

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/jjprot](http://www.elsevier.com/locate/jjprot)

## Review

# Experimental and computational approaches to quantitative proteomics: Status quo and outlook

Alexandre Panchaud<sup>a,b</sup>, Michael Affolter<sup>b</sup>, Philippe Moreillon<sup>a</sup>, Martin Kussmann<sup>b,\*</sup>

<sup>a</sup> Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland

<sup>b</sup> Functional Genomics Group, Department of Bioanalytical Science, Nestlé Research Centre, Vers-chez-les-Blanc, P.O. Box 44, CH-1000 Lausanne 26, Switzerland

### ARTICLE INFO

#### Article history:

Received 1 October 2007

Accepted 18 December 2007

#### Keywords:

Quantitation

Stable isotope labeling

Aniline

Benzoic acid

### ABSTRACT

Proteomics has come a long way from the initial qualitative analysis of proteins present in a given sample at a given time (“cataloguing”) to large-scale characterization of proteomes, their interactions and dynamic behavior. Originally enabled by breakthroughs in protein separation and visualization (by two-dimensional gels) and protein identification (by mass spectrometry), the discipline now encompasses a large body of protein and peptide separation, labeling, detection and sequencing tools supported by computational data processing. The decisive mass spectrometric developments and most recent instrumentation news are briefly mentioned accompanied by a short review of gel and chromatographic techniques for protein/peptide separation, depletion and enrichment. Special emphasis is placed on quantification techniques: gel-based, and label-free techniques are briefly discussed whereas stable-isotope coding and internal peptide standards are extensively reviewed. Another special chapter is dedicated to software and computing tools for proteomic data processing and validation. A short assessment of the *status quo* and recommendations for future developments round up this journey through quantitative proteomics.

© 2007 Elsevier B.V. All rights reserved.

### Contents

1. Proteomics: hopes, promises, deliveries and gaps . . . . .	20
2. Concepts in proteomics . . . . .	21
3. Instrumentation and separation strategies. . . . .	21
4. Computational approaches for protein identification and validation . . . . .	22
5. Quantification approaches and tools . . . . .	23
6. Proteomics —future directions . . . . .	26
References . . . . .	29

\* Corresponding author. Functional Genomics Group, Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland. Tel.: +41 21 785 89 66; fax: +41 21 785 94 86.

E-mail address: [martin.kussmann@rdls.nestle.com](mailto:martin.kussmann@rdls.nestle.com) (M. Kussmann).

## 1. Proteomics: hopes, promises, deliveries and gaps

The rapid rise of Proteomics as a scientific discipline is a good example of how technology has driven biology, whereas traditionally researchers have thought the other way round, in the sense that biology drives technology. The invention of mass spectrometers enabling the analysis of proteins and peptides – molecules that have classically been studied individually or in small numbers – at large-scale and in high-throughput mode appeared to place the proteome – i.e. the total protein complement of a cell, organ or even organism – as the analogue to the genome within reach of analytical means [1]. The “proteomic gold rush” in the later 1990s and early 2000s was mainly fed by the ambition of cataloguing all the proteins in the human body, of generating disease biomarkers and of filling the drug discovery pipelines with candidates. Deciphering the proteome was simply the next logical challenge after the discovery that the human genome with its only 30,000+ genes [2] is way too small to explain the complexity of human biology.

Unfortunately, none of the three hopes (or hypes?), i.e. of identifying all human proteins and of generating many clinical biomarkers and drug (target) candidates, have been fulfilled in the meantime, for several reasons. The goal of charting the human proteome was early questioned by its complexity that had soon turned out to be more difficult than expected, because of both the number of proteins present and their wide concentration range [3]. The analytical challenges of analyzing the human proteome have simply been underestimated. Nevertheless, based on an estimated protein concentration range of  $10^{10}$  to  $10^{12}$  in human plasma and 6 to 8 orders of magnitude in human cells [4], remarkable progress has been achieved in the meantime: current mass spectrometric platforms can cover a dynamic range of up to  $10^4$  [4], depletion of abundant and/or enrichment of less prominent proteins have been combined, and extensive protein and peptide pre-fractionation and sophisticated data acquisition and processing are in place. However, simple calculations based on the numbers above reveal that there are still several orders of magnitudes left until one can cope with the dynamic range of e.g. human cellular, tissue and body fluid proteomes. We are far from looking at the totality of the plasma proteome, to mention the clinically most relevant sample for protein biomarkers. While we are digging more deeply into proteomes, also thanks to biologically driven fractionation down to cell organelles, there are still studies “re-discovering similar proteomes in different samples”. This observation creates the impression of scratching the surface: it seems that many studies still reveal similar “housekeeping proteomes” as found in many cell and tissue types, or that the candidate biomarkers for a specific disease are often markers for a generic phenomenon underlying several diseases, such as inflammation.

These preceding statements on the (in-)completeness of proteome coverage are based on the ever increasing but still too small numbers of proteins revealed by proteomic studies and on the concentration range covered as exemplified by proteins at the top and bottom of such lists, for which non-proteomic concentration data exist. Today’s shotgun proteomic approaches reveal up to some 2000 protein identities. However, the number of true positive protein identifications within those

long lists has enormously increased over the years. This important achievement can be attributed to (a) improved mass spectrometric instrumentation with better sensitivity and specificity, the latter based on superior mass accuracy and resolution (e.g. Orbitrap [5]); and (b) more sophisticated software for peptide sequencing and protein identification (e.g. Phenyx [6], OMSSA [7], X!Tandem [8] or VEMS [9–11]). The computing tools for proteomic data processing have evolved from scoring mass spectra with regard to their fidelity for peptide sequence read out to assessing the trade-off between false-positives (specificity) and false-negatives (sensitivity) in a data-dependent manner, eventually leading to computational data validation (PeptideProphet [12] and ProteinProphet [13]).

The reasons, for which the second and third hope of proteomics, i.e. of rapidly filling the clinical biomarker and drug discovery pipelines, has not become a reality, are manifold and extend beyond the analytical problems discussed above. There is first the issue of converting data into information, translating protein identifications into candidate markers or targets that deserve follow-up. The large amount of protein identifications can render this transformation a daunting task, especially when such candidates have to fit into existing discovery programs in the pharmaceutical or nutritional industry. Second, many of the marker/target validation technologies are still low-throughput and require multi-centric collaboration: *in vivo* validation takes much longer than *in vitro* discovery [14]. Third, the pay-off for protein diagnostic development, i.e. for taking a protein biomarker to the bed-side test, has so far been small in economic terms [14]. However, many if not most clinical tests today measure proteins and the examples of protein-based biomarkers successfully applied in the clinical stage should encourage: (i) Prostate-specific antigen (PSA) for prostate cancer [15]; (ii) liver transaminases for liver cell destruction [16]; and (iii) Troponin I and T for acute myocardial infarction [17]. Furthermore, a couple of successful proteomics-derived biomarkers can be cited: (i) The 14-3-3 proteins in cerebrospinal fluid (CSF) function as markers for types of brain destruction such as found in Creutzfeldt-Jakob Disease (CJD) [18]; (ii) stroke biomarkers have been discovered in the spinal fluid and have been clinically validated in serum [4].

The previous paragraphs have shed some light on the evolution of proteomics as a discovery platform for clinical biomarkers and drug targets. Another aspect of the evolution of proteomic platforms is their meanwhile widespread application that extends beyond markers and targets. Proteomics is deployed for the elucidation of mechanisms in health and disease and has matured from an academic, basic research platform to an analytical tool now widely employed in biological, pharmaceutical and, more recently, also nutritional research. In nutrition, for example, proteomics (in conjunction with transcriptomics and metabolomics) is used for efficacy, quality and safety assessment of food stuffs and specific ingredients [19]. Moreover, proteomic profiles are monitored to follow dietary interventions and characterize their outcome [20]. In this context, proteomics will help bridging from “pure bioavailability” studies for nutrients to demonstration of bioefficacy of these functional ingredients [21]. Last, but certainly not least, the identification of bioactive food proteins and peptides also relies on proteomics [22].

Proteomics in the clinical, medical and pharmaceutical context is mainly expected to deliver markers for disease prognosis, state and treatment outcome and targets for disease treatment [23]. By contrast, Proteomic research in nutrition and health has a different deliverable set: nutritionists apply proteomics to discover biomarkers for early individual disposition towards disease onset and to reveal markers of ingredient/nutrient efficacy [22]. These objectives derive from the fact that modern nutrition aims at health promotion, disease prevention and performance improvement. As nutritional interventions are more subtle than drug treatments and because early disposition for nutritionally actionable diseases like allergy or diabetes is expected to be difficult to detect, nutritionists are rather looking for characteristic proteome profiles than for single protein biomarkers. The discouraging clinical proteomics example that most candidate cancer biomarkers have turned out to rather reflect the general condition of inflammation than that of cancer [4] illustrates the difference between proteomics applied to nutrition as opposed to proteomics deployed for drugs: nutritional modulation of the immune status by e.g. dampening (chronic) inflammation is a major research and consumer care interest and, therefore, these proteomics-derived inflammation markers are of great relevance to nutritional intervention studies [24].

