research articles Journal of proteome research

# Shotgun Proteome Analysis of *Rhodospirillum rubrum* S1H: Integrating Data from Gel-Free and Gel-Based Peptides Fractionation Methods

**Felice Mastroleo,[†,‡] Baptiste Leroy,[†] Rob Van Houdt,[‡] Catherine s' Heeren,[†] Max Mergeay,[‡] Larissa Hendrickx,[‡] and Ruddy Wattiez*[,†]**

*Department of Proteomics and Protein Biochemistry, University of Mons, Mons, Belgium, and Expert group Molecular and Cellular Biology, Belgian Nuclear Research Center (SCK•CEN), Mol, Belgium*

**Received January 5, 2009**

Beside bioreactor modeling studies, the molecular characterization of life-support organisms appeared to be essential to complete their global behavior picture, in particular, culture conditions. Using a combination of LC-MS/MS approaches with gel-free and gel-based peptides fractionation steps, we identified 932 proteins from the α-proteobacterium *Rhodospirillum rubrum* S1H. In addition, abundance data were retrieved using the recently developed emPAI approach which takes into account the number of sequenced peptides per protein. This work has also allowed identification of new and specific proteins for the Rhodospirillaceae family.

**Keywords:** *Rhodospirillum rubrum* S1H • alpha-proteobacterium • bacterial proteome • MudPIT • isoelectric focusing • LC-MS/MS • emPAI • MELiSSA

## 1. Introduction

The Micro-Ecological Life Support System Alternative (MELiSSA) project initiated by the European Space Agency (ESA) aims to sustain the human crew for long haul space exploration missions.[1,2] It consists of 5 interconnected processes including the human compartment and four bioreactors; each bioreactor achieves a particular step in the biotransformation of the waste produced by the crew.[2,3]

Although inspired by natural ecosystems,[2,4] life-support systems remain engineered systems, and as such, the needs and the requirements of their three main partnering groups—the energy source of the primary producers, the nutritional quality of the food sources for the consumers and the efficiency of recycling of the decomposers—will be adapted accordingly.[3,4] Thus, so far, studies on the MELiSSA organisms have mainly focused on bioreactor culture technologies and modeling.[3,5−8] However, a multidisciplinary approach including reactor engineering for space applications, mathematical models, Earth-based test beds, space mission experiment, risk analysis, psychological studies or dietetics, toxicity studies and biological studies is necessary to design and to build a reliable system.[3] Therefore, the molecular characterization of the MELiSSA organisms appeared to be essential to complete the global behavior picture of these organisms in particular culture conditions. In that context, the proteomic study of the α-proteobacterium *Rhodospirillum rubrum* S1H, chosen to take care of the carbon transformation step in the second compartment of the MELiSSA loop, was initiated. While the MELiSSA concept implies photoheterotrophic culture conditions for *R. rubrum*, the present study dealt with chemoheterotrophic culture conditions because it aimed to set up the tools for integrated studies of space samples that already flew to the International Space Station using the latter culture conditions. However, the experimental approaches developed in this work will be used in the future to study the expression of proteins under different bacterial growth conditions (photoheterotrophic *versus* chemoheterotrophic cultures, photoheterotrophic culture in space condition, etc.).

Proteome analysis implies the ability to separate proteins with high resolution and reproducibility prior to characterization by mass spectrometry (MS). Within that purpose, the shotgun proteomic approach, also known as multidimensional protein identification technology (MudPIT) MS was used. In shotgun analyses, protein mixtures are digested to peptides, which then are analyzed by liquid chromatography−tandem mass spectrometry (LC-MS/MS) to identify peptide and protein sequences.[9] Shotgun proteomic platforms use multidimensional peptides separations to fractionate complex peptide mixtures prior to reverse phase LC-MS/MS. Each fraction presents a simplified peptide mixture for LC-MS/MS and this enables acquisition of MS/MS spectra for lower abundance peptides.

