DB} or $MSF_{species}$ by processing FASTA-formatted protein databases (Lipman and Pearson, 1985).

The relation between the masses of the light ^{14}N and heavy ^{15}N versions of an amino acid aa can be expressed by:

$$m(^{heavy}aa) = m(^{light}aa) + \#N(aa) \cdot (m(^{15}N) - m(^{14}N)) \quad (3)$$

where $\#N(aa)$ is the number of nitrogens in amino acid aa .

Let the total mass of all amino acids in a database DB be $totmass_{DB} := \sum_{k=1}^{21} aa_k \cdot m(^{light}aa_k)$, where $|aa_k|$ denotes the absolute number of amino acid aa_k in DB (with Selenocystein there may occur 21 amino acids in protein databases). Then we define the MSF_{DB} of database DB (and $MSF_{species}$ analogously) by:

$$MSF_{DB} = \sum_{j=1}^{21} \frac{m(^{heavy}aa_j)}{m(^{light}aa_j)} \cdot \frac{aa_j \cdot m(^{light}aa_j)}{totmass_{DB}}, \quad (4)$$

where $m(^{heavy/light}aa_j)$ is the mass of the respective isotopic variant of amino acid j . [Eq. 4] thus corresponds to an average mass shift weighted by amino acid-specific total mass proportions.

Calculating MSF_{DB} as described above is only a good estimation for MSF_{pep} , if the sequence database digested contains a sufficient representation of the protein composition of an organism. The MSF_{NCBI} of the NCBI database as of June 1, 2011 is 1.012198 (Table 1). Taxonomy-specific factors $MSF_{species}$ for taxonomy specific sub-databases further improve the estimation of the correct MSF (Table 2).

Detection of the isotopic peak pattern in $^{14}N/^{15}N$ labeling experiments

In $^{14}N/^{15}N$ labeling experiments, reliable detection of peak pairs in MS survey spectra is a challenging endeavor. The reliability of peak pair detection can be improved by considering information about the complete IPP. Figure 1 shows the IPP of the doubly-charged peptide VYGPHFASPK in an MS level 1 line spectrum. Besides the ^{14}N and ^{15}N monoisotopic peaks, there are isotopic peaks with higher masses representing peptides with naturally-occurring ^{13}C , called “ ^{13}C peaks” (other naturally-occurring isotopes like ^{18}O , ^{15}N , and 2H are very rare and negligible here). The peaks with lower m/z than the monoisotopic ^{15}N peak are termed “ ^{15}N satellites.” They result from incomplete integration of ^{15}N into proteins during the labeling process (^{14}N is integrated instead). This is caused by the available ^{15}N reagents having a typical purity of 95–98% (Andreev et al., 2006; Gouw et al., 2008; Hebel et al., 2008). FindPairs efficiently detects IPPs in MS level 1 line spectra in linear computation time depending on input size (mass tolerance for isotopic peaks: parameter “mass tolerance for other isotopic peaks”).

Peptide quantification using the isotopic peak pattern in $^{14}N/^{15}N$ labeling experiments

The intensities of IPP peaks should be utilized for accurate peptide quantification. High-abundant peptides usually will display more isotopic peaks in their measured IPP than lower-abundant peptides (although the theoretical isotopic peak distribution of a protonated peptide mainly depends on its m/z value and abundance). Accordingly, for different peptides, the number of isotopic peaks observable may vary, which has an influence on relative quantification. To address this issue, FindPairs can be configured to consider all detectable peaks of the IPP for quantification or to enclose only peak pairs with at least or exactly a specified number of ^{13}C peaks

TABLE 1. CALCULATION OF THE $^{14}\text{N}/^{15}\text{N}$ MASS SHIFT FACTOR FOR THE NCBI_{NR} DATABASE ($MSF_{\text{NCBI}_{\text{NR}}}$)

Amino Acid	No. of N atoms	^{14}N mass (Da)	^{15}N mass (Da)	Relative mass shift	Absolute number of occurrences	Absolute ^{14}N mass (Da)	Resulting mass shift
A	1	71.04	72.03	1.014035	416,181,307	29,564,317,285.30	0.055375
C	1	103.01	104.01	1.009679	66,156,824	6,814,760,853.21	0.012710
D	1	115.03	116.02	1.008668	259,750,066	29,878,255,256.78	0.055667
E	1	129.04	130.04	1.007726	302,543,016	39,040,934,371.05	0.072670
F	1	147.07	148.07	1.006779	194,915,706	28,665,942,965.45	0.053308
G	1	57.02	58.02	1.017485	345,351,755	19,692,461,283.66	0.037010
H	3	137.06	140.05	1.021823	108,827,666	14,915,801,279.80	0.028153
I	1	113.08	114.08	1.008817	284,204,338	32,138,980,410.65	0.059888
K	2	128.09	130.09	1.015567	258,273,991	33,083,596,546.19	0.062061
L	1	113.08	114.08	1.008817	479,064,445	54,174,552,442.25	0.100949
M	1	131.04	132.04	1.007609	117,630,794	15,414,396,884.85	0.028689
N	2	114.04	116.04	1.017485	201,111,265	22,935,317,916.61	0.043105
P	1	97.05	98.05	1.010273	238,213,785	23,119,305,304.30	0.043143
Q	2	128.06	130.05	1.015572	191,227,245	24,488,289,452.01	0.045937
R	4	156.10	160.09	1.025548	269,718,239	42,103,316,495.15	0.079757
S	1	87.03	88.03	1.011456	337,721,215	29,392,562,915.52	0.054914
T	1	101.05	102.04	1.009867	273,748,005	27,661,600,809.88	0.051598
U	1	150.95	151.95	1.006605	2132	321,833.16	0.000001
V	1	99.07	100.07	1.010064	325,864,101	32,282,838,362.15	0.060230
W	2	186.08	188.07	1.010716	63,807,328	11,873,223,567.18	0.022166
Y	1	163.06	164.06	1.006114	148,056,815	24,142,637,283.09	0.044867

