ANIBAL, Stable Isotope-based Quantitative Proteomics by Aniline and Benzoic Acid Labeling of Amino and Carboxylic Groups*

Alexandre Panchaud‡§, Jenny Hansson§, Michael Affolter§, Rachid Bel Rhlid¶, Stéphane Piu‡, Philippe Moreillon‡, and Martin Kussmann§

Identification and relative quantification of hundreds to thousands of proteins within complex biological samples have become realistic with the emergence of stable isotope labeling in combination with high throughput mass spectrometry. However, all current chemical approaches target a single amino acid functionality (most often lysine or cysteine) despite the fact that addressing two or more amino acid side chains would drastically increase quantifiable information as shown by in silico analysis in this study. Although the combination of existing approaches, e.g. ICAT with isotope-coded protein labeling, is analytically feasible, it implies high costs, and the combined application of two different chemistries (kits) may not be straightforward. Therefore, we describe here the development and validation of a new stable isotope-based quantitative proteomics approach, termed aniline benzoic acid labeling (ANIBAL), using a twin chemistry approach targeting two frequent amino acid functionalities, the carboxylic and amino groups. Two simple and inexpensive reagents, aniline and benzoic acid, in their $^{13}$C and $^{15}$N form with convenient mass peak spacing (6 Da) and without chromatographic discrimination or modification in fragmentation behavior, are used to modify carboxylic and amino groups at the protein level, resulting in an identical peptide bond-linked benzoyl modification for both reactions. The ANIBAL chemistry is simple and straightforward and is the first method that uses a $^{15}$N reagent for a general stable isotope labeling approach of carboxylic groups. In silico as well as in vitro analyses clearly revealed the increase in available quantifiable information using such a twin approach. ANIBAL was validated by means of model peptides and proteins with regard to the quality of the chemistry as well as the ionization behavior of the derivatized peptides. A milk fraction was used for dynamic range assessment of protein quantification, and a bacterial lysate was used for the evaluation of relative protein quantification in a complex sample in two different biological states. Molecular & Cellular Proteomics 7:800–812, 2008.

Quantitative proteomics aims at globally comparing protein abundance levels between two or more biological conditions, e.g. cellular states or different environmental or health versus disease conditions. The determination of changes in protein expression is a key element in proteome research and widely applied for the functional analysis of biological systems and the detection of diagnostic, prognostic, and clinical markers. In the past 30 years, most of the proteomics experiments have been conducted with two-dimensional (2D) PAGE as the main separation technique using gel image comparison to obtain relative quantitative information (1). 2D PAGE analysis is still widely used and yields excellent results. However, global proteomics by 2D gel electrophoresis/MS has always been limited by the discrimination of analytically challenging protein groups, such as membrane proteins, very large or small proteins, and very acidic or basic proteins (2–4).

In 1999, Gygi et al. (5) introduced a gel-free quantitative proteomics technique in conjunction with gel-free protein/peptide separations called ICAT. The advantage of stable isotope-based methods relies on their efficiency when coupled to MS. As a consequence, several strategies have been developed based on stable, non-radioactive isotopes like $^{13}$C, $^{15}$N, and $^{18}$O and can be classified in three stable isotope labeling experiment categories: (i) metabolic stable isotope labeling like stable isotope labeling by amino acids in cell culture (SILAC) using growth medium with stable isotope-labeled amino acids (6) or via isotope-labeled nutrients like $^{15}$N-ammonium salt (7, 8); (ii) isotopic tagging by chemical reaction such as ICAT, isotope-coded protein labeling (ICPL), or isobaric tagging for relative and absolute quantification (iTRAQ) (5, 9, 10); and (iii) enzyme-catalyzed reactions like...
16O/18O exchange in terminal carboxylic groups conferred by trypsin or other proteases (11, 12). Typically two biological conditions (control versus case) are defined and compared with regard to their protein expression profile. For this purpose, two chemically identical tags with different isotopic composition and thus different masses (light and heavy) are utilized. After labeling, the samples are combined and fractionated by liquid chromatography. Finally mass spectrometry is used to identify proteins from the derived peptides, and relative quantification is achieved by comparing ion intensities within the light and heavy labeled peptide pairs. Moreover in the case of iTRAQ, multiplexing is also feasible with simultaneous comparison of up to four different conditions (9).

Although all three categories show specific advantages and drawbacks (13, 14), many proteomics experiments are nowadays performed using isotope tagging by chemical derivatization. This is so certainly because of its universal applicability to any kind of biological sample (cells, tissues, or biological fluids) including \textit{ex vivo} sampling from higher organisms and because of the wide range of available quantification targets, i.e. amino acids amenable to derivatization (Cys, Lys, Met, Trp, His, Asp, Glu, or N/C termini) (15–17). Almost all approaches are based on a single labeling, but Liu and Regnier (18) described for the first time a strategy using a dual chemistry targeting amino and carboxylic groups at the peptide level.

In two previous studies, we have described the development and application of two tags, namely sulfophenyl isothiocyanate and sulfanilic acid, for quantitative proteomics and shown their performance when combined with MALDI ionization (19, 20). Both tags showed efficient ionization in MALDI and simplified fragmentation due to the sulfonate group incorporated by both the amino-directed sulfophenyl isothiocyanate and the carboxyl-directed sulfanilic acid tag into the peptide. However, the tagged peptides revealed decreased ionization efficiency in electrospray mode when compared with their untagged, native counterparts.

Here we describe a new isotope-coded chemical reaction optimized for electrospray ionization called \textit{aniline benzoic acid labeling} (ANIBAL) that relies on a twin/dual chemistry approach tagging carboxylic and amino groups at the protein level. Both reactions are based on carbodiimide chemistry to activate carboxylic groups for further reaction with primary amino groups. Light and heavy tags have a mass shift of 6 Da and do not show any retention time shift because of the 13C isotope substitution in their phenyl ring. The ANIBAL technique was first applied to protein standards for optimization, then to a milk fraction containing six major abundant proteins for dynamic range assessment, and finally to a bacterial whole cell lysate (in a 1:1 ratio) to confirm the applicability to a complex sample.

**EXPERIMENTAL PROCEDURES**

\textbf{Chemicals}—Pyridine, HEPES, ammonium bicarbonate, DTT, iodoacetamide, [13C6]aniline, [13C3]benzoic acid, [13C3]-benzoic acid, N-hydroxysulfosuccinimide sodium salt (NHSS), ethyl acetate, dimethylformamide (DMF), and \textit{N,N’}-dicyclohexylcarbodiimide were purchased from Sigma-Aldrich. 1-Ethyl-3-(3-dimethylaminopropyl) (EDC) and the BCA protein assay kit were purchased from Pierce. Sequencing grade modified trypsin was obtained from Promega (Madison, WI). The ProteoExtract precipitation kit was obtained from Calbiochem.