## 2. Concepts in proteomics

Several decisive factors render proteomics such an essential field of activity in life science research:

- (i) Researchers have realized that even with the availability of complete genome sequences, this information is not sufficient to derive biological function.
- (ii) Proteomics is complementary to genomics because it focuses on the gene products that are the active agents in a cell, and therefore potential targets for drug development.
- (iii) There is no linear relationship between the genome and the proteome of a cell and protein expression has to be determined directly as it does not always correlate with mRNA levels [25].
- (iv) Prediction of genes from genomic data remains difficult even with modern bioinformatic tools [26] and verification of a gene product by proteomic methods is an important annotation step (e.g. PeptideAtlas project (<http://www.peptideatlas.org/>)).
- (v) Protein modifications or protein localization are not visible and barely predictable at DNA sequence level and, therefore, proteomics is indispensable for the elucidation of protein isoforms, post-translational modifications and gene product localization.
- (vi) Finally, protein regulation mechanisms, protein-protein interactions or molecular compositions of cellular structures such as organelles can be determined only at protein level.

The complexity of a proteome – the set of all expressed proteins in a cell, tissue or organism [27] – is fascinating as it is overwhelming. This complexity is due to several events such as gene and protein splicing and post-translational modifica-

tions. To date, more than 100 protein modification types are described [28]. Clearly, post-translational modification is an event with dramatic effect on the proteome complexity.

In addition to this high complexity, the hallmark of a proteome is its ability to regulate protein expression dynamically in response to physiological, developmental and pharmacological conditions or aging. The huge dynamic range, i.e. the wide difference in concentration from the most to the least abundant proteins, renders global protein analysis a challenging task. It is estimated that the dynamic range of a proteome spans six to eight orders of magnitude in a cell and even ten to twelve in the human body with blood plasma as an example [29]. The fact that protein analysis lacks the equivalent of PCR amplification does not make it simpler [30].

Obviously no general method allows to identify and quantify in a single experiment all proteins of a given sample, as opposed to DNA microarrays [31], where global gene expression profiles can be obtained from total extracted RNA deploying automated systems as described by Raymond et al. [32]. While genome-wide microarrays are readily available [33,34], the equivalent of a proteome microarray is missing, mainly for the following reasons: i) as opposed to nucleotide sequences, proteins do not share the same hybridization properties as nucleic acids and, therefore, antibodies are used as capture molecules with their preparation and production being difficult and time-consuming; ii) protein concentrations in a biological sample may span more orders of magnitude than those of mRNAs, making protein on-chip detection more difficult. Consequently, proteomic laboratories have to design experiments according to the question to be addressed and take advantage of an ample biochemical toolbox. Examples are: fractionation vs. depletion vs. enrichment [3,35], total proteome vs. sub-proteomes (e.g. glyco-/phosphoproteins) [36,37], biological fluids vs. tissue, and total cellular proteomes vs. sub-cellular proteomes [38,39].

In addition to these analytical considerations, the potential and the limitations of proteomics technologies should be well understood. Compared to traditional analytical strategies that focus on a single or a few molecules and are based on a pre-conception, the main advantage of proteomics and other omics based approaches relies in the fact that they do not depend on a hypothesis but rather can generate new ones. This said proteomics and related omics technologies produce extremely large amounts of data, which makes result exploitation difficult. Nowadays, these data can be handled and processed thanks to advanced computer hard- and software. Yet, converting these large data sets into interpretable biological information remains a challenge.

## 3. Instrumentation and separation strategies

The mass analyzer is central to mass spectrometric technology, and in the proteomics context, its key parameters are sensitivity, resolution, mass accuracy and ability to produce information-rich fragment mass spectra from peptide ions (tandem mass or MS/MS spectra). There are five basic types of mass analyzers currently used in proteomics [40]: ion trap (IT), time-of-flight (TOF), quadrupole (Q), Fourier transform ion cyclotron resonance (FT-ICR), and the newly developed

Orbitrap system [5]. They are different in conception and performance, each with its own strengths and weaknesses. Often, they work as stand-alone mass analyzer, but the current trend points towards hyphenated systems in order to combine the advantages of different analyzers in one mass spectrometer: triple-Q, Q-IT, Q-TOF, IT-TOF, TOF-TOF, IT-FT-ICR or IT-Orbitrap tandem mass spectrometers are all capable of protein or peptide sequencing. IT-FT-ICRs and IT-Orbitraps are especially efficient when combined with fragmentation techniques such as electron capture dissociation (ECD) [41] or electron-transfer dissociation (ETD) [42].

MS-based proteomics is about identifying and quantifying as many proteins as possible in a single experiment. While the mass spectrometric toolbox is impressive and has diversified and considerably progressed over the last two decades, even with the best detector, only few proteins are identified if no pre-separation and reduction of sample complexity is achieved prior to sample introduction in the mass spectrometer.

The traditional strategy used over the last 20 years for protein separation and visualization is 2D-PAGE. In the gel approach, a protein is resolved by molecular weight and isoelectric point and isolated from other proteins in a single spot except in the case of co-migration. The spot is then excised and in-gel digested. The resulting proteolytic peptides are extracted from the gel pieces and typically analyzed by MALDI-TOF MS or MS/MS [43]. While 2D gels are still popular and powerful, the liquid chromatography approach appears to take over in terms of numbers of studies and, even more so, with regard to number of proteins identified, mainly for reasons of higher throughput.

Apart from the more traditional 2D gels, two main concepts are currently applied to deal with proteome complexity and dynamic range: fractionation using extensive separation techniques such as multidimensional protein identification technology (MudPIT) [44] or enrichment/depletion using affinity-based techniques [45–47].

Most of LC-MS/MS-based proteomic studies rely on on-line LC-ESI systems. However, to combine features of the MALDI technique, where peptides can be immobilized on a target plate, with the separation power of liquid chromatography, LC-MALDI was developed in an automated, but off-line mode [48]. MS and MS/MS acquisition is now time-wise uncoupled from LC separation. Comparisons of LC-ESI-MS/MS and LC-MALDI-MS/MS have revealed the complementarities between the two strategies and clearly highlight the fact that no single approach is capable of unraveling a proteome, but rather that many technologies have to be used in combination.

#### 4. Computational approaches for protein identification and validation

Proteins can be identified with good throughput [49] and high sensitivity [50] based on the set of measured proteolytic peptide masses. This process is known as peptide mass fingerprinting (PMF). The experimental mass profile is matched against those generated *in silico* from the protein sequences in the database using the same enzyme cleavage sites. The proteins are then ranked according to the number of peptide masses matching their sequence within a certain

mass error tolerance. Unfortunately, the lack to sequence data makes identification nowadays ambiguous because of: (i) database size (especially for eukaryotes) [51]; (ii) percentage of experimental peaks unmatched (matrix adducts, keratin contaminants, trypsin autolysis, etc.); and (iii) still too low mass accuracy (especially when there is no FT-ICR or Orbitrap at hand). Examples of PMF algorithms are: Mascot [51], Aldente (<http://www.expasy.org/tools/aldente/>), MS-FIT [52], Pro-Found [53] or VEMS [9–11].

By contrast, MS/MS provides access to sequence data, which allows to more confidently identify peptides. In an MS/MS experiment, a precursor ion with known mass is selected from the previous MS scan and isolated for further collision to produce daughter ions with unique signature. This process is described as peptide mass sequencing (PMS) as opposed to PMF. Different daughter ion types are produced depending on the ion source-(MALDI or ESI) and analyzer type (quadrupoles, ion trap, TOF, etc.), fragmentation method (collision-induced dissociation (CID), ECD, ETD) and collision energy (low vs. high energy). Most common ions are y- or b-ions. Other ion types may be generated through loss of water, NH<sub>3</sub>, CO or by immonium ion formation. A nomenclature has been described by Roepstorff and Fohlmann [54] and subsequently modified by Johnson and colleagues [55] to the commonly known Biemann notation.

Identification of proteins using MS/MS data is nowadays performed using three different approaches: (i) peptide sequence tag, (ii) cross-correlation method and, (iii) probability based matching.

The “peptide sequence tag” approach extracts a short, unambiguous amino acid sequence from the peak pattern. Combined with the mass information, this small sequence is a specific probe to determine the origin of the peptide [56,57]. Examples of such algorithms are GutenTag [58], MS-seq (<http://prospector.ucsf.edu/ucsfhtml4.0/msseq.htm>) or MultiTag [59].

In the “cross-correlation method”, peptide sequences in the database are used to construct theoretical mass spectra and the overlap or cross-correlation of these predicted spectra versus the measured mass spectra determines the best match [60]. A typical example of such an algorithm is Sequest [60] or SALSA [61].

In the “probability based matching”, the calculated fragments from peptide sequences in the database are compared with observed peaks and from this comparison and a score is calculated, which reflects the statistical significance of the match between the spectrum and the sequences present in the database. Typical examples are Mascot [51], OMSSA [7], X! Tandem [8], Phenyx [6] or VEMS [9–11].

All previously described approaches are based on the assumption that a proteomic or translated genomic databases is available to perform the search. *De novo* sequencing uses the MS/MS spectra as the only reference to deduce the peptide sequence. In theory, all the information necessary to reconstruct the peptide sequence should be present in the spectrum. In reality, spectra of high quality are rare because of e.g. ion-suppression effects or poor signal-to-noise ratio. The advantages of *de novo* methods are: (i) no need for reference database and, hence, applicability to unsequenced organisms, (ii) usefulness for sequencing of artificial peptides, (iii) sequencing of peptides for which an error of genome

sequencing has occurred or, (iv) identification of post-translational modifications. For a detailed review on *de novo* sequencing, the readers are referred to the review of Lu and Chen [62].

A recently developed approach based on MS/MS spectra libraries (SpectraST) [63] is a promising alternative: an MS/MS spectra library is created based on previous database identifications on the same sample and is used for identification for following, related experiments. The method outperforms standard database search engines in terms of speed and ability to discriminate good and bad hits, but this method is not a discovery tool.