Used monoisotopic element masses are from <http://www.ionsource.com/Card/Mass/mass.htm> (which is assembled from Audi and Wapstra, 1995; Coplen, 2001; Rosman and Taylor, 1997). Absolute occurrences of amino acids were derived from the NCBI_{NR} FASTA file as of June 1, 2011 (14,261,927 sequence entries). The letter U denotes Selenocystein. “Absolute ^{14}N mass” is the “ ^{14}N mass” times the “Absolute number of occurrences.” The “Resulting mass shift” represents the contribution of a particular amino acid to the observed mass shift between light and heavy versions of an (arbitrary) peptide (without knowledge of its sequence). The mass shift factor for the whole database equals the sum of the resulting mass shifts of all amino acids: $MSF_{\text{NCBI}_{\text{NR}}} = 1.012198$. (Columns 3, 4, and 7 are rounded to two decimals. Columns 5 and 8 are rounded to six decimals to allow the influence of MSFs on three decimals of peak masses around 1000 Da).

and/or ^{15}N satellites (parameter group “consider other isotopic peaks”).

However, the fewer isotopic peaks are required by the parameter setting to form a valid IPP, the more (potentially false-positive) peptide pairs are typically detected, especially in spectra with many noise peaks and low signal-to-noise ratios. As first choice we recommend to consider monoisotopic peaks plus exactly two ^{13}C peaks (for each of the monoisotopic peaks), and exactly one ^{15}N satellite. A more stringent approach is to consider three ^{13}C peaks and two ^{15}N satellites, which may further reduce false-positives. As a typical trade-off also some true-positives may be lost, and the optimal parameter setting depends on the data set i.e., the sample and the actual measurement workflow.

Sometimes the charge states of the peaks in MS survey spectra are not available in total or simply not stored in the spectrum file format used. Therefore, in the “without sequence” case, target charge states can be specified (parameter group “charge states to iterate, if not available in spectra”). In

cases for which the spectrum file format contains charge state information, FindPairs can be restricted to detect a peak pair only from peaks having the same charge state (parameter “build pairs only from peaks with same charge state”). In the “with sequence” case, charge states follow implicitly from the protein digestion parameters (“Digest/Mod” tab).

By default, the quantification values of peak pairs with different charge states are displayed separately in the result tables. They can be merged by selecting the parameter check box “merge pairs with different charge state constellations.” In case of a merge, the intensities of the IPP peaks are summed and the ratio is recalculated for this “merged IPP.”

Processing of MS data and data import

FindPairs can make use of time-dependent extraction of MS level 1 line spectra obtained from LC/MS analyses. If the same peptide is present in two consecutive time segments due to its elution profile from the chromatographic column, it will

TABLE 2. OVERVIEW OF THE $^{14}\text{N}/^{15}\text{N}$ MASS SHIFT FACTORS FOR DIFFERENT ORGANISMS

Organism	Thale Cress	Human	Mouse	Rat
Organism	IPi protein DB	Number of proteins	$^{14}\text{N}/^{15}\text{N}$ mass shift factor	
<i>Arabidopsis thaliana</i>	ARATH v3.82	38,515	1.012162	
<i>Homo sapiens</i>	HUMAN v3.84	90,166	1.012341	
<i>Mus musculus</i>	MOUSE v3.84	59,995	1.012312	
<i>Rattus norvegicus</i>	RAT v3.84	39,928	1.012287	

From IPI databases as of June 1, 2011. The arithmetic mean of the mass shift factor for all organisms shown is 1.012276.

FIG. 1. Zoom-in region showing the isotopic peak pattern (IPP) of the $^{14}\text{N}/^{15}\text{N}$ -labeled doubly-charged peptide VYGFHFASPK (screen shot slightly edited for better readability of mass labels). The left part of the IPP represents the ^{14}N peptide, while the right part represents the ^{15}N peptide. Besides the monoisotopic peaks observed at 551.84 Da (^{14}N) and 558.32 Da (^{15}N), ^{13}C isotopic peaks and ^{15}N satellite peaks are present as indicated (refer to text for details).

appear in consecutive spectra of the data set. Utilizing this fact can help to avoid the merging of unrelated IPPs: FindPairs allows for limiting the condensation of quantification values to only those peptides for which IPPs were detected in adjacent spectra (parameter “only merge pairs from consecutive spectra”). In addition, several pre-processing steps can be performed to enhance the quality of the peak pair detection and quantification process. Among others, this includes application of noise filters and the removal of user-defined contaminants (“Filter” tab of the “Configuration / General” dialog).