**In Silico Prediction**—The human International Protein Index (IPI) database version 3.26 was uploaded from the European Bioinformatics Institute. In a first analysis, the total number of each amino acid was determined, and the mean and median values per protein were calculated for all 67,665 sequences available. Secondly the number of peptides containing quantifiable information and the number of proteins without quantifiable information were determined for each amino acid by sequentially considering each of them as a possible chemical target for a stable isotope chemical labeling strategy. More precisely, all proteins were in silico digested using trypsin in two ways: R/P for labeled lysine and K/R/P for all other labeled amino acids. For each generated peptide, the mass and the amino acid composition were determined and sequentially submitted to a quantification test by considering each amino acid at a time as chemically labeled. The following criteria were necessary to pass the test: (i) containing at least one of the selected targets and (ii) having a mass between 400 and 4000 Da. Finally the same analysis was performed using chemical functional groups as targets. Carboxylic, amino, thiol, and indole groups were considered based on their known reactivity in biochemistry. All combinations of groups were also tested.

**NHSS [13C6]Benzoate and NHSS [13C3]Benzoic Acid Synthesis**—Light and heavy NHSS benzoate was synthesized according to the protocol of Staros (21) with slight modifications. 2.0 mmol of NHSS, 2.2 mmol of \textit{N,N’}-dicyclohexylcarbodiimide, and 2.0 mmol of [13C6]benzoic acid or [13C3]benzoic acid were dissolved in 5.0 ml of dry DMF. The reaction mixture was stirred overnight at room temperature and then cooled to 4 °C and stirred for another 4 h. The precipitated dicyclohexylurea was removed by filtration and washed with 2 ml of dry DMF. The product was precipitated from solution by adding a 20× excess of ethyl acetate and incubation at 4 °C for 2 h. The precipitate was filtered and stored in a desiccator.

**Ionization Efficiency**—500 pmol of BSA (corresponding to 30 nmol of carboxylic groups) was dissolved once in 20 μl of 1 m pyridine, pH 5.0, containing 375 nmol/μl [14C]aniline and once in 20 μl of 1 m pyridine without [14C]aniline. 5 μl of a 3.5 μmol/μl EDC solution was added to both samples to start the reaction. The final COOH:aniline:EDC ratio was approximately 1:250:500. Both samples were incubated for 2 h at room temperature, and the reaction was stopped by the addition of 5 μl of concentrated acetic acid. Both reactions were pooled, and BSA was precipitated with the ProteoExtract kit as described in the manufacturer’s protocol. After precipitation, the pellet was resuspended in 80 μl of 100 mM ammonium bicarbonate, reduced for 30 min at 60 °C by adding 10 μl of 45 mM DTT, alkylated for 30 min at room temperature in the dark by adding 10 μl of 100 mM iodoacetamide, and then digested overnight with trypsin (1:50 (w/w) trypsin: protein ratio) at 37 °C. Finally five-replicate MS analysis was conducted.

**Aniline Labeling of Human Milk**—The human milk fraction was prepared as described previously (22), and the protein concentration was determined by Bradford test. 50 μg of fractionated human milk aliquots (corresponding to approximately 50 nmol of carboxylic groups) were derivatized in 20 μl of 1 m pyridine, pH 5.0, containing a 625 nmol/μl concentration of either light or heavy aniline and sonicated for 10 min. 10 μl of a 2.5 μmol/μl EDC solution were added to start the reaction. The final COOH:aniline:EDC ratio was approximately 1:250:500. After 2 h of derivatization at room temperature, both aliquots (light and heavy) were precipitated separately (ProteoExtract kit). The precipitated pellets were resuspended in 80 μl of 100 mM ammonium bicarbonate, reduced, alkylated, and digested over-
night as described above. Light and heavy derivatized milk fractions were then mixed to produce the following light to heavy ratios: 0.05, 0.10, 0.20, 0.33, 0.50, 1, 2, 3, 5, 10, and 20.

**Benzoic Acid Labeling of Human Milk—**

50 μg of fractionated human milk (corresponding to approximately 25 nmol of amino groups) were derivatized in 40 μl of 200 mM HEPES, pH 8.0, with light and heavy NHSS benzoate (in two different tubes) at an NH₂:NHSS benzoate ratio of 1:500. After 4 h of derivatization at room temperature, both aliquots (light and heavy) were precipitated separately and processed (reduction, alkylation, and digestion with trypsin) as described above. Light and heavy derivatized milk fractions were then mixed to produce the following light to heavy ratios: 0.05, 0.10, 0.20, 0.33, 0.50, 1, 2, 3, 5, 10, and 20.

**Lactococcus lactis Labeling with ANIBAL—**

Four 50-μg aliquots of the soluble proteins of a total cell lysate of *L. lactis* were used (50 μg corresponds to approximately 50 nmol of carboxylic groups and 25 nmol of amino groups). Two aliquots were derivatized in 60 μl of 1 M pyridine, pH 5.0, containing a 200 nmol/μl concentration of either light or heavy aniline and sonicated for 10 min. 10 μl of a 2.5 μmol/μl EDC solution were added to initiate the reaction. The final COOH:aniline:EDC ratio was 1:250:500. Both samples were incubated for 2 h at room temperature, and the reaction was stopped by addition of 5 μl of acetic acid. Samples were then mixed, and proteins were precipitated using the ProteoExtract kit.

In parallel, the remaining two aliquots were derivatized in 40 μl of 200 mM HEPES, pH 8.0, with light and heavy NHSS benzoate at an NH₂:NHSS benzoate ratio of 1:500 (approximately 4 mg of NHSS benzoate (molecular weight (light) = 321; molecular weight (heavy) = 327). After 4 h of derivatization at room temperature the two samples (light and heavy) were mixed (1:1) and precipitated. For both derivatizations the pellets were resuspended and processed (reduction, alkylation, and digestion with trypsin) as described above.

**LC-MS Analysis—**

The LC-MS/MS data were acquired using an HCTultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) coupled on line to an Ultimate 3000 HPLC system (Dionex, Sunnyvale, CA) equipped with an analytical Magic C₁₈ reversed-
phase column (100 × 0.075 mm, 5 μm) (Spectronex, Basel, Switzerland). The scan range was set at m/z 400–1600. For each injection, the peptide mixture was loaded and washed for 10 min with 2% (v/v) acetonitrile, 0.1% (v/v) formic acid on a C18 PepMap100 trapping column (5 × 0.3 mm, 5 μm) (LC Packings, Amsterdam, Netherlands) at a flow rate of 20 μl/min prior to elution with a linear gradient of 5–50% (v/v) acetonitrile, 0.1% (v/v) formic acid at a flow rate of 0.3 μl/min over 30 min (BSA), 60 min (milk fraction), and 90 min (L. lactis). Peptides were analyzed using the "peptide scan" option of the HCTultra system consisting of a full-scan MS spectrum acquisition in "standard-enhanced" mode (8,100 m/z/s) for charge state assignment based on the 13C isotope envelope followed by three MS/MS scans in "ultra scan" mode (26,000 m/z/s) on the three most abundant ions as well as exclusion of singly charged ions with preferred charge state set to double for MS/MS selection.