Recently, an alternative to using strong cation exchange to separate peptides prior to LC-MS/MS was described where trypsin digested proteins were separated over an immobilized pH gradient (IPG) using isoelectric focusing (IEF), and this technique was termed peptide IPG-IEF.[10,11]

* Corresponding author: Department of Proteomics and Protein Biochemistry, 5 avenue du Champs de Mars, University of Mons, B-7000 Mons, Belgium. E-mail: ruddy.wattiez@umh.ac.be. Phone/Fax: +3265373311/20

† University of Mons.

‡ Expertise group for Molecular and Cellular Biology, Belgian Nuclear Research Center (SCK•CEN).



Published on Web 02/25/2009

10.1021/pr900007d CCC: $40.75 

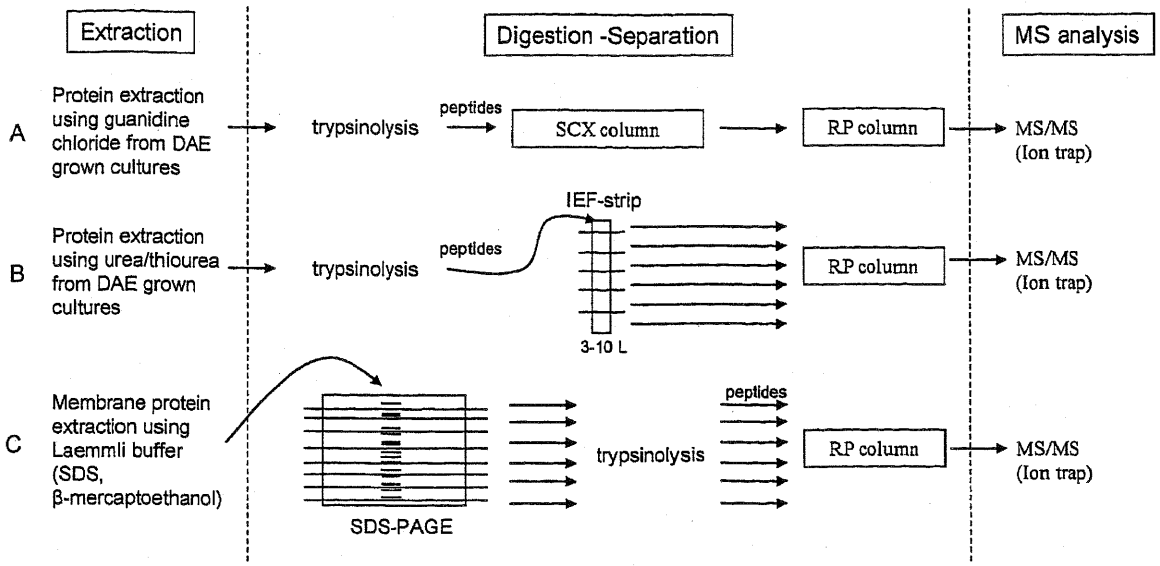



**Figure 1.** Summary representation of the sample workflows compared in this study. Details of each workflow are detailed in the text. DAE, dark aerobic; IEF, isoelectric focusing; MS, mass spectrometry; MS/MS, tandem mass spectrometry; SCX, strong cation exchange; RP, reverse phase; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

Separating peptides directly on the IPG strip is more effective than separating proteins on gel and then performing an in-gel digestion. Indeed, using the latter method, some peptide bonds are not accessible to the enzyme due to the trapping of protein substrate in the gel and not all the peptides produced during the digestion can diffuse freely out from the gel. On the other hand, due to the high hydrophobicity of the membrane fraction, protein separation using mono-dimensional SDS-PAGE followed by in-gel digestion was recommended.[9]

In that context, three peptide fractionation methods were tested: (i) the common approach using on-line strong cation exchange and two alternative approaches, (ii) off-line isoelectric focusing[10,11] and (iii) off-line SDS-PAGE[12] (Figure 1).

