Chromatographic Isolation of Methionine-containing Peptides for Gel-free Proteome Analysis

IDENTIFICATION OF MORE THAN 800 ESCHERICHIA COLI PROTEINS†S

Kris Gevaert‡, Jozef Van Damme, Marc Goethals, Grégoire R. Thomas, Bart Hoorelbeke§, Hans Demol, Lennart Martens¶, Magda Puype, An Staes, and Joël Vandekerckhove

A novel gel-free proteomic technology was used to identify more than 800 proteins from 50 million Escherichia coli K12 cells in a single analysis. A peptide mixture is first obtained from a total unfractionated cell lysate, and only the methionine-containing peptides are isolated and identified by mass spectrometry and database searching. The sorting procedure is based on the concept of diagonal chromatography but adapted for highly complex mixtures. Statistical analysis predicts that we have identified more than 40% of the expressed proteome, including soluble and membrane-bound proteins. Next to highly abundant proteins, we also detected low copy number components such as the E. coli lac operon repressor, illustrating the high dynamic range. The method is about 100 times more sensitive than two-dimensional gel-based methods and is fully automated. The strongest point, however, is the flexibility in the peptide sorting chemistry, which may target the technique toward quantitative proteomics of virtually every class of peptides containing modifiable amino acids, such as phosphopeptides, amino-terminal peptides, etc., adding a new dimension to future proteome research. Molecular & Cellular Proteomics 1:896–903, 2002.

Recent progress in chromatography, mass spectrometry, and bioinformatics combined with access to a multitude of protein and oligonucleotide sequence (genomic) databases has opened the way for the development of proteome methodologies different from conventional two-dimensional gel analyses. An interesting approach starts from peptide mixtures generated from digests of total cellular extracts. Peptides are purified and identified by a combination of multidimensional chromatography and high-throughput mass spectrometry (1, 2). Alternatively, a subset of peptides, highly representative of the constituent proteins, can be isolated by modification and affinity selection (3, 4) or by covalent chromatography (5). Although these techniques have only been recently developed, the number of applications and technical improvements published during the last year undoubtedly illustrates its potential for future proteome research (6–12).

Here we introduce a peptide-based protein identification technique, different from the multidimensional protein identification technology (MudPIT)† or isotope-coded affinity tag method (2, 3) and is based on a modified version of diagonal chromatography (13): combined fractional diagonal chromatography (COFRADICTM). COFRADICTM was applied to select and identify methionine peptides in a total Escherichia coli tryptic digest. Using 50 million cells, we identified at least 800 different proteins. The procedure is ~100 times more sensitive than two-dimensional gel analysis and can be carried out in a fully automated manner. Next to its very high sensitivity, COFRADICTM offers the possibility to select from highly complex mixtures any subset of peptides containing amino acids that can be chemically or enzymatically modified.

EXPERIMENTAL PROCEDURES

E. coli Cells and Peptide Sample Preparation—E. coli HB2151 (K12) cells were grown to saturation densities (14), and a washed pellet corresponding to 250 × 10⁶ cells was taken up in 250 μl of freshly made 4 M urea in 100 mM phosphate buffer (pH 8.0) and lysed by sonication. The urea concentration was subsequently brought to 1 M with 100 mM phosphate buffer (pH 8.0). This E. coli lysate was divided in 200-μl aliquots: one serving for COFRADICTM analysis and the others for control two-dimensional gel analyses.