Data Extraction—Data were extracted using DataAnalysis 3.4 (Bruker Daltonics) generating peak lists in mgf format. Bruker Daltonics yep files were transformed to mzXML using CompassXport 1.3.2. An in-house script called da2tp was used to transform mgf files to new mgf files or dta files used for Mascot (Matrix Science, London, UK) and Sequest (ThermoFisher, San Jose, CA) search and to be compatible for further validation using the Trans-Proteomic Pipeline version 2.9.8 (Institute for Systems Biology, Seattle, WA).

Database Search—Data were searched using Mascot or Sequest against the Swiss-Prot database using the taxonomy "other Mammalia" for BSA and "Homo sapiens" for the milk fraction. The L. lactis subsp. cremoris SK11 (GenBank accession number NC_009004) database was used for the bacterial samples. New modifications for aniline and benzoic acid were created in the Mascot modification file as described in supplemental Data 1.

Mascot search parameters were set as follows:
- Trypsin (KR/P) was used for aniline labeling and Arg-C (R/P) was used for benzoic acid labeling with in both cases two missed cleavages allowed.
- Peptide tolerance was 100 ppm, and MS/MS tolerance was 0.6 Da.
- Fixed modifications used were carbamidomethyl (Cys), aniline_light (Asp), and aniline_light (Glu) for aniline labeling or carbamidomethyl (Cys) and BA_light (Lys) for benzoic acid labeling.
- Variable modifications used were oxidation (Met), deamidation (NQ), aniline_light ("pC-term"), aniline_heavy (Asp), aniline_heavy (Glu), and aniline_heavy (pC-term) for aniline labeling or oxidation (Met), deamidation (NQ), BA_light ("pN-term"), BA_heavy (Lys), and BA_heavy (pN-term) for benzoic acid labeling.

Sequest search parameters were set as follows.
- Trypsin (KR/P) was used for aniline labeling or "trypsin_R" (R/P) was used for benzoic acid labeling with in both cases two missed cleavages allowed.
- Peptide tolerance was 100 ppm, and fragment ion tolerance was 0.8 Da.
- Static modifications used were C+57.0513 and DE+75.1130 for aniline labeling or C+57.0513, K+104.1061, and "Ctermprot"+104.1061 for benzoic acid labeling.
- Differential modifications used were M+15.9994, NQ+0.9840, and DE+6.0000 for aniline labeling or M+15.9994, NQ+0.9840, K+6.0000, and "Cterm"+6.0000 for benzoic acid labeling.

Data Validation and Quantification—The mzXML, Mascot dat files, and Sequest dta/out directory were transferred on a Linux server with the Trans-Proteomic Pipeline version 2.9.8 installed on it. Search results were transformed to pepXML using either Mascot2XML or Out2XML. Data were then analyzed using ProteinProphetTM (23) and filtered using a 0.05 error rate. Quantification was performed using both Xpress (24) and ASAPRatio (25) for the milk sample, whereas only ASAPRatio was used for the bacterial sample. Finally proteins were validated using ProteinProphetTM (26) with again a 0.05 error rate, and protein quantification was checked manually.

RESULTS

Principle of the ANIBAL Approach—The complete work flow is shown in Fig. 1. The ANIBAL strategy relies on a
combined “symmetric” chemistry approach targeting two frequent amino acid side chains resulting in the incorporation of the same peptide-like benzoyl group into the protein as illustrated in Fig. 2. Targeting both amino and carboxylic groups enables a broad proteome coverage for relative quantification experiments. For each pair of biological conditions to be compared, two pairs of samples with equal protein amounts are prepared and immediately derivatized using the two different labels either in their light ($^{13}$C) or heavy ($^{15}$C) forms. Light or heavy aniline is used to derivatize all carboxylic groups of the proteins, namely Asp or Glu residues as well as the C terminus. Similarly, NHSS benzoate, a carbodiimide-activated form of benzoic acid, is used to derivatize all amino groups (Lys and N terminus). After combining both mixtures for each labeling approach, proteins are digested, and peptides are analyzed by high throughput LC-MS/MS. As peptides with identical sequence derived from the two different states only differ in mass, they appear as doublets in the acquired MS spectra. Relative quantification and peptide ratios are determined by performing an extracted ion chromatogram for each mass of the doublets. Proteins are identified by CID MS/MS and searching protein sequence databases with different search algorithms. Finally, results are integrated into one single result file containing identification as well as quantification information of peptides derived from the two different labeling strategies.

**In Silico Survey of Amino Acid Distribution among Proteins**—To assess the suitability of multiple labeling approaches in terms of proteome coverage, a survey of amino acid distribution as well as protein functional groups was performed using the human IPI database version 3.26 as a model. Table I shows the results obtained for each amino acid plus both termini. The first two columns display the total number and the mean and median number of amino acids per protein. The last two columns show the mean and median number of quantifiable peptides per protein as well as the percentage of proteins without any quantification information available again for each chemically targeted amino acid. Peptides were generated *in silico* with trypsin (KR/P), and a peptide was considered quantifiable if it (i) bears at least one of the desired amino acids and (ii) has a mass between 400 and 4000 Da corresponding to the analysis range of doubly/triply charged tryptic peptides accessible with modern MS instruments. In the case of labeling lysine residues, a trypsin_R (R/P) cleavage was defined instead of standard tryptic proteolysis. Table II was generated with the same rules applied as for Table I but with a focus on chemical functional groups for Table I but with a focus on chemical functional groups. Table II was generated with the same rules applied as for Table I but with a focus on chemical functional groups.