The PeakQuant suite supports various MS data formats for importing peak lists (Table 3). It allows opening several thousands of MS spectra simultaneously for automated batch processing.

A typical workflow for processing data generated by LTQ-Orbitrap instruments (Thermo Scientific, Waltham, MA, USA) or QStar instruments (formerly Applied Biosystems, now Life Technologies, Carlsbad, CA, USA) is: (1) extract the MS level 1 spectra from the .raw/.wiff file, (2) apply the manufacturer’s peak picking algorithm (usually “centroid”) to obtain line spectra/peak lists still containing the isotopic peaks, and (3) convert these into a mzXML file. This can be done, for example, by using ProteoWizard (Kessner et al., 2008; <http://proteowizard.sourceforge.net/>) with the command: `msconvert <name>.[raw|wiff] --mzXML --filter ‘‘msLevel 1’’ --filter ‘‘threshold absolute`

`0.1 most intense’’ --filter ‘‘peakPicking true 1’’` (vendor-reader support version of ProteoWizard is necessary).

For the tune mix and the exemplary data files acquired on a QStar instrument shown in this work (see also <http://www.medizinisches-proteom-center.de/PeakQuant>), peak picking via ProteoWizard results in multiple line peaks per profile peak, probably due to insufficient smoothing. These multiple peaks are subsequently merged by the “merge masses” filter of PeakQuant (mass tolerance: 0.1 Da). Noise should be filtered in the “without sequence” case (e.g., by using a threshold of 1% or 2% of the spectrum’s most intense peak).

The configuration of the general properties and the modifications can be saved in and reloaded from *.cfg files (see <http://www.medizinisches-proteom-center.de/PeakQuant> for the configuration used to produce the examples). After loading properties, modifications, and spectra data, the FindPairs module has to be started. Then the automatic detection of peak pairs is initiated by clicking the “find pairs” button on the “Pairs of Spectra” tab. The identified peptide pairs are reported as shown in Figure 2 (one table per spectrum).

To be able to choose optimal values for the manifold parameters controlling the FindPairs algorithm, a “tuning data set” with known properties or results is strongly recommended. Such a standard data set can also be used to estimate the accuracy of the FindPairs quantification workflow. In a “1:1 mixture” data set in the work by Hebel et al. (2008), the

TABLE 3. SUPPORTED FILE FORMATS FOR IMPORTING MS AND MS/MS DATA INTO THE PEAKQUANT SUITE

<i>Instruments/data format</i>	<i>Software</i>	<i>Producer</i>
e.g. QStar, QTrap	Analyst	Applied Biosystems, Foster City, CA, USA
e.g. HCT, ultraflex, autoflex, XML, ASCII and peak list files	DataAnalysis, FlexAnalysis (previously: XMASS)	Bruker Daltonik GmbH, Bremen, Germany
Mascot Generic Format/MGF files	Mascot	Matrix Science, London, UK
ProteinScape export format	ProteinScape	Bruker Daltonik GmbH, Bremen, Germany Protagen AG, Dortmund, Germany
Finnigan LCQ Deca XP/DTA files	SEQUEST/Xcalibur	Thermo Electron, San Jose, CA, USA
mzXML–open standard file format	Open standard	Institute for Systems Biology, Seattle, WA, USA
Compressed peak lists/CPL files	PeakQuant	
Comma-separated text files (CSV) containing peak list information	Export of instrument software	

average of protein ratios was 0.95 with a standard deviation of 0.44, resulting in a coefficient of variation of 0.4680 (analysis with sequence information). We reanalyzed the same “1:1 mixture” data set without using sequence information (default parameters, no optimization) and determined an average of pair ratios of 1.05 with a standard deviation of 1.37, resulting in

a coefficient of variation of 1.3091. As expected, the accuracy of the results obtained without using sequence information is worse than that seen with using sequence information, probably due to more false-positives. In an analogously produced, more recent “1:1 mixture” spectra data set that was analyzed with tuned parameters, in particular the standard deviation

1741	
1742	
1743	
1744	
1745	
1746	
1747	
1748	
1749	
1750	
1751	
1752	Overview of the FindPairs search:
1753	pairs found: 5
1754	arithmetic mean: 0.76
1755	root mean square error: 0.28, i.e. 37.14 %
1756	calculation time (ms): 0
1757	Parameters used for FindPairs:
1758	use shift information to find matching pairs
1759	shift for: IPI.ARATH - 15N
1760	no limiting range for finding monoisotopic pairs
	monoisotopic mass tolerance: 0.30
	other isotopic mass tolerance: 0.10
	isotope peaks detection type: find exact
	number of considered 13C peaks: 3
	number of considered 15N peaks: 2
	consider intensities (monoisotopic peaks should be highest) yes; the max mass where the monoisotopic peak is highest: 1900.0

FIG. 2. Screenshot of peptide pairs found in one spectrum (FindPairs without sequence information). In the table for the detected peptide pairs (here only 5 in spectrum 1741) the charge states, measured masses, and intensity sums of the light and heavy peptide variants (here called “1” and “2”) are listed. Furthermore, the abundance ratios heavy versus light (“I2:I1”) and vice versa are given. Below the table some statistical information and a summary of the most important FindPairs parameters are shown.

could be improved, to an average of pair ratios of 1.10 with a standard deviation of 0.49, resulting in a coefficient of variation of 0.4479. It shows a clear maximum at 1.0 in a histogram (see Supplementary Fig. S1 at www.liebertpub.com/omi).