Except for the study of Selao et al.[13] who used the 2D-gel approach and identified 44 differentially expressed proteins, no high-throughput proteomic studies have been carried out on *R. rubrum* S1H. Therefore, the use of shotgun proteomics, for the first time, will help in the molecular characterization of this bacterium that occupies a key position in the MELiSSA process.

## 2. Material and Methods

**2.1. Strain and Medium.** *R. rubrum* strain S1H was obtained from the American Type Culture Collection (ATCC25903). For each experimental conditions, *R. rubrum* S1H was cultivated chemoheterotrophically in biological triplicates using 75 $cm^2$ cell culture flask with vented cap to ensure consistent gas exchange (BD Falcon), filled with 50 mL of Sistrom medium A[14] containing 20 mM Na-succinate as carbon source. Flasks were packed with aluminum foil and cells were kept rotating at 40 rpm and 30 °C for dark aerobic growth until mid-exponential phase ($OD_{680nm} \approx 0.3$) was reached. Thereafter, the 3 independent cultures were pooled, centrifuged at 7 000 rpm for 10 min at 4 °C and the resulting pellet was kept at −80 °C until use.

**2.2. 2D-LC MS/MS Approach.** Our first MudPIT experimental setup included on-line strong cation exchange followed by reverse phase chromatography and ion-trap MS/MS analysis (Figure 1A).

**2.2.1. Protein Extraction and Quantification.** Before protein extraction, the bacterial pellet was washed 3 times with 50 mM phosphate buffer saline, pH 7.2 (Buffered saline pack, Pierce). Samples were cleared by centrifugation at 13 500 rpm for 15 min at 4 °C. Protein samples were obtained by high power sonication (U50 control, IKA labortechnik, Germany) of the washed bacterial pellet suspended in one pellet volume of 6 M guanidine chloride solution. Sonication was performed with 3 cycles of 10 s (40 % amplitude, cycle 1) followed by 1 min on ice. Samples were cleared by centrifugation at 13 500 rpm for 15 min at 4 °C. Proteins in supernatant were reduced with 5 mM Tris(2-carboxyethyl) phosphine at 60 °C for 30 min and alkylated with 0.4 μM iodoacetamide at 25 °C for 30 min. Proteins were recovered by acetone precipitation (1 h) with a acetone/protein ratio of 4/1. After 15 min centrifugation at 13 500 rpm and acetone evaporation, the resulting pellet was dissolved in 25 mM Tris-HCl (pH 8)/2 M urea. Protein concentration was measured using the Non-Interfering Protein Assay Kit (Calbiochem, Germany) with bovine serum albumin as a protein standard. Overnight enzymatic digestion was carried out with modified sequencing grade trypsin (Promega) at an enzyme/substrate ratio of 1/50 at 37 °C.

**2.2.2. Chromatographic Separation.** Eighteen micrograms of tryptic peptides obtained by enzymatic digestion was separated using an Ultimate 3000 chromatographic system (Dionex). The chromatographic system was composed of two columns connected in series, a strong cation exchange (SCX) column (POROS 10S, 10 cm) and a reverse phase column (C18, 15 cm, i.d. 75 μm) both from Dionex. Peptides were injected in the SCX column and the flow through was directly loaded on a precolumn (C18 Trap, Dionex). The peptides were then washed on the precolumn with the loading solvent (5 % (v/v) acetonitrile (ACN), 0.025 % (v/v) TFA) during 15 min. After washing, the ACN gradient was started and peptides were separated on the reverse column based on their hydrophobicity. The ACN gradient was 4−37% of solvent B (solvent B: 80% ACN, 0.08% formic acid) in 100 min, 37−57% of solvent B in 10 min and 57−90% of solvent B in 10 min; 90% of solvent B was maintained for 10 min and then reset to 4%. Column

equilibration at 4% of solvent B was allowed for 10 min. After this first cycle, a first lot of peptides was eluted from the SCX column by injection of 20 $\mu$L of loading solvent containing 1 mM NaCl. The eluted peptides were loaded on the precolumn and separated as described for the flow through. This sequence of elution from the SCX column with injection of a salt plug followed by separation on the reverse phase column was repeated 9 times with salt plug concentration of 2, 5, 10, 25, 50, 100, 200, 500, and 1000 mM NaCl.