Sorting of Methionine-containing Peptides—A lysate corresponding to 50 × 10⁶ cells was digested overnight at 37 °C with 5 μg of trypsin (Promega Corp., Madison, WI) and stopped by acidification with trifluoroacetic acid. The digest was centrifuged to remove cell debris and insoluble material, and the supernatant was injected into

† The abbreviations used are: MudPIT, multidimensional protein identification technology; COFRADICT, combined fractional diagonal chromatography; LC-MS/MS, liquid chromatography-tandem mass spectrometry; Met-SO, methionine-sulfoxide; MALDI-TOF-MS, matrix-assisted laser dissociation ionization-time of flight-mass spectrometry; HPLC, high pressure liquid chromatography.
a narrow-bore reverse-phase ZORBAX® 300SB-C18 column (2.1 inner diameter × 150 mm, Agilent Technologies, Waldbronn, Germany) coupled to an Agilent 1100 Series Capillary LC system under the control of the Agilent ChemStation software modules. Following injection of the sample, a solvent gradient was developed at a constant flow rate of 80 µl/min. First, the column was rinsed with 0.1% trifluoroacetic acid in water (Baker HPLC analyzed, Mallinkrodt Baker B.V., Deventer, The Netherlands) (solvent A) for 10 min followed by a linear gradient to 70% acetonitrile (Baker HPLC analyzed) in 0.1% trifluoroacetic acid (solvent B) over 100 min (thus an increase of 1% of solvent B/min). We refer to this reverse-phase-HPLC separation as the primary run. Peptides were collected starting from 40 min on (corresponding to a concentration of 30% of solvent B) in a total of 48 fractions of 1 min (or 80 µl) each in a microtiter plate using the Agilent 1100 Series fraction collector. Fractions separated by 12 min were pooled as described in Table I and dried in a centrifugal vacuum concentrator. These dried fractions were redissolved in 70 µl of 1% trifluoroacetic acid in water and placed in the Agilent 1100 Series Well plate sampler. The methionine oxidation reaction proceeded in the injector compartment by transferring 14 µl of a freshly prepared aqueous 3% H2O2 solution to the vial containing the peptide mixture. This reaction proceeded for 30 min at 30 °C after which the sample was immediately injected onto the reverse-phase-HPLC column. Methionine-sulfoxide (Met-SO)-containing peptides elute under the given experimental conditions in a time window from 7 to 11 min in front of the bulk of the unmodified peptides and were collected in eight subfractions per primary fraction (Table I and Fig. 1). Subfractions with the same subscript and derived from the same secondary run (e.g. subfractions 12, 24, 36, and 48, of run 2A were pooled; see Table I and Fig. 1).

**Mass Spectrometry**—Matrix-assisted laser dissociation ionization-time of flight-mass spectrometry (MALDI-TOF-MS) analysis was carried out as described previously (15). This was done on one-quarter of each set of pooled subfractions.

For liquid chromatography-tandem mass spectrometry (LC-MS/MS) identification of methionine-containing peptides, we used 75% of the volume of the pooled subfractions, which were dried in a centrifugal vacuum concentrator and redissolved in 45 µl of 0.05% formic acid in 2:98 acetonitrile:water (by volume). This solution was split into two equal parts, which were used for two consecutive LC-MS/MS runs. Per run, 20 µl was loaded onto a 0.3-mm-inner diameter × 5-mm trapping column (PepMap, LC Packings, Amsterdam, The Netherlands) at a flow rate of 20 µl/min of solvent A. By valve switching, the trapping column was back-flushed, and the sample was loaded onto a nanoscale reverse-phase C18 column (75-µm-inner diameter × 150-mm PepMap column, LC Packings), and a binary solvent gradient was started. The solvent delivery system was run at a constant flow of 60 µl/min, and by using a 1:300 flow splitter (Accurate, LC Packings), 200 nl/min of solvent was directed into the nanocolumn. Peptides were eluted from the stationary phase using a gradient from 0 to 100% solvent B applied in 50 min. The outlet of the nanocolumn was in-line connected to a distal metal-coated fused silica Picotip™ needle (Picotip™ FS360-20-10-o-C7, New Objective, Inc., Woburn, MA) placed in front of the inlet of a Q-TOF mass spectrometer (Micromass UK Ltd., Cheshire, UK). Automated data-dependent acquisition with the Q-TOF mass spectrometer was initiated 20 min after the solvent gradient was started. The acquisition parameters were chosen such that only doubly charged ions were selected for fragmentation. After completion of the first LC-MS/MS run, a mass exclusion list was created containing all the selected ion masses of the peptides that were identified using Mascot (16). This exclusion list was then used for the second LC-MS/MS analysis on the remaining half of the material. The same procedure was used for the analysis of the other pooled subfractions.