Removal of false-positively detected IPPs

To reduce the number of false-positively detected IPPs, the following methods can be applied.

Noise filtering

Noise peaks can lead to false-positively detected IPPs, in particular when detecting peak pairs without utilizing sequence information. They often interfere with correctly detected IPPs. However, they can be identified easily, as they typically show low intensities and/or feature the same masses but different charge states as other peak pairs. By filtering noise (e.g., by ignoring or removing peaks below a certain percentage of the maximum intense peak), the number of false-positive peak pairs is reduced. As mentioned above, this may also reduce some true-positive peak pairs, and the optimal parameter setting for noise reduction depends on the data set, and may be optimized utilizing a “tuning data set.”

Peptide filtering

FindPairs integrates two filters to potentially increase the reliability of peak pair detection.

It allows restricting the quantification process to peak masses within a user-defined mass range (property “use range for monoisotopic peaks”). This can for example be used for excluding m/z below 300 or above 1500, or for focusing on a certain m/z range, where specific pairs are expected.

The height of the peaks of an IPP is based on the distribution of naturally-occurring isotopes (Bohlke et al., 2005). For peptides with masses lower than approximately 1900 Da the highest peak is the monoisotopic one. In case of higher masses—due to the element composition of the amino acids—the probability for having at least one ^{13}C isotope in the peptide is larger than the probability of having only ^{12}C isotopes, resulting in a higher intensity for the ^{13}C peak (Yergey, 1983). FindPairs can reject peptides not fulfilling these characteristics (parameter “consider intensity of monoisotopic peak”).

Using sequence information for removal of false-positively detected IPPs

In the case of $^{14}\text{N}/^{15}\text{N}$ labeling experiments, the use of sequence information leads to exact MSF_{pep} values, and therefore supports correct peak pair detection, even in the presence of peptides that exhibit virtually the same molecular mass but different amino acid sequences. FindPairs utilizes sequence information to calculate the difference ΔTM between the theoretical and the measured mass shift:

$$\Delta\text{TM} = m(\text{light pep})_{\text{theo}} - m(\text{light pep})_{\text{meas}} + m(\text{heavy pep})_{\text{theo}} - m(\text{heavy pep})_{\text{meas}}, \quad (5)$$

where $m(\text{light pep})_{\text{theo}}$ and $m(\text{heavy pep})_{\text{theo}}$ are the m/z values of theoretical light and heavy peptide variants. $m(\text{light pep})_{\text{meas}}$ and $m(\text{heavy pep})_{\text{meas}}$ are the corresponding measured m/z values. ΔTM is then used to reduce false-positive results. In case of ambiguity, FindPairs reports only that sequence with its cor-

responding peak pair, which yield the smallest ΔTM value (parameter “choose only the best pair”).

Detection and filtering of outliers

The identified peptide pairs can be further evaluated for detection and removal of outliers (parameter group “filter outliers”). To this end, a general mechanism capable of detecting multiple outliers was integrated into FindPairs, which is based on the conclusions in Dixon (Dixon, 1953). Peptide pairs are declared as outliers and ignored on the “all Pairs (of Protein)” tab if their intensity ratio exceeds a maximum threshold ($\text{max}_{\text{threshold}}$), or falls below a minimum threshold ($\text{min}_{\text{threshold}}$). Thresholds are calculated by $\text{max}_{\text{threshold}} = \bar{x} + k \cdot \sigma_x$ and $\text{min}_{\text{threshold}} = \bar{x} - k \cdot \sigma_x$, where k describes a user-specified factor, $\bar{x}$ is the arithmetic mean of the ratios of all detected peptide pairs, and σ_x is their standard deviation. This corresponds to the exclusion of values beyond mean plus/minus k times the standard deviation of the current protein in the “with sequence” case (i.e., it is a simple and straightforward mechanism to exclude peptides shared by multiple proteins having different quantitation ratios). Instead of $\bar{x}$ (parameter “use arithmetic mean”), FindPairs also supports the use of the median $\tilde{x}$ (parameter “use median”).

The value of k can be adjusted according to the quality of measurements (parameter “outlier factor for standard deviation”). Applying Chebyshev’s inequality theorem (Chebyshev, 1867), a value of $\sqrt{20}$ for k ensures that 5% of all peptide pairs are categorized as outliers. This holds true independent of the actual distribution. Moreover, elements of a data set with at least a total of 10 elements can be classified as outliers if they are beyond the $4 \sigma_x$ range (Sachs, 2002).

Intensity ratios of the outliers are not considered for the determination of the protein abundance ratios as described in the following.