The best case scenario in terms of identification would be to use only the mass information of a peptide as a unique signature. Such approaches have already been described by Zubarev and colleagues [64] and later on by Conrads and co-workers [65] as the “accurate mass tag” (AMT) approach. In this technique, identification is only based on the peptide mass and high resolution instruments are needed providing sub-ppm mass accuracy (0.1 ppm). But even with such accuracy, only in small eukaryotic systems (e.g., yeast) high levels of confidence in protein identifications can be achieved. However, as the genomic, and thus, the proteomic, complexity of an organism increases with higher organization, the ability to identify proteins (or peptides) on the basis of mass measurements alone decreases. Until mass spectrometers will be capable of low ppb mass accuracy, additional information such as isoelectric point or LC elution time are needed to confidently identify a protein without using tandem mass spectrometry [66–68].

Liquid chromatography and tandem mass spectrometry have become the preferred methods to conduct large-scale characterization of proteomes. Several completely automated MS/MS search engines are available that enable the submission of extremely large amounts of experimental data. However, one of the major problems of such sequence search engines is that they return false positive results and researchers have to find ways to distinguish between correct and incorrect peptide identifications. In the case of small datasets, this can be achieved by manually verifying and validating the spectrum-to-peptide assignment. However, in the case of large datasets containing tens of thousands of spectra, such a time-consuming approach is not feasible. In these more and more contemporary cases, researchers separate correct from incorrect matches by using filtering criteria based upon database search scores or properties of the assigned peptide [44,69,70]. However, the number of rejected correct identifications (true negatives, sensitivity loss) and accepted false identifications (false positives, increased error rate) are not known applying such filtering criteria. Moreover, it remains elusive how those numbers are affected by sample preparation, mass spectrometer type and spectrum quality. Finally, these filtering criteria are differently applied from one laboratory to another, which renders comparison and cross-evaluation of the results difficult. Today, two main approaches have arisen to respond to this problem: the first is based on a robust and accurate statistical model to assess the validity of peptide identifications made by MS/MS and database search as exemplified by PeptideProphet and ProteinProphet in the Trans-Proteomic Pipeline [12,13]. Each spectrum-to-peptide assignment is evaluated with respect to

all other assignments in the datasets including necessarily some incorrect assignments. The method applies machine-learning techniques employing different parameters such as scores and peptide properties and it computes for each spectrum-to-peptide assignment a probability of being correct. The second strategy relies on database search using a target-decoy database [51,71–73]: first an appropriate “target” protein sequence database is generated and then, a “decoy” database preserving the general composition of the target database while minimizing the number of peptide sequences in common (generally done by reversing the target protein sequence) is created. The search is done against both the target and the composite database. Assuming that no correct peptides are found in both the target and decoy entities and that incorrect assignments from target or decoy sequences are equally likely, one can estimate the total number of false positives (FPs). In any case, neither one of the approaches can claim to remove all false positives but they both empower researchers to have a direct estimate of FPs according to the filtering criteria used, which ultimately helps sharing, comparing and publishing data in a more objective manner.

For an extensive review on this topic, readers are referred to the following publications [74,75].

---

## 5. Quantification approaches and tools

As previously described, a large toolbox of analytical methods, instruments and algorithms is available to identify and characterize proteins. However, the mere identification of a protein expressed in a biological system is not sufficient to answer most biological questions because quantitative answers are more and more required (e.g. comparison of protein expression level between to biological conditions (control vs. case).

Quantitative information comes in two forms, the relative change in protein amount between two states or the absolute amount of the protein in a sample determined, for example, in  $\text{ng ml}^{-1}$  of a biomarker or the copy number of a protein per cell. Relative quantification defines the amount of a protein relative to another measure of the same protein in another state (e.g. protein expression changes after drug treatment). In principle, absolute quantification encompasses both absolute and relative comparisons, if the absolute amounts of the proteins are known for both samples.

All current proteomic methods that quantify unknown proteins are relative methods. Absolute quantification of proteins can be achieved with isotopically labeled synthetic peptides and mass spectrometry, a method known as AQUA [76]. The availability of such peptide in accurately known amounts is a limitation to the widespread utilization of this method. An elegant way to partially overcome this limitation is the QCAT [77] or QconCAT [78] technique in which an artificial gene is used for expression, labeling and purification of a corresponding artificial protein, which represents a concatemer of tryptic peptides for several known proteins (Table 1). However such method requires the foreknowledge of the target proteins and preparation of isotopically labeled synthetic peptide for each targeted protein. Because of the Boolean nature of MS peptide detection, so called “proteotypic” peptides – peptides optimally representing the protein in terms of chromatography and MS behavior and being a

**Table 1 – Summary of currently available gel-free strategies for quantitative proteomics**

Target	Name of method or reagent	Isotopes	References
<i>Metabolic stable-isotope labeling</i>			
None	<sup>15</sup> N-labeling ( <sup>15</sup> N-ammonium salt)	<sup>15</sup> N	[82,83]
	Stable isotope labeling by amino acids in cell culture (SILAC)	D, <sup>13</sup> C, <sup>15</sup> N	[84]
	Culture-derived isotope tags (CDIT)	D, <sup>13</sup> C, <sup>15</sup> N	[90]
	Bioorthogonal noncanonical amino acid tagging (BONCAT)	No isotope	[91]
<i>Isotope tagging by chemical reaction</i>			
Sulfhydryl	Isotope-coded affinity tagging (ICAT)	D	[92]
	Cleavable ICAT	<sup>13</sup> C	[141–143]
	Catch-and-release (CAR)	<sup>13</sup> C	[93]
	Acrylamide	D	[144]
	Isotope-coded reduction off of a chromatographic support (ICROC)	D	[145]
	2-vinyl-pyridine	D	[146]
	N-t-butyliodoacetamide	D	[147]
	Iodoacetanilide	D	[147]
	HysTag	D	[148,149]
	Solid-phase ICAT	D	[150]
	Visible isotope-coded affinity tags (VICAT)	<sup>13</sup> C, <sup>14</sup> C and <sup>15</sup> N	[151]
	Acid-labile isotope-coded extractants (ALICE)	D	[152]
	Solid phase mass tagging	<sup>13</sup> C	[153]
Amines	Tandem mass tag (TMT)	D	[154]
	Succinic anhydride	D	[155]
	N-acetoxysuccinamide	D	[156]
	N-acetoxysuccinamide: In-gel Stable-Isotope Labeling (ISIL)	D	[157]
	Acetic anhydride	D	[158]
	Propionic anhydride	D	[159]
	Nicotinoyloxy succinimide (Nic-NHS)	D	[160]
	Isotope-coded protein labeling (ICPL,Nic-NHS)	D	[94]
	Phenyl isocyanate	D or <sup>13</sup> C	[161]
	Isotope-coded n-terminal sulfonation (ICens) 4-sulphophenyl isothiocyanate (SPITC)	<sup>13</sup> C	[162]
	Sulfo-NHS-SS-biotin and <sup>13</sup> C, <sup>3</sup> D <sub>3</sub> -methyl iodide	<sup>13</sup> C and D	[163]
	Formaldehyde	D	[164]
	Isobaric tag for relative and absolute quantification (iTRAQ)	<sup>13</sup> C, <sup>15</sup> N and <sup>18</sup> O	[95]
	Benzoic acid labeling (BA part of ANIBAL)	<sup>13</sup> C	[98]
Lysines	Guanidination (O-methyl-isourea) mass-coded abundance tagging (MGAT)	No isotope	[165]
	Guanidination (O-methyl-isourea)	<sup>13</sup> C and <sup>15</sup> N	[166,167]
	Quantitation using enhanced sequence tags (QUEST)	No isotope	[168]
	2-Methoxy-4,5-1H-imidazole	D	[169]
N-terminusprotein	Differentially isotope-coded N-terminal protein sulphonation (SPITC)	<sup>13</sup> C	[170]
N-terminuspeptide	N-terminal stable-isotope labelling of tryptic peptides (pentafluorophenyl-4-anilino-4-oxobutanoate)	D or <sup>13</sup> C	[171]
Carboxyl	Methyl esterification	D	[96]
	Ethyl esterification	D	[97]
	C-terminal isotope-coded tagging using sulfanilic acid (SA)	<sup>13</sup> C	[172]
	Aniline labeling (ANI part of ANIBAL)	<sup>13</sup> C	[98]
Indole	2-nitrobenzenesulfonyl chloride (NBSCI)	<sup>13</sup> C	[173]
<i>Stable-isotope incorporation via enzyme reaction</i>			
C-terminuspeptide	Proteolytic <sup>18</sup> O-labeling (H <sub>2</sub> <sup>18</sup> O)	<sup>18</sup> O	[99, 100, 104, 174]
	Quantitative cysteinyl-peptide enrichment technology (QCET)	<sup>18</sup> O	[175]
<i>Absolute quantification</i>			
None	Absolute quantification (AQUA)	D, <sup>13</sup> C, <sup>15</sup> N	[76]
	Multiplexed absolute quantification (QCAT)	D, <sup>13</sup> C, <sup>15</sup> N	[77]
	Multiplexed absolute quantification using concatenated signature (QconCAT)	D, <sup>13</sup> C, <sup>15</sup> N	[78]
	Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA)	D, <sup>13</sup> C, <sup>15</sup> N	[176]
<i>Label-free quantification</i>			
None	XIC-based quantification	No isotope	[112,113]
	Spectrum sampling (SpS)	No isotope	[108,177]
	Protein abundance index (PAI)	No isotope	[109]
	Exponentially modified protein abundance index (emPAI)	No isotope	[111]
	Probabilistic peptide scores (PMSS)	No isotope	[178]

unique identifier – have to be synthesized to ensure absolute quantification [79,80]. However, use of this technique for proteome/protein discovery approaches is precluded as previous knowledge of the sample is needed to synthesize all these peptides.