**Peptide and Protein Identification**—The obtained collision-induced dissociation spectra in each LC-MS/MS run were automatically converted to a Mascot (www.matrixscience.com)-acceptable format (pklfomat format) using Proteinlynx available in the Micromass MassLynx software (version 3.4). Per Mascot search a maximum of 300 collision-induced dissociation peak lists were merged and used for peptide identification in a newly created database only containing E. coli methionine-containing tryptic peptides. For this list we used the E. coli K12 proteome (Refs. 17 and 18 and ftp.expasy.ch/databases/complete-proteomes/Ecoli.dat). The following search parameters were used: enzyme: trypsin, maximum number of missed cleavages: 1, variable modification: oxidation (Met), N-formyl (protein) and pyroglutamyl formation (amino-terminal Glu and Gin), peptide tolerance: 0.3 Da, MS/MS tolerance: 0.25 Da, and peptide charge: 2+. Only MS/MS spectra that exceeded Mascot’s significance level were retained.

A second independent method was developed to match experimental spectra with peptide sequences. This method was devised to quantify the probability of a spectrum to match a peptide sequence using two independent methods. As for Mascot, we used as reference the list of methionine-containing tryptic peptides (0 or 1 miscleavage) derived from the complete E. coli K12 proteome. Optional modifications used were oxidation of methionine and the formation of pyroglutamic acid from glutamate and glutamine. From each peptide sequence a list of b and y ions was derived to obtain a theoretical spectrum. Each experimental spectrum containing at least six peaks was scored for similarity with each theoretical spectrum with matching precursor ion mass.

The scoring scheme thus associates two distinct scoring methods that both quantify the likelihood of the experimental spectrum-theoretical spectrum match to occur. The first scoring takes into account the number of matching peaks with reference to the size of the spectra. The G-test was used to evaluate the probability for a match to differ from expectancy. The second scoring focuses on the amplitude of the matching peaks with reference to the relative amplitude of the spectra. The score indicates the deviation from expectancy of the correlation coefficient between spectra. The final score is the product of the two intermediate scores. Because we here combined a probabilistic approach and signal correlation, we named this algorithm Procorr™. Procorr™ is accessible at penyfan.rug.ac.be/procorr/, and details will be published in the future.2

The same MS/MS spectra were subjected to Procorr™ analysis, and the matches differing from expectancy were retained (confidence level of 98%). Since the number of identified proteins is −2.5 times lower than the number of identified spectra, a level of confidence of 98% for spectrum identification corresponds to a level of confidence of at least 95% for protein identification.

**RESULTS**

**The Concept of COFRADIC™**—Our gel-free proteome approach starts from a protein cell lysate that is digested with trypsin. The resulting peptide mixture may contain up to 50,000 and most likely even more different components. Out of this complex mixture, we select a subset of peptides, which is highly representative of the parent proteins originally present in the lysate. In this respect our approach is similar to the isotope-coded affinity tag technique (3) or the covalent chromatography method (5) since we also select for a subset of representative peptides. Technically, however, we do not use

---

2 G. R. Thomas, J. Vandekerckhove, K. Gevaert, and N. Bernitsis, in preparation.
any tagging chemistry combined with affinity selection but rather use a specific chemical or enzymatic modification on peptides containing rare amino acids, thereby altering their chromatographic properties. When such modification reactions are carried out in between two consecutive identical chromatographic runs, the subset of altered peptides will change elution times in the second run, while the non-modified peptides will elute at the same predictable positions. Such a strategy, in which a modification is carried out between two identical runs, to induce a shift in the chromatographic behavior of the modified components, was previously used on peptide mixtures derived from single proteins and was called “diagonal chromatography” (13). Its name and concept was derived from the technique of “diagonal electrophoresis,” which was introduced in a series of elegant articles by the group of Hartley and colleagues (19).