Determination of protein abundance ratios

After peak pair detection, a summary table reporting all results is provided (Fig. 3). If sequence information is used for the detection, the reported peak pairs belong to one protein. Currently, in the protein summary table the peptide ratios are merged (or as this is also called, condensed or aggregated) to the protein ratio by calculating an outlier-filtered mean (outliers are defined by multiples of standard deviation as described above; the factor can be specified with the parameter “outlier factor for standard deviation”). The intensities of the corresponding light and heavy peptides are listed together with their quotient (the resulting ratio). Peptide pairs with the same m/z values from adjacent MS mass spectra can be merged or listed separately (depending on the parameter “only merge pairs from consecutive spectra”).

To determine biologically meaningful protein expression ratios, independent experiments (biological replicates) have to be performed. In such studies, the quantification results of several FindPairs runs can be transferred to ExperiAna, the statistical analysis module of the PeakQuant suite, for further statistical analysis. There, for example, a Student’s t -test and the 95% confidence interval can be calculated to evaluate statistical significance of changes in protein abundance. Moreover, ExperiAna also assesses the overall variance (σ_g^2) of the experiment, which consists of

FIG. 3. Screenshot of quantification results for a given protein (FindPairs with sequence information). Peptides from a theoretical digest of the NCBI-nr-RefSeq sequence NP_180643.1 "glutathione S-transferase" were used for relative peptide/protein digestion. At the top, the protein sequence is shown. In addition, information on the sequence length and the amino acid coverage (number of amino acids found in peptide pairs) is provided. In the table the detected peptides are listed along with their amino acid sequences (modifications like "oxidation of methionine" can also be displayed), measured masses of the light and heavy peptide variants (here called: "1" and "2"), their corresponding intensity sums, charge states, the overall peptide expression ratio, and the spectra indices in which the peptides were found. Finally, a statistical summary of the quantitative protein analysis is shown.

the biological variance (σ_a^2) and the technical variance (σ_e^2) (Bosch, 1998).

Results and Discussion

In this section, the calculation of MSFs is demonstrated and the quality of the obtained quantification data is exemplarily evaluated. Next, the features of FindPairs are compared with the possibilities of other existing software solutions. Finally, the features of the FindPairs algorithm are summarized and examples of successful applications are given.

Examples for peak pair detection of FindPairs

Example without sequence information (using MSF_{DB} values). In case of $^{14}N/^{15}N$ labeling experiments without using sequence information, the database-derived MSF_{DB} value is applied for the detection of peak pairs in MS spectra. In the following example, this approach is demonstrated for the doubly-charged peptide VYGPHFASPK (the IPP is shown in Fig. 1). Monoisotopic peaks of its light and heavy form are expected at m/z 551.79 and 558.27 (including proton masses), respectively. The theoretical mass difference is 6.48 Da. The theoretical MSF ($MSF_{theoretical}$) of this $^{14}N/^{15}N$ -labeled peptide is calculated as follows (proton masses subtracted):

$$MSF_{theoretical} = \frac{m(^{heavy}X_{peptide})}{m(^{light}X_{peptide})} = \frac{m(^{15}N_{peptide})}{m(^{14}N_{peptide})} = \frac{556.26}{549.78} = 1.011787. \quad (6)$$

A theoretical average MSF for the NCBI-nr database (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA>) was determined with [Eq. 4]: $MSF_{NCBI-nr} = 1.012198$ (for detailed results see Table 1).

Using [Eq. 2] with this $MSF_{NCBI-nr}$, the predicted m/z value of the ^{15}N peak is calculated as 558.50 Da. Thus the mass difference between the theoretical and the estimated ^{15}N peak is 0.23 Da. This corresponds to a ^{15}N estimation exactness δ_{15N} of 99.959% according to

$$\delta_{15N} = \frac{m(^{15}N_{theoretical}) - m(^{15}N_{estimated})}{m(^{15}N_{theoretical})}. \quad (7)$$

Hence the application of a theoretical MSF provides a valuable means for detecting $^{14}N/^{15}N$ -coded peak pairs, if no sequence information is available. In the actual analysis, the expected mass error produced by the MS instrument has to be additionally incorporated via the parameter "mass tolerance for other isotopic peaks."

Example without sequence information. When utilizing sequence information, FindPairs uses theoretical m/z values of stable isotope-labeled peptides to detect peak pairs in MS survey spectra. FindPairs can use protein sequences provided via the graphical user interface, or those retrieved from protein databases.

Theoretical digestion of proteins is performed considering the properties of the chosen enzyme, the number of missed cleavage sites, and user-defined modifications (the "Digest/Mod" tab). As described for the $^{14}N/^{15}N$ -labeled peptide pair used as an example above, monoisotopic peaks of the light

and heavy peptide form are expected at m/z 551.79 and 558.27, respectively. Monoisotopic peaks are observed at m/z 551.84 and m/z 558.32, as shown in the mass spectrum depicted in Figure 1. Comparing the calculated and measured monoisotopic masses yields a deviation of 0.05 Da for both peaks. Thus, if sequence information is used, an improved exactness of $\delta_{14N} = \delta_{15N} = 99.991\%$ is reached in this example.