Initially, quantitative and comparative proteome analysis was performed with 2D-PAGE. The staining pattern of proteins from two samples run separately on different gels (with replicate gels for each sample) are compared and “up-“ or “down-regulated” proteins identified by difference in protein spot intensities. A rough estimate of the relative abundances of each protein within the sample is obtained. However, as already described in Section 5, 2D-PAGE has limitations such as low resolution, low dynamic range, bias against categories of proteins (e.g., membrane proteins) and low reproducibility of the gels. The last limitation has been largely solved by differential imaging gel electrophoresis (DIGE) where control and case samples are labeled with different fluorescent dyes and then run and compared on the same gel [81]. While still producing excellent results, “gel-based” quantitative proteomics has been largely superseded by “gel-free” MS-based quantitative proteomics approaches where quantification is performed using the mass spectrometric data. Similar to the gel approach, where *per se* the protein staining intensity within a gel is not proportional to the amount present within sample, in both MALDI- and ESI-MS the relationship between the amount of protein present and the measured signal intensity is complex and incompletely understood. Moreover the reproducibility of a peptide/protein signal between different runs is also complex. Therefore mass spectrometers are inherently poor quantitative devices and techniques to alleviate this problem were needed.

A first solution came with the technique of stable-isotope dilution. This method makes use of the fact that pairs of chemically identical molecules (in this case peptide pairs), but with different stable-isotope composition ( $^{13}\text{C}$  instead of  $^{12}\text{C}$ ,  $^2\text{H}$  instead of  $^1\text{H}$ ,  $^{18}\text{O}$  instead of  $^{16}\text{O}$  or  $^{15}\text{N}$  instead of  $^{14}\text{N}$ ) can be differentiated in a mass spectrometer owing to their mass difference only. Thus the ratio of signal intensities for such peptide pairs should be a direct and accurate measure of the abundance ratio between the two peptides/proteins derived from two different biological conditions. Three main approaches exist today (Table 1), which are: (i) metabolic stable-isotope labeling, (ii) isotope tagging by chemical reaction and, (iii) stable-isotope incorporation via enzyme reaction.

One metabolic stable-isotope technique uses  $^{15}\text{N}$ -ammonium salt to achieve complete labeling of all amino acids within the cells — every observable peptide is therefore quantifiable [82,83]. However, there are two major caveats: (i) the mass difference introduced between labeled and unlabeled peptides depends on the amino acid sequence and peptide identification program that can handle this correctly are rare [75]; (ii) the need for highly enriched nitrogen to avoid partial labeling.

The second metabolic method is stable-isotope labeling by amino acids in cell culture (SILAC) [84], in which amino acids containing stable isotopes, like arginine with six  $^{13}\text{C}$  atoms, are supplied in growth media. Several amino acids have been used like leucine (deuterated form), which labels 70% of tryptic peptides [85], or simultaneously lysine and arginine, with subsequent tryptic digestion resulting in labeling of all peptides but the C-terminal peptide [86]. A principal advan-

tage of metabolic labeling over chemical labeling is the earliest possible introduction of the label into the live cells, immediate pooling of case and control and the concomitant reduction of parallel sample preparation bias. The absence of “harsh” chemistry and side reactions is also an advantage. While these methods can only be applied to cultured cells like bacteria or yeast, recently these organisms have in turn been fed to small multicellular organisms such as *Caenorhabditis elegans*, *Drosophila melanogaster* [87], plants [88] or even a rat by using  $^{15}\text{N}$ -labeled algae [89].

Even more promising is the pair-wise comparison between cultured cell lines and dissected tissues [90]. In this case, a cell line derived from the tissue in question is labeled with SILAC and then spiked into both tissue states (e.g. healthy vs. diseased tissue) to serve as an internal standard and independent reference for both conditions. Thus, if the two ratios (healthy tissue vs. internal standard and diseased tissue vs. internal standard) obtained with the internal standard are different, it directly reflects a change in protein expression between compared tissues. Finally, a recently introduced technique called bio-orthogonal non-canonical amino acid tagging (BONCAT) [91] allows to specifically identify newly synthesized proteins based on co-translational introduction of azide groups into proteins, chemoselective tagging of these azide-labeled proteins and subsequent capture with an affinity tag.

A wide variety of isotopically labeled chemicals has been reported (Table 1). All chemical reagents are targeted toward reactive sites on a protein or peptide and the two proteomes to be compared are labeled with the light and heavy reagent, respectively. Isotope-coded affinity tagging (ICAT) [92] was the first approach described in 1999 by Gygi and co-workers. The reagent consists of a reactive group that is cystein-directed, a polyether linker region with eight deuteria and a biotin group for avidin purification of labeled peptides. Due to compromised co-elution of deuterium-tagged and natural hydrogen peptides, and MS fragmentation problems (large tag) with this first ICAT version, a new version was developed with an acid cleavable site and  $^{13}\text{C}$  atoms instead. Recently, Gygi and colleagues have described a new method called catch-and-release (CAR) [93] that makes use of a cystein-directed reductively cleavable reagent. The tag features a novel disulfide moiety that links biotin and a thiol-reactive group. The disulfide is resistant to reductive conditions during labeling but readily cleaved with tris-(2-carboxyethyl) phosphine (TCEP), therefore simplifying sample handling procedures and reducing non-specific interactions during avidin purification.

Several strategies have been reported that target amines of which two have been applied to experimental biology. The first, isotope-coded protein labeling (ICPL) [94], targets all amino groups at the protein level using nicotinoyloxy succinimide (Nic-NHS) as the reagent. The second, isobaric tag for relative and absolute quantification (iTRAQ) [95], uses the same NHS chemistry as ICPL, but adds an innovative concept, namely a tag that generates a specific reporter ion for quantification in MS/MS spectra (mass 114, 115, 116, 117) but with isobaric mass at MS level. Therefore, mass spectra are relatively simple and differential behavior is only reported after fragmentation. Moreover, multiplexing (currently eight-

plex) is an interesting feature as it allows comparing more than two conditions.

Carboxylic groups have also been labeled using either methyl [96] or ethyl [97] esterification at the peptide level. However, both methods use deuterium atoms and bear the risk of chromatographic discrimination and the mass offsets of 2 Da (methyl) and 4 Da (ethyl) poses problems of isotopic overlap of the peptide pairs. Recently, our group introduced a new technique called ANIBAL [98] to label amino and carboxylic groups at the protein level using two symmetrical tags with 6 times  $^{13}\text{C}$  and alleviate the previously mentioned caveats.

A clear advantage of all these chemical approaches is the multitude of available functional groups in proteins allowing designing almost any kind of quantitative tag. Possible enrichment is also an asset as it allows reducing sample complexity without losing quantitative information. However, reactions have to be specific, proceed to completion and involve minimal sample handling. Side reactions are problematic, too, as they considerably increase the sample complexity. Despite these constraints, chemical stable-isotope labeling has produced most of the quantitative proteome data mainly due its chemical versatility and certainly because of its applicability to any biological sample as opposed to metabolic labeling.

Stable isotopes can also be introduced into the peptide by different proteases such as Trypsin, Lys-N or Glu-C [99–101]. The digestion is performed in  $\text{H}_2^{18}\text{O}$  water and enzymatic oxygen exchange occurs at the carboxyl group of the generated peptides. The advantage of this method is its versatility (virtually any protease-generated peptide is labeled), its applicability to low sample amounts and almost unlimited compatibility with sample preparations. On the other hand, the labeling is performed only at peptide level, and samples have to be processed in parallel until these peptide are generated. One or two oxygens can be exchanged leading to variability in peptide spacing and the mass offset of 2 Da is not sufficient to separate the isotopic envelopes. Recent modifications such as post-digestion incubation of peptides in small volumes of  $\text{H}_2^{18}\text{O}$  or deactivating the protease through reduction/alkylation have addressed these issues [102–104].

As outlined, a large number of stable-isotope-based quantification methods are available. However, if such experiment is undertaken, one needs to be able to quantify thousands of labeled peptides using automatic tools capable of extracting the intensity for both peptides of a pair and report a protein ratio based on all identified and quantified peptides. Even more important, such tools should be able to process data from different instrument manufacturers (compared to software specific to the instrument of acquisition) and should also be able to accept result input from different search engines (Mascot, SEQUEST, X!Tandem, OMSSA, Phenyx, etc.) for optimal sample analysis and comparison. Several open-source softwares have been developed by scientific laboratories performing large-scale quantitative proteomic experiments (Table 2) and in general each of them has been developed according to supported instruments or database search algorithms. Only the Trans-Proteomic Pipeline (TPP as previously described in Section 6 for statistical validation of database searches) supports almost all instrument types by transforming all proprietary formats to an open-source format – mzXML [105] – that can be read and processed by any computer. The

most frequently used algorithms for database search such as Mascot, SEQUEST, X!Tandem or Phenyx are supported as input for subsequent quantification. Two tools – Xpress [70] and ASAPRatio [106] – quantify at MS level and can cope with any labeling strategy for which labeled amino acids are specified and another tool – Libra – is available for quantification at MS/MS level as required for iTRAQ data.

Recently, new promising approaches have emerged that do not use labeling and stable isotopes to obtain quantitative information and these are described as “label-free”. A first strategy is based on peptide match score summation (PMSS) [107]. The method is based on the assumption that a protein score is a sum of identification scores of its peptides and that a high protein score is correlated with a higher abundance, thus yielding semi-quantitative information. Another very similar approach relies on the counting of spectra identifying a protein and is named spectrum sampling (SpS) [108]. Protein abundance indices (PAIs) represent another related method and are believed to be more reliable as they are based on observable parameters. For example, the number of peptides identifying a protein increases with increasing protein amount. As a larger protein will statistically generate more measurable peptides than a smaller one, a simple PAI can be derived by normalizing the number of observed peptides by the number of observable peptides for the protein under consideration [109,110]. Ishihama and colleagues have described an exponentially modified PAI (emPAI) by observing a logarithmic relationship between the number of peptides observed and the protein amount within given sample [111].