The total frequency of each amino acid is displayed for all 67,687 entries of the IPI human database version 3.26 as well as the mean (median) number of amino acids per protein in the first two columns. The mean (median) quantifiable peptides per protein as well as the percentage of proteins without quantifiable information are shown in the last two columns. Nter, N terminus; Cter, C terminus.

| Amino acid | Total | Mean (median) amino acid per protein | Mean (median) quantifiable peptides per protein | Protein without quantifiable information % |
|------------|-------|------------------------------------|----------------------------------------------|------------------------------------------|
| Ala        | 1,990,878 | 29.41 (19)                       | 12.43 (8)                                        | 2.54                                       |
| Cys        | 649,682   | 9.69 (6)                          | 5.55 (4)                                        | 9.21                                       |
| Asp        | 1,330,324 | 19.65 (12)                        | 10.43 (6)                                        | 5.52                                       |
| Glu        | 2,003,916 | 29.61 (17)                        | 13.36 (8)                                        | 3.64                                       |
| Phe        | 1,013,942 | 14.98 (10)                        | 8.38 (5)                                         | 5.69                                       |
| Gly        | 1,894,148 | 27.98 (18)                        | 11.87 (8)                                        | 2.66                                       |
| His        | 747,969   | 11.05 (7)                         | 6.74 (4)                                         | 7.76                                       |
| Ile        | 1,217,968 | 17.99 (11)                        | 9.57 (6)                                         | 5.69                                       |
| Lys        | 1,608,200 | 23.76 (14)                        | 6.99 (4)                                         | 7.96                                       |
| Leu        | 2,802,298 | 41.40 (27)                        | 15.92 (10)                                       | 1.35                                       |
| Met        | 614,394   | 9.08 (6)                          | 5.66 (4)                                         | 6.53                                       |
| Asn        | 1,001,225 | 14.79 (8)                         | 8.38 (5)                                         | 7.46                                       |
| Pro        | 1,840,691 | 27.19 (16)                        | 11.25 (7)                                        | 2.70                                       |
| Gln        | 1,364,108 | 20.15 (12)                        | 10.43 (6)                                        | 3.91                                       |
| Arg        | 1,641,606 | 24.25 (16)                        | 12.34 (8)                                        | 3.17                                       |
| Ser        | 2,395,221 | 35.39 (21)                        | 13.75 (9)                                        | 1.48                                       |
| Thr        | 1,542,480 | 22.79 (14)                        | 11.17 (7)                                        | 2.76                                       |
| Val        | 1,701,214 | 25.13 (16)                        | 11.88 (8)                                        | 2.90                                       |
| Trp        | 362,501   | 5.36 (4)                          | 3.66 (2)                                         | 16.09                                      |
| Tyr        | 733,834   | 10.84 (7)                         | 6.65 (4)                                         | 10.92                                      |
| Nter       | 67,665    | 1.00 (1)                          | 0.62 (1)                                         | 61.72                                      |
| Cter       | 67,665    | 1.00 (1)                          | 0.52 (1)                                         | 52.14                                      |

* A peptide is considered quantifiable if it has at least one amino acid and a mass between 400 and 4000 Da.
Four chemically amenable amino acid side chain groups (carboxylic (COOH), amino (NH$_2$), thiol, and indole groups) were considered as chemical targets for stable isotope labeling either alone or in combination of two, three, or four. The total frequency of each functional group is displayed for all 67,687 entries of the IPI human database version 3.26 as well as the mean (median) number of groups per protein in the first two columns. The mean (median) quantifiable peptides per protein as well as percentage of proteins without quantifiable information are shown in the last two columns.

| Functional group combination | Total       | Mean (median) functional groups per protein | Mean (median) quantifiable peptides per protein | Protein without quantifiable information | %  |
|-----------------------------|-------------|--------------------------------------------|-----------------------------------------------|------------------------------------------|-----|
| COOH                        | 3,401,927   | 50.26 (30)                                 | 17.62 (11)                                    |                                           | 1.26 |
| NH$_2$                      | 1,675,887   | 24.76 (15)                                 | 10.71 (7)                                     |                                           | 4.24 |
| Thiol                       | 649,682     | 8.6 (6)                                    | 5.55 (4)                                      |                                           | 9.21 |
| Indole                      | 362,501     | 5.36 (4)                                   | 3.66 (2)                                      |                                           | 16.09 |
| COOH:NH$_2$                 | 5,077,814   | 75.92 (45)                                 | 28.32 (18)                                    |                                           | 0.77 |
| COOH:thiol                  | 4,051,609   | 59.86 (36)                                 | 23.16 (15)                                    |                                           | 0.86 |
| COOH:indole                 | 3,764,428   | 55.62 (34)                                 | 21.28 (13)                                    |                                           | 0.99 |
| NH$_2$:thiol                | 2,225,569   | 34.36 (21)                                 | 16.25 (11)                                    |                                           | 1.51 |
| NH$_2$:indole               | 2,038,388   | 30.11 (19)                                 | 14.37 (10)                                    |                                           | 2.21 |
| Thiol:indole                | 1,012,183   | 14.95 (10)                                 | 9.21 (7)                                      |                                           | 4.61 |
| COOH:NH$_2$:thiol           | 5,727,496   | 84.62 (61)                                 | 33.87 (21)                                    |                                           | 0.59 |
| COOH:NH$_2$:indole          | 5,440,315   | 80.37 (49)                                 | 31.99 (20)                                    |                                           | 0.66 |
| COOH:thiol:indole           | 4,414,110   | 65.21 (40)                                 | 26.82 (17)                                    |                                           | 0.76 |
| NH$_2$:thiol:indole         | 2,688,070   | 39.71 (25)                                 | 19.92 (13)                                    |                                           | 1.11 |
| COOH:NH$_2$:thiol:indole    | 6,089,997   | 89.97 (55)                                 | 37.53 (24)                                    |                                           | 0.56 |

* A peptide is considered quantifiable if it has at least one functional group and a mass between 400 and 4000 Da.

In silico assessment of four protein functional groups as targets for stable isotope labeling

In silico analysis clearly reveals that although each single chemical approach suffers from certain limitations due to amino acid distribution among proteins combinations thereof show interesting complementarities and emerge as a very powerful alternative.

Labeling of Standard Peptides and Proteins—Any protein labeling strategy has to address three critical issues: first, the completeness of the modification reaction; second, the influence of the labeling on the protein solubility; and third, the influence on the LC-MS/MS behavior (retention, ionization, and fragmentation).

To assess the completeness of the reaction for both aniline and benzoic acid, model peptides were first used for validation as already described (20). Briefly model peptides were derivatized and analyzed in short LC-MS runs to identify the different peptide derivatives after the reaction. Conditions were optimized for both labeling strategies until only a single peak corresponding to the fully derivatized peptide was obtained, indicating completeness of the reaction.