Feature comparison with existing protein quantification software

Several other software tools have been developed for the automatic determination of peptide/protein quantification (e.g., MSQuant [Schulze and Mann, 2004]; Xpress [Han et al., 2001]; ASAPratio [Li et al., 2003]; QN [Andreev et al., 2006]; VEMS [Matthiesen et al., 2005; Rodriguez-Suarez et al., 2010]; RelEx/Census [MacCoss et al., 2003; Park et al., 2008]; ProRata [Pan et al., 2006]; ZoomQuant [Halligan et al., 2005]; i-Tracker [Shadforth et al., 2005]; MaxQuant [Cox and Mann, 2008]; Multi-Q [Lin et al., 2006]; MFPAQ [Bouyssié et al., 2007]; SILACTor [Hoopmann et al., 2011]).

Some of these tools were developed to meet the specific needs of distinct labeling technologies such as SILAC or iTRAQ, which often makes them difficult to adapt to other labeling methods. Multi-Q and i-Tracker are used to quantify data sets that were generated using isobaric tags (iTRAQ). Quantification of data generated using ICAT as well as methyl esterification (Goodlett et al., 2001) can be performed with ASAPratio. SILAC experiments with incorporated heavy amino acids can be evaluated with either MaxQuant or MSQuant. MSQuant also supports processing of $^{14}N/^{15}N$ labeling data, but needs the help of extra software (as stated at <http://msquant.alwaysdata.net/> under the heading “N15 labeling”). MFPAQ allows one to quantitatively analyze both ICAT and SILAC data. ZoomQuant supports the quantitative analysis of $^{16}O/^{18}O$ -labeled peptides. QN is a specialist for the calculation of relative protein abundance ratios using the $^{14}N/^{15}N$ labeling technique. SILACTor can handle data sets with several time points.

In contrast to these software tools, software applications such as Xpress, VEMS, and Census may be referred to as all-rounders, providing the ability to process data from most of the available stable isotope labeling techniques.

Some of the above-mentioned software tools are only suitable for a specific ionization method (e.g., ESI [Fenn et al., 1989] or MALDI [Karas and Hillenkamp, 1988]), or for specific MS instruments. For example, QN is especially adapted for the use of data obtained from high-resolution instruments (LTQ-Orbitrap, Q-TOF, or LTQ-FT). Sometimes a specific infrastructure in the lab is a prerequisite, such as special protein search engines (e.g., SEQUEST [Eng et al., 1994] or Mascot [Perkins et al., 1999]), or particular computer operating systems. For example, VEMS and MSQuant require a Windows operating system. ProRata and the standard quantitative proteomics procedure of Xpress are related to ESI and protein identifications obtained from SEQUEST database searches (Han et al., 2001).

The PeakQuant suite, including FindPairs, is characterized by flexible use of the software due to many adjustable parameters. Because the PeakQuant suite is implemented using the Java programming language, the software is executable on

all major computer platforms. Additionally, PeakQuant can be started directly from an intra- or Internet web server (“Java web start” mechanism).

Most software available for automated quantitative analysis data generated by using $^{14}N/^{15}N$ labeling require protein sequence information. Therefore, the approach proposed here using average MSFs is most suitable for the analysis of meta-proteomics data.

There is a strong necessity to repeat proteomics experiments on independent samples in order to cope with biological variations in protein abundance ratios. However, available quantification software frequently does not include adequate statistical methods. Further investigation of the obtained results needs to be carried out with additional computer programs. In contrast, FindPairs takes direct advantage of the statistical modules of the PeakQuant suite, and therefore supports automatic statistical analysis across multiple independent biological experiments.

Conclusions and Outlook

The PeakQuant suite provides a unique platform for reliable and efficient proteomics investigations. In this article, the features of the PeakQuant module FindPairs for time-efficient determination of protein abundance ratios from stable isotope-labeled peptide mass spectra have been described. The focus here was in particular on the applications of FindPairs to $^{14}N/^{15}N$ labeling. FindPairs implements two techniques for peak detection, differing in whether or not sequence information is considered. The detection of the ^{15}N peak using sequence information has considerably higher precision than using average MSFs. As expected, an initial identification step of the peptides/proteins thus leads to more precise quantification results. However, calculation of average MSFs provides a promising alternative approach for quantifying peptides in cases for which limited or no sequence information is available about the samples analyzed (e.g., in the meta-proteomics field).

The algorithm was comprehensively tested for the MS-based determination of protein abundance ratios employing $^{14}N/^{15}N$ labeling and SILAC (data not shown). High applicability of PeakQuant to $^{14}N/^{15}N$ labeling experiments was successfully shown by performing automated relative protein quantification on a data set acquired to study early leaf senescence of *Arabidopsis thaliana* (Hebeler et al., 2008; the algorithm was formerly termed “Peakardt”). In this previous work, protein quantification data obtained via $^{14}N/^{15}N$ labeling combined with MS and two-dimensional DIGE (Unlu et al., 1997) analysis were compared, demonstrating good correlation of the resulting quantitative data sets in general. Statistical evaluation of multiple biological experiments was also performed with the ExperiAna module of the PeakQuant suite to determine significant changes in protein abundance in wild-type and mutant plants. In addition, the PeakQuant suite was successfully applied for the quantification of iTRAQ-labeled proteins (Wiese et al., 2007). As PeakQuant is an active project, users are encouraged to request functionality extensions (see <http://www.medizinisches-protem-center.de/software>). Future extensions of the PeakQuant. FindPairs module will, for example, comprise a weighted consideration of peaks with different charge states when merging pairs, a mass-range-dependent tolerance, and the

integration of an importer for the mzML standard format (Martens et al., 2011). Furthermore, the performance of FindPairs and other $^{14}\text{N}/^{15}\text{N}$ quantification tools should be evaluated on the basis of complex standard data sets produced on several state-of-the-art instruments.