Unfortunately, it is not possible to predict the MS detector response to a particular peptide because of unknown extraction and peptide ionization properties and, therefore, extracted ion currents (XICs) from different peptides of the same protein are also very different even if they are present at the same concentration.

Although directly comparing intensities between different peptides is not possible for the reasons previously mentioned. These sources of error do not apply when comparing the same peptide in different chromatographic runs using identical experimental conditions. Thus two proteomes can be compared when analyzed one after the other and in exactly the same way [112,113]. Software exists to extract the intensities of the same peptide observed in two separate runs to compare and determine their relative abundance (e.g. MSight [114], SuperHirn [115], MapQuant [116], SpecArray [117] or VEMS [9–11]). A clear advantage of such method is the absence of any label and the applicability to any type of instrument. Clear disadvantages are the multiple occasions for quantification error to occur during parallel sample processing, analysis and the need for very accurate and reproducible LC and MS runs.

---

## 6. Proteomics —future directions

Proteomics will further progress and deliver as the related instrumentation advances, the analytical strategies mature and as it is combined with complementary technologies. The methods for biochemical sample preparation and fractionation (depletion, enrichment, gel-, LC- and GeLC separation) are manifold and already quite efficient. Mass spectrometers have

**Table 2 – Summary of open-source softwares available for quantification in stable-isotope-based quantification experiments**

Name	MS level	Environment	Reference	Comment
MSquant	MS <sup>1</sup>	Windows	[179]	Support QSTAR, LTQ-FT & Qtof instrument using a Mascot input file. Web: <a href="http://msquant.sourceforge.net/">http://msquant.sourceforge.net/</a>
ZoomQuant	MS <sup>1</sup>	Windows	–	Developed for <sup>18</sup> O labeling and support Sequest search with Thermo instrument. Web: <a href="http://proteomics.mcw.edu/zoomquant/">http://proteomics.mcw.edu/zoomquant/</a>
TOPP	MS <sup>1</sup>	Linux	[180]	A series of module to analyse proteomics data and perform quantification. Support most instrument format, mzXML and mzData. Only Mascot supported for the moment. Web: <a href="http://open-ms.sourceforge.net/index.php">http://open-ms.sourceforge.net/index.php</a>
Xpress	MS <sup>1</sup>	Windows (cygwin) and Linux	[70]	
ASAPRatio	MS <sup>1</sup>	Windows (cygwin) and Linux	[106]	Part of the Trans-Proteomic Pipeline (TPP). Support almost any instrument type. Mascot, SEQUEST, X!Tandem and Phenyx results are supported as input. Web: <a href="http://tools.proteomecenter.org/XPRESS.php">http://tools.proteomecenter.org/XPRESS.php</a> . Web: <a href="http://tools.proteomecenter.org/ASAPRatio.php">http://tools.proteomecenter.org/ASAPRatio.php</a> . Web: <a href="http://tools.proteomecenter.org/Libra.php">http://tools.proteomecenter.org/Libra.php</a>
Libra	MS <sup>2</sup>	Windows (cygwin) and Linux	–	
i.tracker	MS <sup>2</sup>	Window and Linux	[181]	Takes MGF or OUT files as only input for quantification. Ratio can then be linked to Mascot or SEQUEST results. Web: <a href="http://www.silsoe.cranfield.ac.uk/dasi/download/itracker.htm">http://www.silsoe.cranfield.ac.uk/dasi/download/itracker.htm</a>
MFPaQ	MS <sup>1</sup>	Windows	[182]	Takes mascot (DAT) result files as input for parsing and Analyst Wiff files for quantification. Web: <a href="http://mfpaq.sourceforge.net/">http://mfpaq.sourceforge.net/</a>
RelEx	MS <sup>1</sup>	Windows and Linux	[183]	Needs Thermo files, Xcalibur installed and DTASelect to extract data in the correct format for RelEx. Web: <a href="http://fields.scripps.edu/relex/">http://fields.scripps.edu/relex/</a>
Multi-Q	MS <sup>2</sup>	Windows	[184]	Uses mzXML files from WIFF, RAW and BAF files with Mascot, Sequest and X!Tandem result files to create a Multi-Q file for further validation and quantification. Web: <a href="http://ms.iis.sinica.edu.tw/Multi-Q/index.jsp">http://ms.iis.sinica.edu.tw/Multi-Q/index.jsp</a>
ProRata	MS <sup>1</sup>	Windows and Linux	[185,186]	Uses mzXML files and Sequest DTASelect files for processing and quantification Web: <a href="http://www.msprorata.org/">http://www.msprorata.org/</a>
VEMS	MS <sup>1</sup>	Windows	[9–11]	Supports Micro Mass, mzXML and VEMS raw format to perform database search, processing and quantification. Web: <a href="http://yass.sdu.dk/">http://yass.sdu.dk/</a>

seen quantum leaps in sensitivity and accuracy and more of these can be expected. Informatics for data processing has evolved and will further improve to tools not only for data processing, but also for data validation and correlation. This said the complexity and dynamic range of a mammalian proteome can most likely not be met solely by developments at technical level.

We believe that proteomics will largely benefit from developments beyond technical improvements of its essential building blocks, i.e. separation means, mass spectrometers and software. These developments mean essentially the integration of proteomics with other comprehensive molecular biology disciplines and should encompass (i) improved analytical strategies; (ii) the integration of proteomics with transcriptomics and metabolomics; (iii) the consideration of genetics in Omics-driven biomarker studies; and (iv) the consideration of regulation at transcriptional, translational, post-translational and at (v) epigenetic level.

- (i) The establishment and use of indicative cellular and animal models and specific samples derived thereof is key for the initial discovery phase [4]. We need downstream assays for *ex vivo* and *in vivo* validation, that are less broad but more sensitive and specific and can therefore afford to be less invasive than proteomics-based methods. This translation will determine much of

the success of proteomics for future biomarker studies. In other words: the earlier the study phase, the more invasive one can and the closer to the expected signal one should be (e.g. by analyzing tissues), whereas candidate markers derived from this phase must be validated in robust assays performed on readily accessible samples (e.g. urine or plasma) [4].

Besides these choices of model, subject and sample and a separation of the discovery from the validation phase, a paradigm change on how to perform the proteomic experiment emerges: Anderson [118] and also Aebersold [80] promote a multiple-reaction-monitoring (MRM) strategy of protein quantification focused on a defined proteome subset and based on proteotypic peptides (PTPs). This strategy, also termed “multiplexed MS-ELISA” because of its specificity and sensitivity for a target set of proteins [36], is opposed to the classical “shotgun” way of identifying and quantifying as many proteins as possible. The latter has often turned to be redundant and under-sampling, and may therefore not be suitable for a near-to-complete coverage of a given proteome. Classical, high-performance shotgun proteomics can deliver up to 2000 or more protein identities, a number still below the one of predicted proteins of a microbe for example. While being less comprehensive, the MRM-PTP strategy appears to be more sensitive: low

attomolar amounts of proteins, or – in other terms – 50–250 copies per cell (depending on the degree of fractionation prior to LC-MS/MS) appear to be amenable to identification and quantification in complex samples [119]. But the “MS-ELISA” comes at the expense of being a targeted approach, which depends on stable-isotope labeled peptides as internal standards for each protein of interest. In a near future, one could imagine that a complete understanding of what makes a peptide proteotypic, as already started and described by Mallick et al. [80], should allow *in silico* prediction of all these peptides from gene/protein sequences. Therefore, those targeted approaches could be transformed back into discovery strategies by synthesizing whole stable-isotope peptide libraries of predicted peptides (e.g. AQUA peptide libraries [76]).

- (ii) Apart from improvements directly related to the technology itself, the success of proteomics will in our opinion largely depend on the integration with other Omics platforms, i.e. transcriptomics and metabolomics [19] and genotyping [120]. Seeking causality between events at gene and protein expression and metabolic breakdown level can provide deeper insights into mechanisms and may provide more consolidated biomarker profiles by spanning all three Omic levels. While such Omics correlation is appealing and an element of Systems Biology (see below), the challenge is how to correlate transcript, protein and metabolite events, the interrelated timing of which is difficult to dissect: depending on the biological question under scrutiny, it may not be obvious how to best time the sampling for transcriptomics, proteomics and metabolomics, respectively [121]. Furthermore, software tools for correlating Omics data still need to mature, although there are promising examples (e.g. SBEAMS [<http://www.sbeams.org/>]). A recent example of how bioinformatics can identify regulatory nodes in networks that were neither detected at transcript nor at protein level was given by Grafström et al.: the group investigated epithelial cell lines as models for different stages of cancer and the bioinformatics correlation of transcript and protein data revealed a transcription factor as common regulatory motif explaining the observed changes at gene chip and proteomic level [122].
- (iii) A third means of improving the diagnostic and prognostic power of proteomic biomarkers is, besides the above discussed technology improvements and Omics integration, the consideration of genetic disposition. Single-nucleotide polymorphisms (SNPs) can directly alter protein concentration, structure and function, if they are present in protein-coding gene sequences. This shall be illustrated with a nutritionally relevant example: Siffert et al. have identified and characterized metabolically relevant SNPs in G proteins, the latter representing an important “funnel” of cellular signaling. These polymorphisms predispose individuals of different ethnicity to having a higher risk of developing hypertension, atherosclerosis, metabolic syndrome or functional dyspepsia [123,124]. As a consequence, nutritional intervention studies to prevent or alleviate

symptoms of these conditions should take this genetic disposition into account for subject selection, cohort definition and monitoring the outcome of such a study at different Omic levels. Also in the field of direct DNA analysis, as exemplified with SNPs, mass spectrometry is becoming an important analytical player due to the high-throughput capabilities of such instruments [125,126].