**2.2.3. Mass Spectrometry Analysis.** The separated peptides were on-line analyzed using an ion-trap mass spectrometer (HCT ultra PTM discovery, Bruker Daltonics, Germany). This mass spectrometer was used in the positive mode and spectra were acquired in a data dependent manner: MS scan range = 300−1 500 $m/z$, maximum accumulation time = 200 ms, Ion Charge Current (ICC) target = 200 000. The top 4 most intense ions in the MS scan were selected for MS/MS in dynamic exclusion mode: mode = ultrascan, absolute threshold = 75 000, relative threshold = 1%, excluded after spectrum count = 1, exclusion duration = 0.15 min, averaged spectra = 5, ICC target = 200 000.

**2.2.4. Data Extraction and Database Search.** Peptide peaks were detected and deconvoluted automatically using the Data Analysis 3.4 software (Bruker Daltonics, Germany). The mass list was generated automatically in the Mascot Generic Files format and searched against a local copy of the NCBInr released 20080608 database using an in-house Mascot 2.2 server (Matrix Science) for protein identification ($p < 0.05$). The default search parameters used were taxonomy = *Rhodospirillum rubrum* S1 (ATCC11170), enzyme = trypsin, maximum missed cleavages = 1, fixed modification = Carbamidomethyl (C), variable modification = Oxidation (M), peptide mass tolerance ±1.5 Da, fragment mass tolerance ±0.5 Da, peptide charge = 1+, 2+ and 3+; instrument = ESI-TRAP. Only sequence identified with a Mascot score of at least 50 were considered. For each protein identified from one single peptide, MS/MS spectra were manually evaluated. In addition, the confidence level of the latter protein identification was increased by including in the analysis information about redundant proteins coming from replicas of the same experiment or from different experimental approaches.

The false discovery rate was estimated using the Mascot 'decoy' option. If TP was the true positive matches and FP was the false positive matches, the number of matches in the target database was TP + FP and the number of matches in the decoy database was FP. The quantity that was reported was the False Discovery Rate (FDR) = FP/(FP + TP).[15]

**2.3. Peptide IPG-IEF Prior to LC MS/MS Separation. 2.3.1. Protein Extraction and Quantification.** Protein samples were obtained by high power sonication (U50 control, IKA labortechnik, Germany) of the bacterial pellets suspended in one pellet volume 6 M urea/2 M thiourea solution in 10 mM HEPES (pH 8). Sonication and sample clearing by centrifugation was performed as mentioned above. The supernatants were first reduced with 10 mM DL-dithiothreitol and then alkylated with 55 mM iodoacetamide, respectively. Protein concentration of the supernatants was measured by the Bradford method,[16] according to the Bio-Rad Protein Assay kit, with bovine $\gamma$-globulin as a protein standard. A total of 100 $\mu$g from each biological replicate was pooled. Prior to overnight digestion by trypsin (Promega) at an enzyme/substrate ratio of 1/50 at 37 °C, the urea/thiourea concentration was reduced to 2 M by dilution with 10 mM $NH_4HCO_3$. A total of 300 $\mu$g of digested proteins was then desalted using HyperSep SpinTip C18 (Thermo electron) following manufacturer's instructions. The sample was evaporated to dryness in a vacuum centrifuge (Heto, Drywinner, Denmark) and resuspended in 350 $\mu$L of 8 M urea supplemented with a trace of bromophenol blue. A second protein sample was prepared following the same procedure to increase the peptide coverage.