In Fig. 1 we illustrate the adaptation of diagonal chromatography to the sorting of subsets of peptides from very complex mixtures. During the first chromatographic step (run 1), peptides are separated and collected in fractions of appropriate time intervals (Fig. 1A). A specific modification reaction is then carried out in every fraction, altering the properties of a subset of peptides. Every fraction could then be rerun under the same chromatographic conditions, referred to as the secondary run (run 2). The altered peptides will now shift in run 2 compared with their original positions in run 1. The unaltered peptides do not show this shift (Fig. 1B). The shifted peptides can be collected for analysis.

The number of secondary runs, which in principle should be equal to the number of fractions collected in the primary run, can be reduced by combining primary fractions. This is done in such a way that the shifting peptides of a given fraction do not overlap with the non-shifting peptides of neighboring fractions. Depending on the extent of the shifts, up to four or more primary fractions can be combined, thereby reducing the number of secondary runs by the same factor (Fig. 1C). The entire sorting procedure can further be shortened, if necessary, by using two or more columns operating in a synchronous mode. The unaltered peptides are mostly discarded, while the sorted peptides are either on-line analyzed by mass spectrometry or collected for identification in a ternary LC-MS-coupled system. The procedure, in which fractions of the first chromatographic step are combined, modified, and run in a diagonal chromatographic manner is therefore called COFRADIC™.

Application of COFRADIC™ to the E. coli Proteome—An analysis of the predicted proteome of different model organisms revealed that methionine-containing peptides provided the best representation of the predicted proteins. For instance, for the E. coli proteome, between 99.7 and 95.8% of the predicted proteins contained at least one methionine residue (depending whether the initiator methionine is counted or not). Only 85.4% of the proteins contained cysteine. The same trend in amino acid representation is also observed in other model organisms (data not shown). We therefore decided to select for Met-containing peptides and used the oxidation of methionine to Met-SO as the sorting vehicle since the sulfoxide is more hydrophilic than the non-modified peptide.

COFRADIC™ was therefore used to sort Met peptides
present in a tryptic digest of a total, unfraccionated, 4 M urea extract of 50 × 10^6 E. coli cells. Forty-eight fractions of 80 μl (1 min) each were collected during the primary reverse-phase-HPLC run. The first fraction was taken between 40 and 41 min (number 48) following the start of the run (Fig. 1A), and the last fraction was taken between 87 and 88 min (number 48) following the start of the run (Fig. 1A). In every fraction we converted the methionine peptides to their sulfone derivative by a simple oxidation step. Conditions were established in which neither Cys nor Trp residues were oxidized and where Met residues were not converted into their sulfones. In the chromatographic conditions used, the oxidized Met peptides generally display a hydrophilic shift ranging from 1 to 7 min. The extent and the range of the hydrophilic shifts were similar for early- and late-eluting peptides.

The sorted Met-SO peptides were collected during a 6-min broad interval starting 7 min before the elution time of the unaltered peptides. This window is thus 6 times broader than that in which peptides eluted during the primary run. This is an important aspect of COFRADICTM, the sorted peptides elute in a less compressed manner, thereby facilitating their identification by further LC-MS/MS analysis.