- (iv) While genetics will further empower Omics-driven approaches to better understand biology, it is important to admit at this stage, that the various other levels of biological regulation have not yet been included in this consideration: apart from transcriptional regulation, translational and post-translational contributions add complexity to the understanding of the dynamic behavior of a cell for instance. Addressing this problem, Lu et al. recently published an absolute protein expression profiling study, based on which they could estimate the relative contributions of transcriptional and translational gene regulation in the yeast proteome [127]. Regulatory levels that are not genetically encoded like post-translational modifications (PTMs) such as protein phosphorylation and glycosylation cannot be addressed by genomic means but need to be tackled by large-scale biochemical techniques, most of which rely on mass spectrometry. While examples for large-scale MS studies of phosphorylation-based regulation exist [34,128,129], the disciplines glycomics and glycoproteomics have not yet achieved comparable throughput as the analysis of the enormous variety of glycans (both free and protein-bound) is analytically more challenging [130,131]. N-glycans encompass a large variety of complex structures and render structure elucidation difficult, and O-glycans are often subjected to fast changes much like phosphorylation.
- (v) Apart from modifications of expressed proteins, epigenetic regulation like DNA methylation (gene silencing) and histone acetylation (chromatin structure) should ideally be included in Systems Biology studies, as these mechanisms strongly influence gene transcription and expression. Remarkably, MS-based proteomic methods have started to contribute also in this regard: the Jensen group presented a quantitative analysis of human histone PTMs [132] whereas Bonenfant et al. focused on the histone codes of H2A and H2B variants [133].

Transcriptomics, Proteomics and Metabolomics and their correlation form an integral data source for Systems Biology. The term Systems Biology remains, however, loosely defined [134]. In our opinion, Systems Biology is the comprehensive understanding of all components, their interactions and dynamic behavior in a biological system, be it a cell, an organ, an organism or even an ecosystem, like the human intestine with its 500 to 1500 bacterial strains residing inside [135]. While Systems Biology approaches typically study these systems under different conditions/perturbations in order to learn more about their dynamic response, the data describing this behavior should ultimately enable the modeling of such a system and the prediction of its reaction towards external stimuli. Impressive examples of such models are the E-Cell, i.e. the *in silico* model of a cell [136–138] and the establishment

of an *in silico* network of neurons representing a simplified model of the brain [139]. In the latter case, neurons are modeled based on e.g. *in vitro* data of ion channel expression levels because these ion channels are major players in signal exchange between neurons.

Proteomics will continue to play a major role in Systems Biology, as it not only can identify and quantify the “molecular robots” that do all the work in biological systems, but also can map the networks of their physical interactions, among each other and with nutrients, drugs and other small molecules. An impressive example of such a thematic network establishment has been given by Bantscheff et al.: the group revealed mechanisms of action of clinical kinase inhibitors by mass spectrometric profiling of small-molecule interactions with hundreds of endogenously expressed protein kinases [140].

One cannot study everything at once. In order to stay pragmatic, Systems Biology, despite being per definition comprehensive, has to deal with biological (sub-)systems (cells, organelles) or certain pathways and must elucidate aspects of their dynamic behavior in order to further deliver proofs of concept at somewhat limited complexity. However, we believe that integrated Omics linked with genetics can already provide systems-like insights into biological contexts and provides a reasonable data source for the developing field of Systems Biology.

## REFERENCES

- [1] Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 1996;13:19–50.
- [2] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
- [3] Rose K, Bougueleret L, Baussant T, Bohm G, Botti P, Colinge J, et al. Industrial-scale proteomics: from liters of plasma to chemically synthesized proteins. *Proteomics* 2004;4:2125–50.
- [4] Lescuyer P, Hochstrasser D, Rabilloud T. How shall we use the proteomics toolbox for biomarker discovery? *J Proteome Res* 2007;6:3371–6.
- [5] Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham CR. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 2005;40:430–43.
- [6] Colinge J, Masselot A, Giron M, Dessingy T, Magnin J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* 2003;3:1454–63.
- [7] Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. *J Proteome Res* 2004;3:958–64.
- [8] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004;20:1466–7.
- [9] Matthiesen R, Bunkenborg J, Stensballe A, Jensen ON, Welinder KG, Bauw G. Database-independent, database-dependent, and extended interpretation of peptide mass spectra in VEMS V2.0. *Proteomics* 2004;4:2583–93.
- [10] Matthiesen R, Lundsgaard M, Welinder KG, Bauw G. Interpreting peptide mass spectra by VEMS. *Bioinformatics* 2003;19:792–3.
- [11] Matthiesen R, Trelle MB, Hojrup P, Bunkenborg J, Jensen ON. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res* 2005;4:2338–47.
- [12] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74:5383–92.
- [13] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–58.
- [14] Alaiya A, Al-Mohanna M, Linder S. Clinical cancer proteomics: promises and pitfalls. *J Proteome Res* 2005;4:1213–22.
- [15] Wang MC, Papsidero LD, Kuriyama M, Valenzuela LA, Murphy GP, Chu TM. Prostate antigen: a new potential marker for prostatic cancer. *Prostate* 1981;2: 89–96.
- [16] Amacher DE. Serum transaminase elevations as indicators of hepatic injury following the administration of drugs. *Regul Toxicol Pharmacol* 1998;27:119–30.
- [17] Antman EM, Tanasijevic MJ, Thompson B, Schactman M, McCabe CH, Cannon CP, et al. Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. *N Engl J Med* 1996;335:1342–9.
- [18] Hsieh G, Kenney K, Gibbs CJ, Lee KH, Harrington MG. The 14-3-3 brain protein in cerebrospinal fluid as a marker for transmissible spongiform encephalopathies. *N Engl J Med* 1996;335:924–30.
- [19] Kussmann M, Raymond F, Affolter M. OMICS-driven biomarker discovery in nutrition and health. *J Biotechnol* 2006;124:758–87.
- [20] Kussmann M, Affolter M, Fay LB. Proteomics in nutrition and health. *Comb Chem High Throughput Screen* 2005;8:679–96.
- [21] Kussmann M, Affolter M, Nagy K, Holst B, Fay LB. Mass spectrometry in nutrition: understanding dietary health effects at the molecular level. *Mass Spectrom Rev* 2007.
- [22] Kussmann M, Affolter M. Proteomic methods in nutrition. *Curr Opin Clin Nutr Metab Care* 2006;9:575–83.
- [23] Mischak H, Apweiler R, Banks RE, Conaway M, Coon J, Dominiczak A, et al. Clinical proteomics: a need to define the field and to begin to set adequate standards. *PROTEOMICS-Clinical Applications* 2007;1:148–56.
- [24] Kussmann M, Blum S. OMICS-derived targets for inflammatory gut disorders: opportunities for the development of nutrition related biomarkers. *Endocr Metab Immune Disord Drug Targets* 2007;7:271–87.
- [25] Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 1999;19: 1720–30.
- [26] Teufel A, Krupp M, Weinmann A, Galle PR. Current bioinformatics tools in genomic biomedical research. (review) *Int J Mol Med* 2006;17:967–73.
- [27] Pennington SR, Wilkins MR, Hochstrasser DF, Dunn MJ. Proteome analysis: from protein characterization to biological function. *Trends Cell Biol* 1997;7:168–73.
- [28] O'Donovan C, Apweiler R, Bairoch A. The human proteomics initiative (HPI). *Trends Biotechnol* 2001;19:178–81.
- [29] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–67.
- [30] Domon B, Broder S. Implications of new proteomics strategies for biology and medicine. *J Proteome Res* 2004;3:253–60.
- [31] Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–70.
- [32] Raymond F, Metairon S, Borner R, Hofmann M, Kussmann M. Automated target preparation for microarray-based gene expression analysis. *Anal Chem* 2006;78:6299–305.
- [33] MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000;289: 1760–3.
- [34] Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, et al. Global analysis of protein phosphorylation in yeast. *Nature* 2005;438:679–84.