**2.3.2. IEF Separation.** Digested proteins (300 $\mu$g) were used to passively rehydrate linear pH 3−10 18 cm IPG strips (GE Healthcare) overnight. Isoelectric focusing was conducted on a Protean IEF cell (Bio-Rad) with a current limit of 50 $\mu$A per strip at 20 °C with the following focusing program: 300 V for 1 h, a gradient to 1000 V for 1 h, a gradient to 4000 V for 3 h, a gradient to 8000 V for 3 h and 8000 V until 100 000 V/h was reached. This last step lasted approximately 9 h.

The strips were then cut (with plastic backing still in place) with a scalpel blade in 30 equal length pieces (i.e., 5 mm). Peptides were extracted from each fraction by incubation in 100 $\mu$L of 0.1 % (v/v) formic acid for 1 h at room temperature. The extraction was repeated twice and subsequently combined with the initial fraction. Combined peptide extracts from each fraction were concentrated in a vacuum centrifuge (Heto, Drywinner, Denmark) to approximately 25 $\mu$L. Each fraction was desalted using Hypersep ziptip C18 tip (Thermo electron) following the manufacturer's instructions.

**2.3.3. Liquid Chromatography-Mass Spectrometry.** The peptides were analyzed by nanoLC-MS/MS using an LC Ultimate 3000 system (Dionex) coupled with an HCT ultra plus mass spectrometer (Bruker Daltonics, Germany). A total of 6 $\mu$L of each fraction was loaded onto a precolumn (C18 Trap, 300 $\mu$m i.d. × 5 mm, Dionex) using the Ultimate 3000 system, delivering a flow rate of 20 $\mu$L/min of loading solvent. After a 10 min desalting step, the precolumn was switched online with the analytical column (75 $\mu$m i.d. × 15 cm PepMap C18, Dionex) equilibrated in 96% solvent A (0.1% (v/v) formic acid in HPLC-grade water) and 4% solvent B (80% (v/v) ACN, 0.1% (v/v) formic acid in HPLC-grade water). Peptides were eluted from the precolumn to the analytical column and then to the mass spectrometer with a gradient from 4 to 57% solvent B for 50 min and 57−90% solvent B for 10 min, at a flow rate of 0.3 $\mu$L/min delivered by the Ultimate pump.

The separated peptides were then analyzed on-line using an ion-trap mass spectrometer. This mass spectrometer was used in the positive mode and spectra were acquired in a data dependent manner: MS scan range = 300−1500 $m/z$, maximum accumulation time = 200 ms, ICC target = 200 000. The top 5 most intense ions in the MS scan were selected for MS/MS in dynamic exclusion mode: mode = ultrascan, absolute threshold = 75 000, relative threshold = 1%, excluded after spectrum count = 1, exclusion duration = 0.3 min, averaged spectra = 5, ICC target = 200 000. Data extraction and database search were performed as described above.

**2.4. SDS-PAGE Prior to LC-MS/MS. 2.4.1. Protein Extraction and Quantification.** The bacterial pellet was resuspended in extraction buffer (150 mM Tris-HCl, pH 7.2/300 mM NaCl with a trace of Mini EDTA-free protease inhibitor cocktail) (Roche, Belgium). Cells were ruptured by three passages through a French Press (Thermo) at 500 psi.

After centrifugation at 13 200 rpm at 4 °C for 15 min, the supernatant was removed and the pellet was washed three times with the extraction buffer. After centrifugation, the pellet was resuspended in Laemmli buffer (2% SDS, 10% glycerol, 5% $\beta$-mercaptoethanol, 0.002% bromophenol blue and 0.125 M