In the COFRADICTM mode (see Table I), we combined the primary fractions 1 (40–41 min), 13 (52–53 min), 25 (64–65 min), and 37 (76–77 min) and collected the sorted Met-SO peptides in the secondary run during the intervals 33–39 min (11-8), 45–51 min (13-8), 57–63 min (25-8), and 69–75 min (37-8) in eight subfractions. Thus during the secondary run, we collected the sorted Met-SO peptides in eight “subfractions” (indexed 1–8) linked with the number of the primary fraction. For instance, 1-8 indicates the eighth secondary Met-SO subfraction derived from the primary fraction number 1. Identically subindexed subfractions were pooled and dried prior to MALDI or LC-MS/MS analysis. Thus we pooled 11, 13, 25-1, and 37-1 and further 12, 13, 25-2, and 37-2, etc. In Fig. 1C we show the UV absorption profile (214 nm) during the Met-SO-sorting procedure on the combined primary fractions 8 (47–48 min), 20 (59–60 min), 32 (71–72 min), and 44 (83–84 min) and the collection of the Met-SO peptides during intervals 40–46 min (11-8), 52–58 min (25-8), 64–70 min (32-8), and 76–82 min (44-8). A complete fraction collection protocol and detailed time table for the methionine peptide-sorting procedure is provided as additional information in Table I. The entire sorting procedure thus includes one primary run followed by 12 secondary runs (2A–2L), which can be completed in less than 24 h.

An aliquot (1/4) of the combined secondary subfractions was analyzed by MALDI-TOF-MS during which Met-SO peptides could be recognized by their typical neutral loss of methanesulfenic acid (loss of 64 atomic mass units). In this way, we detected at least 1720 different tryptic peptides, 1618 of which contained at least one oxidized methionine residue (data not shown). Thus, less than 6% of the sorted peptides were either not recognized as Met peptides due to lack of specific fragmentation or did not contain methionine and slipped through during the sorting process.

For further individual peptide and protein identification we used a LC-MS/MS configuration using the remaining 3/4 of the material. For this, we carried out 2 × 96 ternary runs in an automated manner. The obtained information was probed against an E. coli K12 database consisting of only Met-containing peptides. This database consisted of 31,746 peptides in the mass range between 780 and 2400 Da and was generated allowing 0 and 1 miscleavage for trypsin in the predicted K12 proteome. The database size reduction was possible because of the low number of non-methionine peptides identified in an initial MALDI-MS screening exercise (see above).

The Mascot search algorithm assigned 2167 MS/MS spec-
118 proteins were found with Procorr™ but not with Mascot, found by both algorithms and are therefore highly relevant, total, 872 different proteins were identified: 689 proteins were thus proteins could have been falsely assigned (Table II). In case a maximum of 43 spectra (0.02 98% for 1350 peptides and for 807 different proteins. In this identification algorithm providing an overall confidence of cell shown in Fig. 2.

A number of important aspects of the COFRADIC™ approach become apparent by further data analysis (Table II). The Mascot column lists all data obtained by applying the Mascot algorithm. The two other columns summarize the data obtained by applying the Procorr™ algorithm: one column with a peptide scoring confidence of 95% and a second column with a peptide scoring confidence of 98%. Only the last column was retained for further discussions in the text.

|                      | Mascot | Procorr™ (95%) | Procorr™ (98%) |
|----------------------|--------|---------------|---------------|
| Number of identified spectra | 2167   | 2312          | 2147          |
| Number of identified peptides | 1326   | 1457          | 1350          |
| Number of identified proteins | 754    | 882           | 807           |
| Number of proteins in K12 proteome | 4349   | 4349          | 4349          |
| Proportion of proteome identified (%) | 17.34  | 20.28         | 18.56         |
| Proportion of expressed proteome identified (estimation) (%) | 34.67  | 40.56         | 37.11         |
| Number of identified proteins identified by Mascot | Unknown | 695           | 689           |
| Number of identified proteins present in the SWISS-2DPAGE database | 155/268 | 163/268       | 155/268       |
| Number of identified cytoplasmic proteins | 290    | 309           | 297           |
| Number of identified periplasmic proteins | 39     | 42            | 41            |
| Number of identified inner membrane proteins | 56     | 68            | 56            |
| Number of identified outer membrane proteins | 11     | 14            | 12            |
| Number of identified proteins with GRAVY ≥ 0.3 | 17     | 32            | 26            |
| Expected number of wrong spectrum identifications (false positives) | Unknown | 115.6         | 42.9          |
| Average number of identified spectra per protein | 2.87   | 2.62          | 2.66          |
| Average number of identified peptides per protein | 1.75   | 1.85          | 1.87          |