- [35] Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. *Proteomics* 2003;3: 1397–407.
- [36] Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, Aebersold R, et al. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 2007;6:1809–17.
- [37] Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* 2007;4:231–7.
- [38] Taylor SW, Fahy E, Ghosh SS. Global organellar proteomics. *Trends Biotechnol* 2003;21:82–8.
- [39] Taylor SW, Fahy E, Zhang B, Glenn GM, Warnock DE, Wiley S, et al. Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* 2003;21: 281–6.
- [40] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207.
- [41] Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 2000;72:563–73.
- [42] Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 2004;101:9528–33.
- [43] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2299–301.
- [44] Washburn MP, Wolters D, Yates III JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–7.
- [45] Righetti PG, Boschetti E, Lomas L, Citterio A. Protein equalizer technology: the quest for a “democratic proteome”. *Proteomics* 2006;6:3980–92.
- [46] Righetti PG, Boschetti E. Sherlock Holmes and the proteome—a detective story. *FEBS J* 2007;274:897–905.
- [47] Righetti PG, Castagna A, Antoniolli P, Boschetti E. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* 2005;26:297–319.
- [48] Zhen Y, Xu N, Richardson B, Becklin R, Savage JR, Blake K, et al. Development of an LC-MALDI method for the analysis of protein complexes. *J Am Soc Mass Spectrom* 2004;15:803–22.
- [49] Pappin DJ. Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol Biol* 1997;64:165–73.
- [50] Schuerenberg M, Luebbert C, Eickhoff H, Kalkum M, Lehrach H, Nordhoff E. Prestructured MALDI-MS sample supports. *Anal Chem* 2000;72:3436–42.
- [51] Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–67.
- [52] Gras R, Muller M, Gasteiger E, Gay S, Binz PA, Bienvenut W, et al. Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* 1999;20:3535–50.
- [53] Zhang W, Chait BT. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* 2000;72:2482–9.
- [54] Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984;11:601.
- [55] Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem* 1987;59:2621–5.
- [56] Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 1994;66:4390–9.
- [57] Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* 1993;22:338–45.
- [58] Tabb DL, Saraf A, Yates III JR. GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 2003;75:6415–21.
- [59] Sunyaev S, Liska AJ, Golod A, Shevchenko A, Shevchenko A. MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry. *Anal Chem* 2003;75:1307–15.
- [60] Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–89.
- [61] Hansen BT, Jones JA, Mason DE, Liebler DC. SALSA: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal Chem* 2001;73:1676–83.
- [62] Lu B, Chen T. Algorithms for *de novo* peptide sequencing using tandem mass spectrometry. *Drug Discov Today: BIOSILICO* 2004;2:85–90.
- [63] Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 2007;7:655–67.
- [64] Zubarev RA, Hakansson P, Sundqvist B. Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Anal Chem* 1996;68:4060–3.
- [65] Conrads TP, Anderson GA, Veenstra TD, Pasa-Tolic L, Smith RD. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem* 2000;72:3349–54.
- [66] Cargile BJ, Stephenson Jr JL. An alternative to tandem mass spectrometry: isoelectric point and accurate mass for the identification of peptides. *Anal Chem* 2004;76:267–75.
- [67] Palmblad M, Ramstrom M, Bailey CG, McCutchen-Maloney SL, Bergquist J, Zeller LC. Protein identification by liquid chromatography-mass spectrometry using retention time prediction. *J Chromatogr B Analyt Technol Biomed Life Sci* 2004;803:131–5.
- [68] Norbeck AD, Monroe ME, Adkins JN, Anderson KK, Daly DS, Smith RD. The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J Am Soc Mass Spectrom* 2005;16: 1239–49.
- [69] Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17: 676–82.
- [70] Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 2001;19:946–51.
- [71] Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003;2:43–50.
- [72] Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2005;2:667–75.
- [73] Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4:207–14.
- [74] Lisacek F, Cohen-Boulakia S, Appel RD. Proteome informatics II: bioinformatics for comparative proteomics. *Proteomics* 2006;6:5445–66.
- [75] Matthiesen R. Methods, algorithms and tools in computational proteomics: a practical point of view. *Proteomics* 2007;7: 2815–32.
- [76] Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins

- from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003;100:6940–5.
- [77] Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* 2005;2: 587–9.
- [78] Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protocols* 2006;1: 1029–43.
- [79] Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol* 2005;6:577–83.
- [80] Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 2007;25:125–31.
- [81] Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997;18:2071–7.
- [82] Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 1999;96:6591–6.
- [83] Conrads TP, Alving K, Veenstra TD, Belov ME, Anderson GA, Anderson DJ, et al. Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and <sup>15</sup>N-metabolic labeling. *Anal Chem* 2001;73:2132–9.
- [84] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1: 376–86.
- [85] Foster LJ, De Hoog CL, Mann M. Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci U S A* 2003;100:5813–8.
- [86] Ibarrola N, Kalume DE, Gronborg M, Iwahori A, Pandey A. A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Anal Chem* 2003;75:6043–9.
- [87] Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, et al. Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* 2003;21:927–31.
- [88] Ippel JH, Pouvreau L, Kroef T, Gruppen H, Versteeg G, van den Putten P, et al. *In vivo* uniform (<sup>15</sup>N)-isotope labelling of plants: using the greenhouse for structural proteomics. *Proteomics* 2004;4:226–34.
- [89] Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates III JR. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* 2004;76:4951–9.
- [90] Ishihama Y, Sato T, Tabata T, Miyamoto N, Sagane K, Nagasu T, et al. Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat Biotechnol* 2005;23:617–21.
- [91] Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* 2006;103:9482–7.
- [92] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–9.
- [93] Gartner CA, Elias JE, Bakalarski CE, Gygi SP. Catch-and-release reagents for broadscale quantitative proteomics analyses. *J Proteome Res* 2007;6:1482–91.
- [94] Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels (vol. 5, Issue 1, pp. 4–15). *Proteomics* 2005;5:826.
- [95] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004;3:1154–69.
- [96] Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, et al. Differential stable isotope labeling of peptides for quantitation and *de novo* sequence derivation. *Rapid Commun Mass Spectrom* 2001;15:1214–21.
- [97] Syka JE, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res* 2004;3:621–6.
- [98] Panchaud A, Hansson J, Affolter M, Bel Rhlid R, Piu S, Moreillon P, et al. ANIBAL — stable-isotope-based quantitative proteomics by ANline and Benzoic acid labeling of amino and carboxylic groups. *Mol Cell Proteomics* in press; M700216–MCP200.
- [99] Mirgorodskaya OA, Kozmin YP, Titov MI, Korner R, Sonksen CP, Roepstorff P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (<sup>18</sup>O)-labeled internal standards. *Rapid Commun Mass Spectrom* 2000;14:1226–32.
- [100] Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic <sup>18</sup>O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 2001;73:2836–42.
- [101] Rao KC, Palamalai V, Dunlevy JR, Miyagi M. Peptidyl-Lys metalloendopeptidase-catalyzed <sup>18</sup>O labeling for comparative proteomics: application to cytokine/lipopolysaccharide-treated human retinal pigment epithelium cell line. *Mol Cell Proteomics* 2005;4:1550–7.
- [102] Bantscheff M, Dimpfelfeld B, Kuster B. Femtomol sensitivity post-digest (<sup>18</sup>O) labeling for relative quantification of differential protein complex composition. *Rapid Commun Mass Spectrom* 2004;18:869–76.
- [103] Staes A, Demol H, Van Damme J, Martens L, Vandekerckhove J, Gevaert K. Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-<sup>18</sup>. *J Proteome Res* 2004;3:786–91.
- [104] Miyagi M, Rao KC. Proteolytic <sup>18</sup>O-labeling strategies for quantitative proteomics. *Mass Spectrom Rev* 2007;26:121–36.
- [105] Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 2004;22:1459–66.
- [106] Li XJ, Zhang H, Ranish JA, Aebersold R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem* 2003;75:6648–57.
- [107] Allet N, Barrillat N, Baussant T, Boiteau C, Botti P, Bougueleret L, et al. *In vitro* and *in silico* processes to identify differentially expressed proteins. *Proteomics* 2004;4:2333–51.
- [108] Liu H, Sadygov RG, Yates III JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76:4193–201.
- [109] Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 2002;12: 1231–45.
- [110] Sanders SL, Jennings J, Canutescu A, Link AJ, Weil PA. Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIIID by multidimensional mass spectrometry. *Mol Cell Biol* 2002;22:4723–38.
- [111] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in

- proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005;4:1265–72.
- [112] Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* 2002;1:317–23.
- [113] Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, et al. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 2002;419:537–42.
- [114] Palagi PM, Walther D, Quadroni M, Catherinet S, Burgess J, Zimmermann-Ivol CG, et al. MSight: an image analysis software for liquid chromatography-mass spectrometry. *Proteomics* 2005;5:2381–4.
- [115] Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, et al. SuperHirm — a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 2007.
- [116] Leptos KC, Sarracino DA, Jaffe JD, Krastins B, Church GM. MapQuant: open-source software for large-scale protein quantification. *Proteomics* 2006;6:1770–82.
- [117] Li XJ, Yi EC, Kemp CJ, Zhang H, Aebersold R. A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol Cell Proteomics* 2005;4:1328–40.
- [118] Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 2006;5:573–88.
- [119] Picotti P, Aebersold R, Domon B. The implications of proteolytic background for shotgun proteomics. *Mol Cell Proteomics* 2007;6:1589–98.
- [120] Kussmann M. How to comprehensively analyse proteins and how this influences nutritional research. *Clin Chem Lab Med* 2007;45:288–300.
- [121] Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 2002;1:153–61.
- [122] Staab CA, Ceder R, Jagerbrink T, Nilsson JA, Roberg K, Jornvall H, et al. Bioinformatics processing of protein and transcript profiles of normal and transformed cell lines indicates functional impairment of transcriptional regulators in buccal carcinoma. *J Proteome Res* 2007;6:3705–17.
- [123] Holtmann G, Siffert W, Haag S, Mueller N, Langkafel M, Senf W, et al. G-protein beta 3 subunit 825 CC genotype is associated with unexplained (functional) dyspepsia. *Gastroenterology* 2004;126:971–9.
- [124] Siffert W. G protein polymorphisms in hypertension, atherosclerosis, and diabetes. *Annu Rev Med* 2005;56:17–28.
- [125] Jurinke C, Oeth P, van den Boom D. MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. *Mol Biotechnol* 2004;26:147–64.
- [126] Ragoussis J, Elvidge GP, Kaur K, Colella S. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet* 2006;2:e100.
- [127] Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 2007;25:117–24.
- [128] Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004;101:12130–5.
- [129] Nuhse TS, Stensballe A, Jensen ON, Peck SC. Large-scale analysis of *in vivo* phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2003;2:1234–43.
- [130] Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* 2005;2:817–24.
- [131] Sun B, Ranish JA, Utleg AG, White JT, Yan X, Lin B, et al. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics* 2007;6:141–9.
- [132] Beck HC, Nielsen EC, Matthiesen R, Matthiesen R, Jensen LH, Sehested M, Finn P, et al. Quantitative proteomic analysis of post-translational modifications of human histones. *Mol Cell Proteomics* 2006;5:1314–25.
- [133] Bonenfant D, Coulot M, Towbin H, Schindler P, van Oostrum J. Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Mol Cell Proteomics* 2006;5:541–52.
- [134] Lisacek F, Appel RD. Systems biology: a loose definition. *Proteomics* 2007;7:825–7.
- [135] Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JL. Host-bacterial mutualism in the human intestine. *Science* 2005;307: 1915–20.
- [136] Takahashi K, Arjunan SN, Tomita M. Space in systems biology of signaling pathways-towards intracellular molecular crowding *in silico*. *FEBS Lett* 2005;579:1783–8.
- [137] Takahashi K, Ishikawa N, Sadamoto Y, Sasamoto H, Ohta S, Shiozawa A, et al. E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics* 2003;19:1727–9.
- [138] Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics* 1999;15: 72–84.
- [139] Markram H. The blue brain project. *Nat Rev, Neurosci* 2006;7: 153–60.
- [140] Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 2007;25: 1035–44.
- [141] Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. Mass spectrometric analysis of protein mixtures at low levels using cleavable <sup>13</sup>C-isotope-coded affinity tag and multidimensional chromatography. *Mol Cell Proteomics* 2003;2:299–314.
- [142] Li J, Steen H, Gygi SP. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics* 2003;2: 1198–204.
- [143] Oda Y, Owa T, Sato T, Boucher B, Daniels S, Yamanaka H, et al. Quantitative chemical proteomics for identifying candidate drug targets. *Anal Chem* 2003;75:2159–65.
- [144] Sechi S, Chait BT. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal Chem* 1998;70:5150–8.
- [145] Shen M, Guo L, Wallace A, Fitzner J, Eisenman J, Jacobson E, et al. Isolation and isotope labeling of cysteine- and methionine-containing tryptic peptides: application to the study of cell surface proteolysis. *Mol Cell Proteomics* 2003;2:315–24.
- [146] Sebastiano R, Citterio A, Lapadula M, Righetti PG. A new deuterated alkylating agent for quantitative proteomics. *Rapid Commun Mass Spectrom* 2003;17:2380–6.
- [147] Pasquarello C, Sanchez JC, Hochstrasser DF, Corthals GL. N-t-butyliodoacetamide and iodoacetanilide: two new cysteine alkylating reagents for relative quantitation of proteins. *Rapid Commun Mass Spectrom* 2004;18:117–27.
- [148] Olsen JV, Andersen JR, Nielsen PA, Nielsen ML, Figeys D, Mann M, et al. HysTag—a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol Cell Proteomics* 2004;3:82–92.
- [149] Nielsen PA, Olsen JV, Podtelejnikov AV, Andersen JR, Mann M, Wisniewski JR. Proteomic mapping of brain plasma membrane proteins. *Mol Cell Proteomics* 2005;4:402–8.
- [150] Zhou H, Ranish JA, Watts JD, Aebersold R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat Biotechnol* 2002;20:512–5.
- [151] Lu Y, Bottari P, Turecek F, Aebersold R, Gelb MH. Absolute quantification of specific proteins in complex mixtures

- using visible isotope-coded affinity tags. *Anal Chem* 2004;76:4104–11.
- [152] Qiu Y, Sousa EA, Hewick RM, Wang JH. Acid-labile isotope-coded extractants: a class of reagents for quantitative mass spectrometric analysis of complex protein mixtures. *Anal Chem* 2002;74:4969–79.
- [153] Shi Y, Xiang R, Crawford JK, Colangelo CM, Horvath C, Wilkins JA. A simple solid phase mass tagging approach for quantitative proteomics. *J Proteome Res* 2004;3:104–11.
- [154] Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 2003;75: 1895–904.
- [155] Wang S, Regnier FE. Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography. *J Chromatogr A* 2001;924:345–57.
- [156] Ji J, Chakraborty A, Geng M, Zhang X, Amini A, Bina M, et al. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr, B, Biomed Sci Appl* 2000;745:197–210.
- [157] Asara JM, Zhang X, Zheng B, Christofk HH, Wu N, Cantley LC. In-Gel Stable-Isotope Labeling (ISIL): a strategy for mass spectrometry-based relative quantification. *J Proteome Res* 2006;5:155–63.
- [158] Che FY, Fricker LD. Quantitation of neuropeptides in Cpe (fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry. *Anal Chem* 2002;74:3190–8.
- [159] Zhang X, Jin QK, Carr SA, Annan RS. N-terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. *Rapid Commun Mass Spectrom* 2002;16: 2325–32.
- [160] Munchbach M, Quadroni M, Miotto G, James P. Quantitation and facilitated *de novo* sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* 2000;72:4047–57.
- [161] Mason DE, Liebler DC. Quantitative analysis of modified proteins by LC-MS/MS of peptides labeled with phenyl isocyanate. *J Proteome Res* 2003;2:265–72.
- [162] Lee YH, Han H, Chang SB, Lee SW. Isotope-coded N-terminal sulfonation of peptides allows quantitative proteomic analysis with increased *de novo* peptide sequencing capability. *Rapid Commun Mass Spectrom* 2004;18:3019–27.
- [163] Hoang VM, Conrads TP, Veenstra TD, Blonder J, Terunuma A, Vogel JC, et al. Quantitative proteomics employing primary amine affinity tags. *J Biomol Tech* 2003;14:216–23.
- [164] Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 2003;75: 6843–52.
- [165] Cagney G, Emili A. *De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat Biotechnol* 2002;20: 163–70.
- [166] Brancia FL, Openshaw ME, Kumashiro S. Investigation of the electrospray response of lysine-, arginine-, and homoarginine-terminal peptide mixtures by liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 2002;16:2255–9.
- [167] Brancia FL, Montgomery H, Tanaka K, Kumashiro S. Guanidino labeling derivatization strategy for global characterization of peptide mixtures by liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* 2004;76: 2748–55.
- [168] Beardsley RL, Reilly JP. Quantitation using enhanced signal tags: a technique for comparative proteomics. *J Proteome Res* 2003;2:15–21.
- [169] Peters EC, Horn DM, Tully DC, Brock A. A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun Mass Spectrom* 2001;15: 2387–92.
- [170] Guillaume E, Panchaud A, Affolter M, Desvergnès V, Kussmann M. Differentially isotope-coded N-terminal protein sulphonation: combining protein identification and quantification. *Proteomics* 2006;6: 2338–49.
- [171] Fedjaev M, Trudel S, Tjon APN, Parmar A, Posner BI, Levy E, et al. Quantitative analysis of a proteome by N-terminal stable-isotope labelling of tryptic peptides. *Rapid Commun Mass Spectrom* 2007;21:2671–9.
- [172] Panchaud A, Guillaume E, Affolter M, Robert F, Moreillon P, Kussmann M. Combining protein identification and quantification: C-terminal isotope-coded tagging using sulfanilic acid. *Rapid Commun Mass Spectrom* 2006;20:1585–94.
- [173] Kuyama H, Watanabe M, Toda C, Ando E, Tanaka K, Nishimura O. An approach to quantitative proteome analysis by labeling tryptophan residues. *Rapid Commun Mass Spectrom* 2003;17:1642–50.
- [174] Rose K, Simona MG, Offord RE, Prior CP, Otto B, Thatcher DR. A new mass-spectrometric C-terminal sequencing technique finds a similarity between gamma-interferon and alpha 2-interferon and identifies a proteolytically clipped gamma-interferon that retains full antiviral activity. *Biochem J* 1983;215:273–7.
- [175] Liu T, Qian WJ, Strittmatter EF, Camp II DG, Anderson GA, Thrall BD, et al. High-throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. *Anal Chem* 2004;76:5345–53.
- [176] Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res* 2004;3:235–44.
- [177] Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 2005;4:1487–502.
- [178] Colinge J, Chiappe D, Lagache S, Moniatte M, Bougueleret L. Differential proteomics via probabilistic peptide identification scores. *Anal Chem* 2005;77:596–606.
- [179] Schulze WX, Mann M. A novel proteomic screen for peptide-protein interactions. *J Biol Chem* 2004;279:10756–64.
- [180] Kohlbacher O, Reinert K, Gropf C, Lange E, Pfeifer N, Schulz-Trieglaff O, et al. TOPP—the OpenMS proteomics pipeline. *Bioinformatics* 2007;23:e191–7.
- [181] Shadforth IP, Dunkley TP, Lilley KS, Bessant C. i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics* 2005;6: 145.
- [182] Bouyssie D, de Peredo AG, Mouton E, Albigo R, Roussel L, Ortega N, et al. Mascot File Parsing and Quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomics study of membrane proteins from primary human endothelial cells. *Mol Cell Proteomics* 2007;6: 1621–37.
- [183] MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates III JR. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem* 2003;75:6912–21.
- [184] Lin WT, Hung WN, Yian YH, Wu KP, Han CL, Chen YR, et al. Multi-Q: a fully automated tool for multiplexed protein quantitation. *J Proteome Res* 2006;5:2328–38.
- [185] Pan C, Kora G, McDonald WH, Tabb DL, VerBerkmoes NC, Hurst GB, et al. ProRata: a quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem* 2006;78:7121–31.
- [186] Pan C, Kora G, Tabb DL, Pelletier DA, McDonald WH, Hurst GB, et al. Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics. *Anal Chem* 2006;78:7110–20.