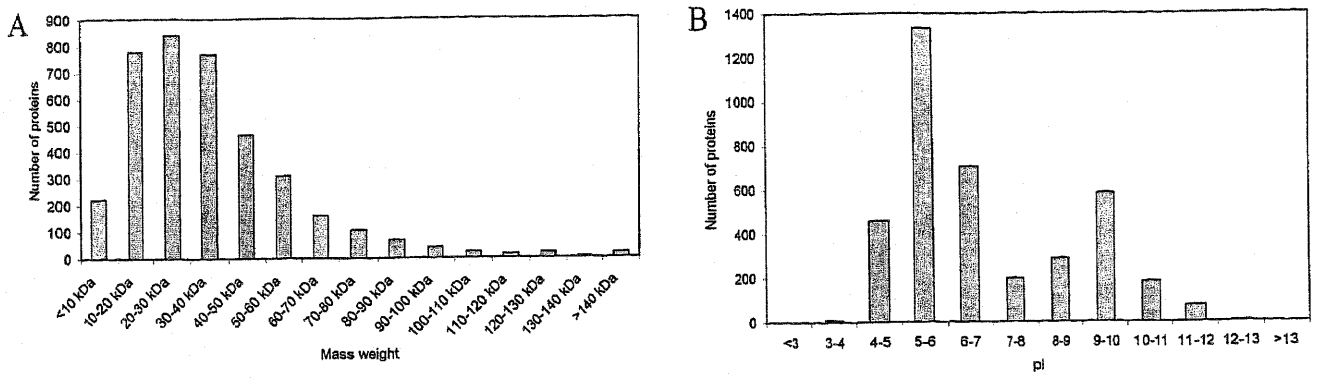



**Figure 2.** *In silico* predicted protein mass weight (MW) (A) and isoelectric point (p*I*) (B) of the *R. rubrum* S1 whole-proteome.

Tris-HCl, pH 6.8) and sonicated in water bath six times 1 min at room temperature. After 1 min incubation at 90 °C, the lysate was centrifuged at 14 000 rpm at room temperature for 15 min to pellet the cell debris. The membrane containing supernatant was loaded on a Precast polyacrylamide mini-gel 4–20% (Pierce) and was run at 150 V and 75 mA/gel. Staining of the gel was achieved using Imperial Protein Stain (Thermo) according to the manufacture's instructions. The corresponding gel lane containing proteins was cut in 29 pieces of 1 mm each.

**2.4.2. Trypsinolysis.** Enzymatic digestion was performed by the addition of 10 $\mu$L of modified sequencing grade trypsin (0.02 mg/mL) (Promega) in 25 mM $NH_4HCO_3$ to each gel piece. The samples were placed for 15 min at 4 °C and incubated overnight at 37 °C. The reaction was stopped with 1 $\mu$L of 5% (v/v) formic acid. At this point, the gel pieces were either stored at −20 °C or directly used for downstream analysis.

LC-MS/MS analysis was performed as described in the IPG-IEF paragraph.

**2.5. Protein Subcellular Localization Prediction.** The *R. rubrum* S1 genome comprises a chromosome (4.35 Mb; GenBank ID: CP000230) and a small plasmid (53.7 kb; GenBank ID: CP000231). Protein subcellular localization prediction for *R. rubrum* S1 was achieved using the Web-based tool, P-CLASSIFIER[17] built on top of multiple support vector machines and available at http://protein.bii.a-star.edu.sg/localization/gram-negative/index.html. Protein sequence information in FASTA format was retrieved from the Joint Genome Institute, Department of Energy (United States) (http://genome.jgi-psf.org/finished_microbes/rhoru/rhoru.home.html).

**2.6. Exponentially Modified Protein Abundance Index (emPAI).** The emPAI is a convenient and easily obtained index that can be used to produce protein expression data from any LC-MS/MS runs. This quantification index is based on the number of sequenced peptides per protein. It has been shown to be directly proportional to protein content and has been defined by Ishihama et al.[18]

This index was directly calculated by the Mascot research server but had to be normalized from one experiment to another to avoid bias inherent to data based acquired spectra already mentioned above. This was achieved by calculating the following percentage:[18]

$$\text{Normalized emPAI} = (\text{emPAI}_{\text{protein}} / \text{emPAI}_{\text{total}}) \times 100$$

where $\text{emPAI}_{\text{protein}}$ and $\text{emPAI}_{\text{total}}$ are, respectively, the emPAI for a given protein and the sum of the emPAI of all the identified proteins within a given experiment.