Acknowledgments

The authors wish to thank Kai A. Reidegeld for implementing the previous version of the program (called "Peakardt"). This study was supported in part by the German Bundesministerium für Bildung und Forschung with grant 0313318B. M.E. and C.S. are funded by P.U.R.E. (Protein Unit for Research in Europe), a project of Nordrhein-Westfalen, a federal state of Germany. M.K. is funded by BioNRW.PRO-FILE, Förderkennzeichen 005-1006-0050, Projektträger Jülich (PTJ) project number z0911bt004f. R.H. was supported by the European Graduate College 795 "Regulatory Circuits in Cellular Systems: Fundamentals and Biotechnological Applications." The research of B.W. and S.W. is supported by the Excellence Initiative of the German Federal and State Governments (EXC 294 BIOS).

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

- Andersen, J.S., Lam, Y.W., Leung, A.K., et al. (2005). Nucleolar proteome dynamics. *Nature* 433, 77–83.
- Andreev, V.P., Li, L., Rejtar, T., Li, Q., Ferry, J.G., and Karger, B.L. (2006). New algorithm for N-15/N-14 quantitation with LC-ESI-MS using an LTQ-FT mass spectrometer. *J Proteome Res* 5, 2039–2045.
- Audi, G., and Wapstra, A.H. (1995). The 1995 update to the atomic mass evaluation. *Nuclear Physics A* 595, 409.
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Analy Bioanal Chem* 389, 1017–1031.
- Beynon, R.J., and Pratt, J.M. (2005). Metabolic labeling of proteins for proteomics. *Mol Cell Proteomics* 4, 857–872.
- Bohlke, J.K., De Laeter, J.R., De Bièvre, P., et al. (2005). Isotopic compositions of the elements, 2001. *J Phys Chem Ref Data* 34, 57–67.
- Bosch, K. (1998). *Statistik-Taschenbuch*, 3rd ed. München: Oldenbourg, Wien.
- Bouyssié, D., De Peredo, A.G., Mouton, E., et al. (2007). Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses. *Molec Cellular Proteomics* 6, 1621–1637.
- Cagney, G., and Emili, A. (2002). De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat Biotechnol* 20, 163–170.
- Chakraborty, A., and Regnier, F.E. (2002). Global internal standard technology for comparative proteomics. *J Chromatogr A* 949, 173–184.
- Chebyshev, P.L. (1867). Des valeurs moyennes. *Liouville's J Math Pures Appl* 12, 177–184.
- Coplen, T.B. (2001). Atomic weights of the elements 1999. *Pure Appl Chem* 73, 667–683.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol* 26, 1367–1372.
- Dixon, W. (1953). Processing data for outliers. *Biometrics* 9, 74–89.
- Eng, J.K., and McCormack, A.L., 3rd (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 976–989.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64–71.
- Foster, L.J., De Hoog, C.L., and Mann, M. (2003). Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci USA* 100, 5813–5818.
- Goodlett, D.R., Keller, A., Watts, J.D., et al. (2001). Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Spectrom* 15, 1214–1221.
- Gouw, J.W., Tops, B.B., Mortensen, P., Heck, A.J., and Krijgsveld, J. (2008). Optimizing identification and quantitation of ^{15}N -labeled proteins in comparative proteomics. *Anal Chem* 80, 7796–7803.
- Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994–999.
- Halligan, B.D., Slyper, R.Y., Twigger, S.N., Hicks, W., Olivier, M., and Greene, A.S. (2005). ZoomQuant: an application for the quantitation of stable isotope labeled peptides. *J Am Soc Mass Spectrom* 16, 302–306.
- Han, D.K., Eng, J., Zhou, H.L., and Aebersold, R. (2001). Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 19, 946–951.
- Hebeler, R., Oeljeklaus, S., Reidegeld, K.E., et al. (2008). Study of early leaf senescence in *Arabidopsis thaliana* by quantitative proteomics using reciprocal N-14/N-15 labeling and difference gel electrophoresis. *Molec Cellular Proteomics* 7, 108–120.
- Hoopmann, M.R., Chavez, J.D., and Bruce, J.E. (2011). SILACTor: Software To Enable Dynamic SILAC Studies. *Anal Chem* 83, 8403–8410.
- Ibarrola, N., Kalume, D.E., Gronborg, M., Iwahori, A., and Pandey, A. (2003). A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Anal Chem* 75, 6043–6049.
- Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60, 2299–2301.
- Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24, 2534–2536.
- Kierszniowska, S., Walther, D., and Schulze, W.X. (2009). Ratio-dependent significance thresholds in reciprocal ^{15}N -labeling experiments as a robust tool in detection of candidate proteins responding to biological treatment. *Proteomics* 9, 1916–1924.
- Lin, W.T., Hung, W.N., Yian, Y.H., et al. (2006). Multi-Q: A fully automated tool for multiplexed protein quantitation. *J Proteome Res* 5, 2328–2338.
- Lipman, D.J., and Pearson, W.R. (1985). Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441.
- Li, X.J., Zhang, H., Ranish, J.A., and Aebersold, R. (2003). Automated statistical analysis of protein abundance ratios from