Gel-free Proteomics Using Isolated Methionine Peptides

Although the original complexity of the peptide mixture is reduced by approximately a factor 5, the flux of peptides passing into the ion source of the mass spectrometer is still too high for individual peptide detection. Given this situation, proteins represented by a large number of Met peptides are expected to be detected with higher probability than proteins with few methionines. Such a bias toward Met-rich proteins is indeed observed by relating the percentage of identified proteins with the number of methionines in these proteins. We observe a nearly linear increase from 18% for the total predicted proteome to 43% for proteins with 10 or more Met residues (data not shown). This percentage is still increasing, reaching a plateau at 60%, when proteins with more than 17 methionines are considered (the latter value is statistically weak because of the low number of proteins involved). Based on these observations and since it seems unlikely that Met-rich proteins are differently represented in the cell lysate compared with Met-poor proteins, we assume that approximately the same percentage (about 50%) of the predicted E. coli proteins may be detectable in our system. This means that we most likely identified ∼37% of the proteins actually present in the cell lysate.

DISCUSSION

Our proteome approach is a peptide-based approach. The methionine peptides are selected by two repeated reverse-phase-HPLC runs with an oxidation step in between. There are no protein premodification steps. This very simple procedure therefore guarantees a high overall sensitivity because peptide losses due to manipulations are limited, although
there is still some room for improvement by, for instance, omitting vacuum drying steps, which result in sample loss, and by downsampling the column dimensions used for the chromatographic isolation. The COFRADICTM analysis was carried out on an extract (no prefractionation was done) of 50 million E. coli cells, corresponding in volume and protein content to ~50,000 hepatocytes. Thus COFRADICTM offers a perfect tool to study the protein profile of biological samples, which could not be addressed previously, using two-dimensional gel analysis. For instance, very small groups of cells displaying defined biological functions such as small biops, early stages of embryonic development, or even parts of individual cells can be studied. COFRADICTM also allows the detection of very abundant proteins such as ribosomal proteins, as well as proteins, known to be expressed at very low copy number, e.g. lac repressor. Thus a simultaneous detection of proteins present in ratios of 1:10,000 or even more is now possible, illustrating the high dynamic range of our technology.

However, it should be clear that methionine COFRADICTM, as most other described peptide-based proteomic technologies, does not allow studying protein modification, protein processing/degradation, and the determination of different protein isoforms on a global scale. Typically, these topics have been addressed by separating protein mixtures on two-dimensional gels followed by Western blotting using specific antibodies. Nevertheless, it should be noted that the concept of COFRADICTM allows the isolation of different representative peptides if these can be specifically modified. We have recently altered the sorting chemistry such that amino-terminal peptides of all proteins present in a mixture can be isolated. These types of peptides now allow the analysis of protein amino-terminal processing on a global scale in a gel-free manner. Similarly, we are developing a sorting chemistry to specifically isolate phosphorylated peptides out of protein digestion mixtures.

Following our analysis, we detected an unexpectedly large number of membrane proteins. This is a particular property of the peptide-based approach and was previously noticed by the group of J. Yates (2) using the MudPIT approach. Indeed, proteins can be in situ trimmed at their extramembranous parts, generating a set of hydrophilic peptides, some of which can function as signature peptides. This approach, in which COFRADICTM may play a crucial role, may offer a valuable alternative to procedures in which membrane proteins are isolated using new types of detergents or novel extraction protocols prior to gel separation (21, 22).

Several peptide-based proteome approaches were recently described. The MudPIT technology of the group of J. Yates (2) identified more than 1500 different proteins from yeast. This impressive number was reached by accumulating data from three prefractionated lysates each containing up to 400 µg of protein. This represents at least 100 times the amount of starting material used in our studies. In addition, MudPIT does not include any presorting step, which makes it very difficult to separate the high number of peptides, thereby suffering from experiment to experiment reproducibility.