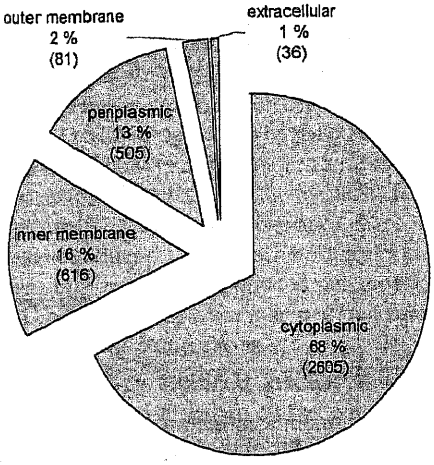



**Figure 3.** *In silico* protein localization prediction of *R. rubrum* whole-proteome (3829 candidate proteins).

## 3. Results and Discussion

**3.1. Whole-Proteome *in Silico* Prediction.** *In silico* prediction tools are powerful tools that permit to easily predict biological characteristics of proteins based on amino acid sequence. These data can be used to set up protocols and to assess the relevance of experimental results.[17,19]

To characterize *R. rubrum* S1 proteome, the protein mass weight (MW) and isoelectric point (p*I*) were predicted using the "Compute pI/Mw tool" from Expasy (http://www.expasy.ch/tools/pi_tool.html). *In silico* predicted protein MW showed that about 90% of *R. rubrum* S1 proteins are distributed from below 10 kDa to 100 kDa (Figure 2A). On the other hand, *in silico* predicted protein p*I* showed a classical bimodal distribution (Figure 2B), as previously observed for other bacterial and archaeal proteomes.[20] This distribution was highly similar to that corresponding to the *Synechocystis* sp. proteome,[20] another phototrophic bacterium.

The P-classifier Web-tool was used to predict the localization of the 3829 candidate proteins included in the *R. rubrum* S1 proteome (Figure 3). This analysis predicted 68% cytoplasmic, 16% inner membrane, 13% periplasmic, 2% outer membrane and 1% extracellular proteins. The amount of predicted membrane proteins was situated at the lower limit of the range (18–29%) reported by Kihara and Kanehisa[21] who investigated 11 bacterial complete genomes.

**3.2. 2D-LC MS/MS Approach.** With the use of the 2D-LC MS/MS approach and starting from 18 $\mu$g of *R. rubrum* digested proteins, 226 proteins including 37 hypothetical proteins were identified with a Mascot score of at least 50 (Supplementary