Once a quantitative reaction was achieved at the peptide level, BSA was chosen as the model protein to adapt the reaction to the protein level. To keep similar chemical ratios, the amount of reagent was adapted to the number of carboxylic or amino groups in BSA (100 × COOH and 61 × NH$_2$). BSA was derivatized using light reagents only, digested, and analyzed by means of LC-MS. Completeness of the reaction was first assessed by performing Mascot-driven database searches in two ways: (i) aniline or benzoic acid set as fixed modifications to identify only fully derivatized peptides and (ii) aniline or benzoic acid set as variable modifications to reveal incomplete reaction through the presence of native or partially derivatized peptides. In addition to the directed Mascot searches for identified labeled peptides, corresponding native or partially derivatized peptide masses were extracted from LC-MS/MS chromatograms to check whether low amounts of native or partially labeled peptides were remaining for which no MS/MS had been acquired. For both aniline (ANI) and
ANIBAL: Aniline Benzoic Acid Labeling

Fig. 3. Statistics of stable isotope labeling targets by in silico analysis using the IPI human database. A, each amino acid was evaluated as a potential target for stable isotope labeling, and the mean amino acid content per protein was calculated. Proteins were digested in silico with trypsin, and each peptide was submitted to a quantification test applying the following rules. A peptide was considered as quantifiable if it (i) contained at least one targeted amino acid and (ii) exhibited a mass between 400 and 4000 Da. Each of the 20 amino acids was considered as a potential target, and thereby each peptide was submitted to the test for all 20 cases. For each protein, the number of amino acids and the number of quantifiable peptides were stored in an array. Finally as shown graphically, the mean amino acid content and the mean number of quantifiable peptides per protein as well as the number of proteins without quantification were calculated. B, functional groups (carboxylic, amino, thiol, and indole groups) instead of amino acids were assessed in this case either alone or in combination of two, three, or even four. Processing was identical to that in A. Nter, N terminus; Cter, C terminus.

benzoic acid (BA) labeling approaches, exclusively fully labeled peptides and no native or partially labeled peptides were observed in the extracted ion chromatograms (data not shown).

It is important to note that any solvents, reagents, or buffers that contain free amino groups (e.g. Tris and urea) would quench the reaction for both labeling approaches and should therefore be strictly avoided. Similarly solvents, reagents, or buffers containing carboxylic groups would compete for the aniline reaction and should be avoided too. Concerning the solubility of the proteins after labeling, no precipitation of the modified protein was observed in aqueous solution.

Finally the influence of ANIBAL on LC-MS behavior was evaluated by comparing native versus labeled BSA peptides. Identical amounts of BSA were treated in parallel under identical conditions in the presence or absence of the tagging reagent (aniline or benzoic acid). After derivatization, the reaction was stopped, and samples were mixed before protein digestion. Trypsin was used in the case of aniline labeling as only carboxylic groups had been derivatized. On the contrary, Arg-C was used for benzoic acid labeling as all lysines had been derivatized. If trypsin had been used in the case of benzoic acid labeling, labeled versus unlabeled BSA would have generated different and incomparable peptides. The digest was then analyzed by LC-MS (five-replicate analysis), and peptides were identified by database search with the tags defined as variable modifications (meaning that labeled and unlabeled peptides could be identified). At last, selected analytical parameters (Mascot score, intensity, scan number, and retention time) for 11 BSA peptides in their unlabeled and labeled forms were extracted as shown in Table III for aniline derivatization. Fig. 4B shows that similar Mascot scores were observed for both unlabeled and labeled peptides indicating that the tagging as such does not deteriorate the fragmentation of the peptides and therefore had no detrimental effect on the database searches. Fig. 4A illustrates the ionization efficiency of the 11 peptides in both unlabeled and labeled forms and indicates that the ionization efficiency of labeled peptides is in general unaltered, and thereby the tags do not compromise peptide/protein identification. Finally the peptide hydrophobicity was increased by the benzoyl group incorporation as observed by the prolonged retention times for ANI- and BA-derivatized peptides compared with their native counterparts (approximately 3.5 min per modified amino acid; Table III). All peptides, derivatized or not, eluted before the gradient had reached 50% acetonitrile. The increase in hydrophobicity conferred by ANI and BA tagging reveals an advantageous property for both tags, namely to retain very short or hydrophilic peptides that normally do not bind to the column. This latter effect was also demonstrated by Julka and Regnier (27) by means of benzoyl derivatization to improve the retention of hydrophilic peptides in tryptic peptide mapping.

Dynamic Range Assessment of ANIBAL Technique on Abundant Milk Proteins—Besides the three already considered critical issues in protein labeling (reaction completeness, derivative solubility, and LC-MS/MS behavior of derivatives), a fourth quality criterion needs to be addressed for any labeling approach, namely the achievable dynamic range. For this purpose, a human milk fraction generated and analyzed in our laboratory in a previous study (22) was used. This fraction comprises a few very abundant proteins (albumin, lactotransferrin, immunoglobulins, etc.) besides many low abundant
species and is thus in this regard similar to blood-derived samples such as fractions of serum or plasma. We took advantage of these sample properties and the available information on its composition to select it as a “complex protein mixture of known proteins” and, at the same time, as a real life biological sample.

For both labeling approaches, the same protein amount was derivatized separately with light and heavy ANI and BA tags, respectively. The reaction conditions were adapted according to an in silico prediction of the human IPI protein database for which the mean molecular mass (approximately 48 kDa) and the number of amino (approximately 25 per protein) and carboxylic (approximately 50 per protein) groups were determined. After labeling, light and heavy labeled fractions were mixed, and proteins were precipitated and digested with trypsin. Finally different dilutions were performed to obtain ratios ranging from 0.05 to 20 (light versus heavy). Dilutions were performed after digestion to avoid precipitation problems due to different protein amounts that would arise if dilution was performed directly after labeling. Peptide mixtures were analyzed by LC-MS, and proteins were quantified using both the Xpress (24) and the ASAPRatio (25) tool from the Trans-Proteomic Pipeline (Institute for Systems Biology). Six of the most abundant proteins were assessed for the quantification, and each peptide spectrum was checked manually to ensure the reliability of both methods regarding dynamic range. Fig. 5 illustrates the results obtained for these six proteins over the assessed dynamic range (1:20 to 20:1 light versus heavy). Fig. 5, A and B, represent results for aniline labeling obtained with Xpress and ASAPRatio, respectively; Fig. 5, C and D, report on the benzoic acid labeling approach. Logarithmic curves clearly show that both the ANI and BA method are linear over 2 orders of magnitude (error <15% between relative ratios of 0.1 and 10). For ratios below 0.05 or above 20, it was no longer possible for Xpress and ASAPRatio to precisely calculate the ratios of the proteins because it was critical to accurately acquire the mass signals of the smaller peaks in the peptide pairs. Moreover and log-