COFRADICTM follows a preselection step introduced to reduce the number of peptides. This is similar to the isotope-coded affinity tag approach (3) and to the covalent chromatography method (5). However, COFRADICTM is much more versatile than previous methods because any peptide carrying a group that can be specifically and quantitatively modified can in principle be sorted. In the example shown here, we have used one of the simplest modification reactions in pro-

\[ \text{K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G. R. Thomas, and J. Vandekerckhove, submitted.} \]
tein chemistry: the conversion of a methionine side chain to its more hydrophilic sulfoxide derivative. A similar one-step reaction could also be proposed for the selection of cysteine peptides: for instance the reduction of -S-S-R groups to the more hydrophilic thiol groups will provoke a hydrophilic shift for SH-containing peptides during the second chromatographic step. An additional advantage is that all these different sorting protocols can be carried out by the same robots in a fully automated manner.

While peptide-based proteomics clearly offers aspects of high sensitivity, broad protein coverage, and full automation, protein identification is, more than in conventional two-dimensional gel approaches, dependent on the confidence by which peptides are identified. Thus both the quality of the MS/MS fragmentation spectra of the individual peptides and peptide identification algorithms to interpret these spectra are therefore of utmost importance.

To provide more confidence to protein identification, we used Mascot as the first searching algorithm, but we additionally used a second in-house-developed algorithm, Procorr™. Since the latter is based on a combination of parameters, we could also use higher stringency criteria while still identifying more proteins. Using the latter algorithm, we identified at least 807 different proteins with 95% probability. This is 53 proteins more than with Mascot. Taken together, both algorithms identify 872 different proteins, which can be classified into three categories: those identified by both algorithms, those identified with a peptide probability score of at least 98% by Procorr™, and finally those identified with a peptide probability score of 95% by Mascot only. These 872 proteins represent almost 40% of the estimated expressed E. coli proteome as calculated from the identification score of Met-rich proteins (see above).

Fig. 3 relates the distribution curves for the acidic proteins detected by COFRADIC™ with the total predicted proteome and those reported in the SWISS-2DPAGE database. COFRADIC™ detects more than 4 times more proteins. The difference between the two data sets is even more striking when the basic proteins are considered where a large number of proteins found by COFRADIC™ are missing in the two-dimensional gel approaches.

For differential quantitative analysis we can use the isotope labeling of the peptide COOH terminus. This trypsin-catalyzed water-oxygen incorporation has been found for some time (see Ref. 23, for instance) but has only been applied recently for peptide and protein quantification (24). This procedure fits extremely well in the COFRADIC™ protocol because it does not need additional labeling reactions and purifications. The procedure as well as the application will be the subject of a separate article in which basic questions such as quantitative aspects of oxygen incorporation, possible back-exchange during the peptide sorting process, co-elution of ¹⁶O and ¹⁸O isopeptides, and exact measurements of the ratio of the isotope variants will be addressed.

In conclusion, we have demonstrated that COFRADIC™ constitutes a valid alternative for peptide-based proteomics. It is very sensitive and is characterized by a broad protein coverage, including abundant and rare; large and small; and acidic, basic, and hydrophobic proteins. In the example of the E. coli proteome we identified 872 different proteins with very high probability scores. This number could have been considerably larger but was restricted by the limited capacity of peptide ion selection in the mass spectrometer used in this study. We are therefore confident that novel high-throughput machines may provide a complete coverage of peptides and thus a full coverage of the expressed proteome.

COFRADIC™ offers more to proteomics than just a methionine peptide-sorting technology. COFRADIC™ is a total concept that may become an indispensable tool for future proteomics. The high sensitivity of COFRADIC™-based proteome analysis clearly allows analysis of only minute amounts of biological material, which was until now not possible using “classical” proteomic technologies. COFRADIC™ will make it possible to carry out targeted forms of proteomics such as detecting and measuring protein cleavage and processing in total cellular lysates or post-translational modifications.

Acknowledgment—We thank Dr. Nikos Berntenis for computational support.