- data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem* 75, 6648–6657.
- Maccoss, M.J., Wu, C.C., Liu, H., Sadygov, R., and Yates, J.R., 3rd (2003). A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem* 75, 6912–6921.
- Martens, L., Chambers, M., Sturm, M., et al. (2011). mzML—a community standard for mass spectrometry data. *Mol Cell Proteomics* 10, R110 000133.
- Matthiesen, R., Trelle, M.B., Hojrup, P., Bunkenborg, J., and Jensen, O.N. (2005). VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res* 4, 2338–2347.
- Munday, D.C., Hiscox, J.A., and Barr, J.N. (2010). Quantitative proteomic analysis of A549 cells infected with human respiratory syncytial virus subgroup B using SILAC coupled to LC-MS/MS. *Proteomics* 10, 4320–4334.
- Oda, Y., Huang, K., Cross, F.R., Cowburn, D., and Chait, B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA* 96, 6591–6596.
- Ong, S.E., Blagoev, B., Kratchmarova, I., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376–386.
- Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nature Chem Biol* 1, 252–262.
- Ong, S.E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* 1, 119–126.
- Pan, C.L., Kora, G., McDonald, W.H., et al. (2006). ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem* 78, 7121–7131.
- Park, S.K., Venable, J.D., Xu, T., and Yates, J.R. (2008). A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 5, 319–322.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.
- Peters, E.C., Horn, D.M., Tully, D.C., and Brock, A. (2001). A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun Mass Spectrom* 15, 2387–2392.
- Qiu, Y., Sousa, E.A., Hewick, R.M., and Wang, J.H. (2002). Acid-labile isotope-coded extractants: a class of reagents for quantitative mass spectrometric analysis of complex protein mixtures. *Anal Chem* 74, 4969–4979.
- Rodriguez-Suarez, E., Gubb, E., Alzueta, I.F., et al. (2010). Virtual expert mass spectrometrist: iTRAQ tool for database-dependent search, quantitation and result storage. *Proteomics* 10, 1545–1556.
- Rosman, K.J.R., and Taylor, P.D.P. (1997). Isotopic compositions of the elements 1997. *Int Union Pure Appl Chem*.
- Ross, P.L., Huang, Y.N., Marchese, J.N., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154–1169.
- Sachs, L. (2002). *Angewandte Statistik*, 10th ed. New York: Springer.
- Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4–15.
- Schnolzer, M., Jedrzejewski, P., and Lehmann, W.D. (1996). Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* 17, 945–953.
- Schulze, W.X., and Mann, M. (2004). A novel proteomic screen for peptide-protein interactions. *J Biol Chem* 279, 10756–10764.
- Shadforth, I.P., Dunkley, T.P.J., Lilley, K.S., and Bessant, C. (2005). i-Tracker: For quantitative proteomics using iTRAQ (TM). *BMC Genomics* 6.
- Skirycz, A., Memmi, S., De Bodt, S., et al. (2011). A reciprocal ^{15}N -labeling proteomic analysis of expanding *Arabidopsis* leaves subjected to osmotic stress indicates importance of mitochondria in preserving plastid functions. *J Proteome Res* 10, 1018–1029.
- Soufi, B., Kumar, C., Gnad, F., Mann, M., Mijakovic, I., and Macek, B. (2010). Stable isotope labeling by amino acids in cell culture (SILAC) applied to quantitative proteomics of *Bacillus subtilis*. *J Proteome Res* 9, 3638–3646.
- Thompson, A., Schafer, J., Kuhn, K., et al. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75, 1895–1904.
- Thorn, A.M., and Orians, C.M. (2011). Partial defoliation and hydraulic integration in *Ocimum basilicum* (Lamiaceae): Testing a model for sectored xylem flow using ^{15}N labeling. *Am J Bot* 98, 1816–1824.
- Unlu, M., Morgan, M.E., and Minden, J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071–2077.
- Wiese, S., Reidegeld, K.A., Meyer, H.E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7, 340–350.
- Yao, X., Freas, A., Ramirez, J., Demirev, P.A., and Fenselau, C. (2001). Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 73, 2836–2842.
- Yergey, J.A. (1983). A general approach to calculating isotopic distributions for mass spectrometry. *Int J Mass Spectrometry Ion Physics* 52, 337–349.

Address correspondence to:

Bettina Warscheid

Funktionelle Proteomik

Fakultät für Biologie und BIOSS Centre

for Biological Signalling Studies

Universität Freiburg, Schänzlestr 1

Freiburg, Germany 79104

E-mail: bettina.warscheid@biologie.uni-freiburg.de

Christian Stephan

Medizinisches Proteom-Center

Ruhr-Universität Bochum

Universitätsstr 150 (building ZKF E.043)

D-44801 Bochum, Germany

E-mail: christian.stephan@rub.de