| Sequence | Mean Mascot score | Mean Intensity | Mean Retention Time (RT) | RT Difference | Number of carboxylic group(s) |
|----------|-------------------|----------------|--------------------------|---------------|------------------------------|
| K.LVNELTEFAK.T | 45.20 ± 5.26 | 30754344 ± 8 % | 41.58 ± 0.19 | 3.88 | 2 |
| K.LVNELTEFAK.T | 45.40 ± 8.99 | 64609605 ± 20 % | 37.70 ± 0.16 | | |
| K.SLHTLFGLDELCK.V | 42.36 ± 5.33 | 9054000 ± 13 | 42.68 ± 0.13 | 7.53 | 2 |
| K.SLHTLFGLDELCK.V | 39.65 ± 4.52 | 7535654 ± 9 | 35.15 ± 0.15 | | |
| K.KLYEIRAR.R | 37.20 ± 4.38 | 69531038 ± 19 % | 39.49 ± 0.18 | 4.90 | 1 |
| K.KLYEIRAR.R | 37.20 ± 2.68 | 45490174 ± 4 % | 34.59 ± 0.14 | | |
| K.YCDNGQDSSKL.K | 48.80 ± 4.32 | 11892965 ± 22 | 44.83 ± 0.41 | 11.54 | 3 |
| K.YCDNGQDSSKL.K | 53.00 ± 10.22 | 6106327 ± 46 % | 33.29 ± 0.56 | | |
| K.KLYEIRAR.R | 52.20 ± 7.26 | 46360242 ± 22 | 39.56 ± 0.20 | 9.70 | 2 |
| K.KLYEIRAR.R | 70.60 ± 6.88 | 12370009 ± 2 % | 29.86 ± 0.14 | | |
| K.KLYEIRAR.R | 42.32 ± 3.65 | 42656322 ± 13 % | 39.01 ± 0.09 | 6.96 | 2 |
| K.KLYEIRAR.R | 59.54 ± 4.56 | 29653244 ± 16 % | 32.05 ± 0.10 | | |
| K.KLYEIRAR.R | 67.20 ± 6.42 | 11503225 ± 45 % | 43.97 ± 0.30 | 3.60 | 1 |
| K.KLYEIRAR.R | 81.00 ± 6.28 | 77716295 ± 21 % | 40.37 ± 0.15 | | |
| K.KLYEIRAR.R | 57.60 ± 6.88 | 87566893 ± 12 % | 40.52 ± 0.37 | 1.34 | 1 |
| K.KLYEIRAR.R | 63.80 ± 1.64 | 97435021 ± 23 % | 36.78 ± 0.14 | | |
| K.KLYEIRAR.R | 73.00 ± 5.10 | 208639933 ± 13 % | 39.11 ± 0.13 | 3.42 | 1 |
| K.KLYEIRAR.R | 73.60 ± 11.74 | 206834693 ± 5 % | 35.69 ± 0.14 | | |
| K.KLYEIRAR.R | 43.40 ± 6.54 | 125324101 ± 21 % | 42.70 ± 0.14 | 3.30 | 1 |
| K.KLYEIRAR.R | 59.00 ± 6.93 | 40844562 ± 17 % | 39.41 ± 0.13 | | |
| K.KLYEIRAR.R | 44.60 ± 3.36 | 91961844 ± 50 % | 41.64 ± 0.16 | 4.09 | 2 |
| K.KLYEIRAR.R | 42.40 ± 5.94 | 23102607 ± 11 % | 37.55 ± 0.15 | | |

TABLE III
Assessment of native versus aniline-derivatized BSA peptides

Eleven peptides of BSA are displayed with their respective Mascot score, intensity, and retention time (RT) in their native (white) or aniline-labeled (gray) form. Differences in retention time between unlabeled and labeled peptides as well as number of carboxylic groups (equals number of modifications) are shown. Periods in sequences represent cleavage sites.

ANIBAL: Aniline Benzoic Acid Labeling
The quantification accuracy depended on the ionization efficiency. Peptides with high ion intensities are easier to quantify over a broad dynamic range. However, the quantitative information for peptide pairs showing a 20-fold up- or down-regulation in either one of the conditions is not lost because the more abundant labeled peptide is nevertheless detected as a singlet. The difficult task in this latter case, however, is to determine whether the identification is a true positive with a 20-fold up- or down-regulation explaining the singlet or whether it is a false-positive peptide with a sequence containing a labeled amino acid and for which no labeled peptide partner is detectable in this case.

Table IV summarizes all peptides considered for identification and quantification of lactotransferrin at a light-to-heavy ratio of 1 using ANIBAL. It shows individual ratios obtained for both tagging reagents as well as the mean ratio with the S.D. for the protein. Two observations can be made. First, the ratios of the 19 identified isotopically tagged peptides (nine for aniline and 10 for benzoic acid) (indicated with an asterisk (*)) are all close to the expected ratio of 1 with a mean Xpress and ASAPRatio of 1.01 ± 0.12 and 1.03 ± 0.09, respectively. Second, the sets of identified peptides obtained by either ANI or BA chemistry undoubtedly reveal the complementarity of both methods because ANI and BA applied in parallel clearly increase the overall protein coverage (>56% (401 amino acids) for a 78-kDa protein) as well as improve the quantification statistics. The advantages are mainly due to the different labeled amino acids and to the different cleavage sites obtained with trypsin, e.g. free lysines with ANI labeling and blocked lysines with BA labeling.

**Analysis of the L. lactis Proteome in a 1:1 Ratio**—Finally as the main proof-of-principle experiment, the ANIBAL method was applied to the soluble fraction of a total cell lysate of *L. lactis*. Two aliquots of 100 μg were derivatized with ANIBAL in a 1:1 ratio, that is 50 μg versus 50 μg for light and heavy aniline and 50 μg versus 50 μg for light and heavy benzoic acid. The light and heavy labeled reaction mixtures of each derivatization were subsequently combined, digested, and analyzed by LC-MS/MS running a 90-min gradient. For each branch of the ANIBAL work flow (see Fig. 1), data were extracted, database searches were performed with Mascot and Sequest with their corresponding isotopic tags, and peptides were quantified and validated using Xpress, ASAPRatio, and PeptideProphet (23). Peptide results from both labeling and search procedures were combined, protein ratios were calculated, and protein probabilities were computed with ProteinProphet (26) to yield the final ANIBAL result. A 5% false-positive rate was accepted for protein identification, and quantifiable proteins with this 95% identification confidence were further assessed with regard to the calculated ratio. Applying a single 90-min LC gradient (5–50% acetonitrile, 0.1% formic acid), a total number of 188 proteins (43 with a single peptide hit) were identified of which 101 revealed quantitative information with an average of 3.3 quantifiable peptides per protein.

Fig. 6 displays the presence or absence of the 188 identified proteins with regard to identification alone or identification plus quantification achieved by the two labeling strategies of the ANIBAL approach. Fig. 6, A, B, and C, highlight the presence/absence of proteins identified with the combined ANIBAL approach (100% of the identifications corresponding to 188 proteins; Fig. 6A), with aniline labeling alone (Fig. 6B), and with benzoic acid labeling alone (Fig. 6C). Fig. 6, D, E, and F, show the presence/absence of quantifiable proteins with the combined ANIBAL approach (Fig. 6D), with the aniline labeling alone (Fig. 6E), and with the benzoic acid labeling alone (Fig. 6F). The complementarity of both tagging methods is evident with the aniline approach resulting in overall more protein identifications and quantifications. The bias of the latter approach can be partly explained by the fact that tryptic cleavage generates more peptides in the case of aniline la-
beling (cleaves at KRE) than in the case of benzoic acid derivatization (cleaves after R£), thus yielding more peptides amenable to analysis. Regarding quantification, the better yield for ANI compared with BA may be explained by the more frequent carboxylic groups compared with amino groups (on average 2-fold more COOH than NH$_2$ groups per protein). This experimental finding confirms the in silico prediction of achievable quantification coverage as discussed previously (Table II).