* This work was supported by Fund for Scientific Research-Flanders (F.W.O.-Vlaanderen) Grants G.0044.97, G.0225.98, and

---

4 K. Gevaert, A. Staes, J. Van Damme, H. Demol, and J. Vandekerckhove, in preparation.
G.0050.02 and by European Commission Grant QLK2-200-31536. 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.

§ The on-line version of this article (available at http://www.mcponline.org) contains Supplemental Table III.

‖ Present address: VIB Proteomics Core Facility, Rijvisschestraat 118, B-9052 Zwijnaarde, Belgium.

¶ A postdoctoral fellow of the Fund for Scientific Research-Flanders (Belgium) (F.W.O.-Vlaanderen).

§§ Present address: VIB Proteomics Core Facility, Rijvisschestraat 118, B-9052 Zwijnaarde, Belgium.

‖‖ A research assistant of the Fund for Scientific Research-Flanders (Belgium) (F.W.O.-Vlaanderen).

¶¶ To whom correspondence should be addressed: Dept. of Biochemistry, Flanders Interuniversity Inst. for Biotechnology, Ghent University, A. Baertsoenkaai 3, B-9000 Ghent, Belgium. Tel.: 32-93313303; Fax: 32-93313597; E-mail: joel.vandekerckhove@rug.ac.be.

REFERENCES

1. Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K., Pasa-Tolic, L., Veenstra, T. D., Lipton, M. S., Udseth, H. R., and Smith R. D. (2001) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. Anal. Chem. 73, 1766–1775

2. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–248

3. 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

4. Geng, M., Ji, J., and Regnier, F. E. (2000) Signature-peptide approach to detecting proteins in complex mixtures. J. Chromatogr. A 870, 295–313

5. Wang, S., and Regnier, F. E. (2001) Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography. J. Chromatogr. A 924, 345–357

6. 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

7. Oda, Y., Nagasu, T., and Chat, B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat. Biotechnol. 19, 379–382

8. Zhou, H., Watts, J. D., and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. Nat. Biotechnol. 19, 375–378

9. Zhou, H., Ranish, J. A., Watts, J. D., and Aebersold, R. (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat. Biotechnol. 20, 512–515

10. Cagney, G., and Emili, A. (2002) De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. Nat. Biotechnol. 20, 163–170

11. Faraco, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat. Biotechnol. 20, 301–305

12. Yu, L.-R., Johnson, M. D., Conrads, T. P., Smith, R. D., Morrison, R. S., and Veenstra, T. D. (2002) Proteome analysis of camptothecin-treated cortical neurons using isotope-coded affinity tags. Electrophoresis 23, 1591–1598

13. Cruickshank, W. H., Malchy, B. L., and Kaplan, H. (1974) Diagonal chromatography for the selective purification of tyrosyl peptides. Can. J. Biochem. 52, 1013–1017

14. Rossen, S., Dewitte, D., Vandekerckhove, J., and Ampe, C. (1997) A phage display technique for a fast, sensitive, and systematic investigation of protein-protein interactions. J. Protein Chem. 16, 499–503

15. Gevaert, K., Demol, H., Puype, M., Broekaert, D., De Boeck, S., Houthaever, T., and Vandekerckhove, J. (1998) A peptide concentration and purification method for protein characterization in the subpicomole range using matrix assisted laser desorption/ionization-postsource decay (MALDI-PSD) sequencing. Electrophoresis 18, 2950–2960

16. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567

17. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462

18. Bairach, A., and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–48

19. Brown, J. R., and Hartley, B. S. (1966) Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A. Biochem. J. 101, 214–228

20. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132

21. Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. Electrophoresis 20, 660–663

22. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? Electrophoresis 21, 1054–1070

23. Schmøier, M., Jedrzejewski, P., and Lehmann, W. D. (1996) Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Electrophoresis 17, 945–953

24. Migorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Korner, R., Sonksen, C. P., and Roepstorff, P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using 18O-labeled internal standards. Rapid. Commun. Mass Spectrom. 14, 1226–1232