As equal amounts of *L. lactis* protein were derivatized with the light and heavy tag, all labeled peptide pairs, and consequently all protein ratios, should yield a light-to-heavy ratio of 1:1. Fig. 7 plots all quantified proteins with their respective ratios as calculated by ASAPRatio. All determined ratios are close to the expected 1:1 ratio with an average ratio for all proteins of 1.03 ± 0.16 representing a mean difference between expected and observed ratios of less than 17% using automatic data processing tools.

**DISCUSSION**

We have developed and validated a novel approach based on twin labeling for global identification and quantification of proteins in complex mixtures. The ANIBAL concept is based on separate and complementary labeling of proteins at their carboxylic and amino groups using two very similar tags in a symmetrical chemical reaction. Both reactions are based on carbodiimide chemistry to activate carboxylic groups and result in an amide bond formation with primary amino groups. For the aniline reaction, protein carboxylic groups are activated and react with the amino group of aniline present in excess to avoid any competitive reaction due to protein amino groups. In the benzoic acid method, the tag itself contains the carboxylic group, is activated, and reacts with primary amino groups of the proteins. Both reactions result in peptide bond formation between the benzoyl reagents and the protein carboxylic or amino group, thus avoiding the introduction of any new chemical bond in the protein.

The peptide ionization efficiency is preserved if not increased. A retardation of the peptide elution time is observed and may help to retain very hydrophilic peptides on reversed-phase material as already shown by Julka and Regnier (27). Performance assessment with peptide and protein standards has demonstrated a quantitative, robust, and reproducible chemistry. The successful relative quantification experiment performed with a bacterial sample has confirmed the applicability of the ANIBAL technique to a complex sample.

Differential isotope incorporation by chemical labeling was first introduced in 1999 by Gygi et al. (5) with ICAT. Meanwhile other approaches have been developed like ICPL or iTRAQ (9,
These approaches rely on a chemical reaction with often harsh conditions (compared with “soft” metabolic labeling like SILAC (6)) for the protein sample. Nevertheless chemical labeling is nowadays the most frequently used technique in quantitative proteomics mainly because of its applicability to any biological sample (cells, tissues, body fluids, etc.), for its selectivity toward certain amino acids, and for the possible peptide enrichment thanks to affinity tags. However, as demonstrated by our in silico analysis, each of these approaches suffers from limited proteome coverage in terms of availability of accessible groups, resulting in a significant number of proteins with a limited amount of quantifiable peptides or even without quantifiable information. This constraint becomes even more important at the peptide level when LC and MS amenability differ a lot. Therefore, a labeling strategy that would increase the chances of labeling a “proteotypic” pep-

| Peptide position^a | Peptide mass (Da) | Peptide sequence identified^b | Strategy used^c | PeptideProphet probability | Xpress ratio (L/H) | ASAP ratio (L/H) |
|--------------------|------------------|------------------------------|-----------------|----------------------------|------------------|-----------------|
| 1 23 - 37          | 1689.91          | RSQWCASVQPE@ATK             | ani             | 0.9807                     | 1.01             | 1.03            |
| 2 24 - 37          | 1533.72          | SVQWCAVSQPE@ATK             | ani             | 0.9910                     | 1.10             | 1.11            |
| 3 24 - 43          | 2382.7           | SVQWCAVSQPEATK*CFOQWR       | ba              | 0.9996                     | 0.89             | 0.93            |
| 4 50 - 58          | 1956.17          | GPPVSCISK*R                 | ba              | 0.9970                     | 0.92             | 0.98            |
| 5 59 - 72          | 1557.74          | DSIPOCQCAIAENR              | ba              | 0.9999                     | not labeled      | not labeled     |
| 6 93 - 105         | 1460.65          | LRPVAAEYGGTER               | ba              | 0.9860                     | not labeled      | not labeled     |
| 7 119 - 139        | 2273.6           | K*GGSFQNLEQGLKSCHTLGK      | ba              | 0.9870                     | 0.97             | 1.01            |
| 8 120 - 132        | 1390.56          | GGSFQLNE*LQLKL             | ba              | 0.9009                     | 0.86             | 0.93            |
| 9 171 - 182        | 1228.39          | FFSASCVPGAD*K               | ani             | 0.9967                     | 1.01             | 1.01            |
| 10 171 - 190       | 2144.45          | FFSASCVPGAD*K*GQFPNLCR      | ani             | 0.9201                     | 0.92             | 0.78            |
| 11 183 - 190       | 934.08           | GQFPNLCR                    | ani             | 0.9771                     | not labeled      | not labeled     |
| 12 220 - 229       | 1020.11          | DGAQDVAIR                   | ani             | 0.9999                     | not labeled      | not labeled     |
| 13 230 - 243       | 1626.85          | ESTFEDLSDEAER               | ba              | 0.9907                     | not labeled      | not labeled     |
| 14 256 - 268       | 1556.95          | K*VPDKF*K*DCHLAR            | ba              | 0.9971                     | 1.09             | 1.07            |
| 15 278 - 291       | 1614.82          | SVNGK*EDAIWWWL             | ani             | 0.9983                     | 0.98             | 0.92            |
| 16 321 - 328       | 851.91           | D*SAIGFSR                   | ani             | 0.9867                     | 0.84             | 1.08            |
| 17 333 - 351       | 2088.35          | IDSGLYGSGYFTAIQNL          | ani             | 0.9999                     | not labeled      | not labeled     |
| 18 353 - 360       | 889.92           | SE*EE*EVAAR                 | ba              | 0.9614                     | 0.87             | 1.04            |
| 19 364 - 375       | 1388.6           | VVWCAVGEQELR               | ba              | 0.9995                     | not labeled      | not labeled     |
| 20 424 - 435       | 1305.55          | CGLVPLAE*NYK               | ani             | 0.9957                     | 1.02             | 1.10            |
| 21 462 - 484       | 2576.87          | RSDTSLTWSNK*GK*K*SCHTAVDR  | ani             | 0.8165                     | 1.29             | 1.26            |
| 22 463 - 473       | 1237.33          | SD*TLTWSNK                 | ani             | 0.9222                     | 1.30             | 1.12            |
| 23 463 - 484       | 2420.58          | SALTWSNK*GK*K*SCHTAVDR     | ani             | 0.9130                     | 1.03             | 1.08            |
| 24 544 - 552       | 1097.19          | YGYTVEHFR                  | ba              | 0.9878                     | not labeled      | not labeled     |
| 25 607 - 621       | 1592.95          | SCHLAMAPNHAVSR             | ani             | 1.0000                     | not labeled      | not labeled     |
| 26 628 - 641       | 1666             | LK*QVLLHQAK*FGR            | ba              | 0.9989                     | 0.95             | 1.04            |
| 27 650 - 658       | 1102.27          | FCLOESEQ*K                 | ani             | 0.9847                     | 0.90             | 0.96            |
| 28 681 - 694       | 1536.79          | YLGPOQYVAGITNLK           | ani             | 0.9979                     | not labeled      | not labeled     |
| 29 695 - 709       | 1697             | K*CSSPLLEEACEFLR          | ani             | 0.9997                     | 1.00             | 1.05            |

Mean ratio ± SD: 1.01 ± 0.12 1.03 ± 0.09

Sequence coverage: 56.5% (all peptides)
Sequence coverage: 36.7% (only labeled peptides)
Sequence coverage: 14.4% (only aniline labeled peptides)
Sequence coverage: 24.0% (only benzoic labeled peptides)

^a Start and end position of the peptide in the protein sequence.
^b Peptide sequence identified by peptide mass sequencing. Asterisks indicate site of isotope tag.
^c “ani” stands for aniline labeling; “ba” stands for benzoic acid labeling.
Fig. 6. Coverage of the 188 identified proteins with regard to being identified only or both identified and quantified by the two labeling strategies of the ANIBAL approach. A, all proteins identified using the ANIBAL approach. B, proteins identified using aniline labeling alone. C, proteins identified using benzoic acid labeling alone. D, protein identified and quantified by ANIBAL. E, protein identified and quantified with aniline labeling alone. F, protein identified and quantified using benzoic acid labeling only. Proteins shared by the two labeling strategies are displayed in white boxes; those unique to one strategy are shown in black boxes.

Fig. 7. Ratio distribution for the 101 proteins identified and quantified using ANIBAL on L. lactis. Light-to-heavy (L/H) ratios for each of the 101 proteins identified, for which ratios were manually validated, are shown using a logarithmic scale. The mean ratio among the 101 proteins with S.D. is displayed as well as the mean number of quantifiable peptides per protein.

One objective of the ANIBAL method was to overcome this limitation by using cheaper reagents, thereby drastically decreasing the costs per sample to approximately $5 per experiment with aniline and $15 for benzoic acid, resulting in a total of approximately $20 for each combined ANIBAL experiment in which two isotope labelings are performed.

In summary, our ANIBAL approach exhibits the following features: (i) a reaction that is simple, straightforward, and quantitative; (ii) symmetrical twin chemistry targeting protein amino and carboxylic groups with identical peptide bond-linked benzoyl modification; (iii) convenient peak spacing (6 Da) and no chromatographic heavy/light discrimination ($^{13}$C reagents); (iv) high proteome coverage through targeting of two frequent protein functionalities; (v) identical MS and MS/MS behavior; (vi) increased retention time of modified peptides for binding of very hydrophilic peptides; and (vii) inexpensive reagents.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[5] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

To whom correspondence should be addressed. Tel.: 41-21-785-89-66; Fax: 41-21-785-94-86; E-mail: martin.kussmann@rdls.nestle.com.

REFERENCES

1. O’Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007–4021
2. Gorg, A., Weiss, W., and Dunn, M. J. (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4, 3665–3685
3. Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2, 3–10
4. Lahm, H. W., and Langen, H. (2000) Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis 21, 2105–2114
5. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999
6. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1, 376–386
7. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. Proc. Natl. Acad. Sci. U. S. A. 96, 6591–6596
8. Conrads, T. P., Alving, K., Veenstra, T. D., Belov, M. E., Anderson, G. A., Anderson, D. J., Lipton, M. S., Pasa-Toic, L., Udseth, H. R., Christer, W. B., Thrall, B. D., and Smith, R. D. (2001) Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and $^{15}$N-metabolic labeling. Anal. Chem. 73, 2132–2139
9. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154–1169
10. Schmidt, A., Kellermann, J., and Lottspeich, F. (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4–15

11. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842

12. Yao, X., Afonso, C., and Fenselau, C. (2003) Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *J. Proteome Res.* **2**, 147–152

13. Canas, B., Lopez-Ferrer, D., Ramos-Fernandez, A., Camafelita, E., and Calvo, E. (2006) Mass spectrometry technologies for proteomics. *Brief. Funct. Genomics Proteomics* **4**, 295–320

14. Chen, X., Sun, L., Yu, Y., Xue, Y., and Yang, P. (2007) Amino acid-coded tagging approaches in quantitative proteomics. *Expert Rev. Proteomics* **4**, 25–37

15. Julka, S., and Regnier, F. (2004) Quantification in proteomics through stable isotope coding: a review. *J. Proteome Res.* **3**, 350–363

16. Leitner, A., and Lindner, W. (2004) Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **813**, 1–26

17. Moritz, B., and Meyer, H. E. (2003) Approaches for the quantification of protein concentration ratios. *Proteomics* **3**, 2208–2220

18. Liu, P., and Regnier, F. E. (2002) An isotope coding strategy for proteomics involving both amine and carboxyl group labeling. *J. Proteome Res.* **1**, 443–450

19. Guillaume, E., Panchaud, A., Affolter, M., Desvergnes, V., and Kussmann, M. (2006) Differentially isotope-coded N-terminal protein sulphonation: combining protein identification and quantification. *Proteomics* **6**, 2338–2349

20. Panchaud, A., Guillaume, E., Affolter, M., Robert, F., Moreillon, P., and Kussmann, M. (2006) Combining protein identification and quantification: C-terminal isotope-coded tagging using sulfanilic acid. *Rapid Commun. Mass Spectrom.* **20**, 1585–1594

21. Storey, J. V. (1982) N-Hydroxysulfosuccinimide esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers. *Biochemistry* **21**, 3950–3955

22. Panchaud, A., Kussmann, M., and Affolter, M. (2005) Rapid enrichment of bioactive milk proteins and iterative, consolidated protein identification by multidimensional protein identification technology. *Proteomics* **5**, 3836–3846

23. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392

24. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951

25. Li, X. J., Zhang, H., Ranish, J. A., and Aebersold, R. (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.* **75**, 6648–6657

26. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658

27. Julka, S., and Regnier, F. E. (2004) Benzoyl derivatization as a method to improve retention of hydrophilic peptides in tryptic peptide mapping. *Anal. Chem.* **76**, 5799–5806

28. Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25**, 125–131

29. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods* **4**, 231–237