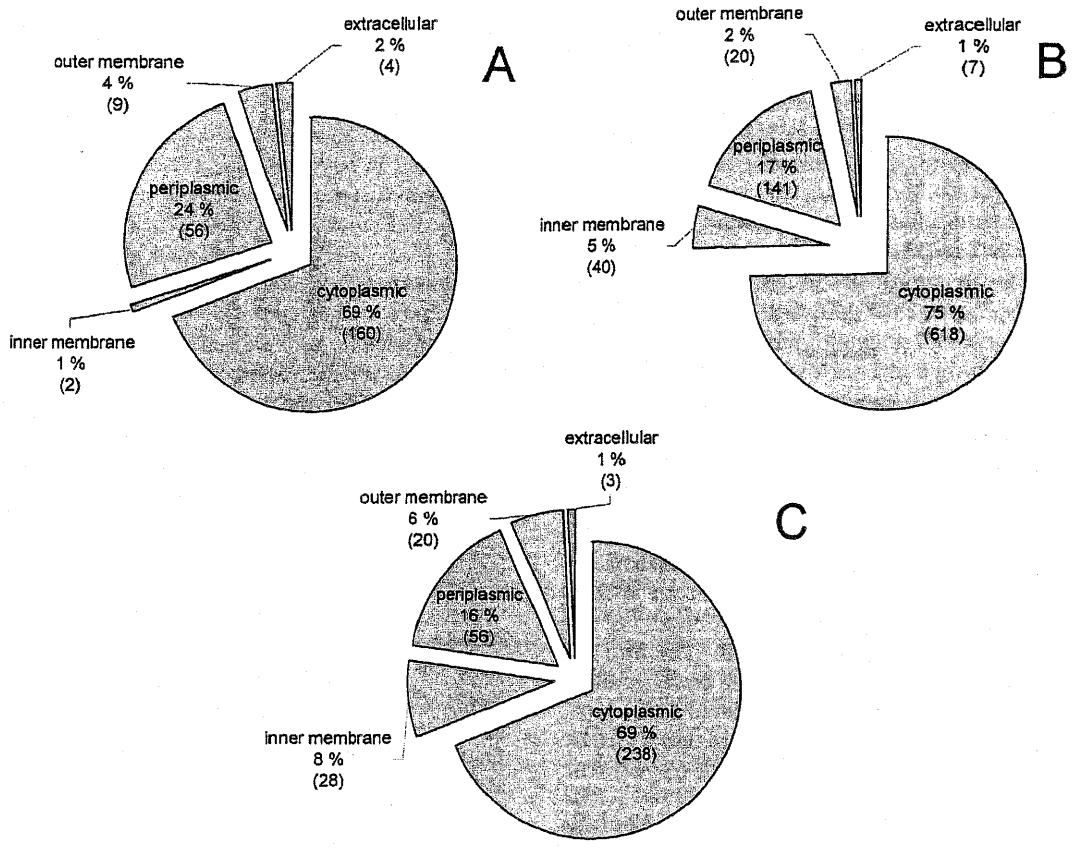



**Figure 4.** *In silico* protein localization prediction of the *R. rubrum* proteins identified using the 2D-LC (A) the IPG-IEF-LC (B) and the SDS-PAGE-LC (C) MS/MS approaches (see Material and Methods). Respectively, 226, 826, and 345 proteins were identified with a Mascot score above 50.

Table 1 in Supporting Information). The computed FDR was 0.34% estimated at the peptide level. Among the 226 proteins, 69% were predicted to be cytoplasmic, 1% inner membrane, 25% periplasmic, 4% outer membrane and 1% extracellular proteins (Figure 4A).

Our results show a fairly good coverage for cytosolic, periplasmic and extracellular proteins. However, this coverage was highly biased towards the identification of cytosolic proteins, presumably due to a better solubility of cytosolic proteins compared to membrane proteins, especially in our experimental conditions. Indeed, the protein extraction was realized, without detergent, in presence of 6 M guanidine chloride. Moreover, just before the trypsin treatment, the proteins were precipitated in presence of acetone and were resolubilized in 2 M urea. During these steps, it might occur that some hydrophobic proteins do not precipitate or do not resolubilize in 2 M urea.

**3.3. Peptide IPG-IEF-LC MS/MS Approach.** The extraction protocol (including a urea/thiourea extraction buffer) using twice 300 $\mu$g of *R. rubrum* digested proteins, resulted in the identification of 826 nonredundant proteins including 104 hypothetical proteins (Supplementary Tables 2 and 3 in Supporting Information). The distribution of the 826 proteins in regard to subcellular localization prediction showed 75% cytoplasmic, 5% inner membrane, 17% periplasmic, 2% outer membrane and 1% extracellular proteins (Figure 4B).

**3.4. SDS-PAGE-LC MS/MS Approach.** From the 29 analyzed SDS-PAGE fractions (Figure 1C), we could retrieve 345 nonredundant proteins including 63 hypothetical proteins identified by a Mascot score of at least 50 (Supplementary Table 4 in Supporting Information). The computed FDR was 0.82% estimated at the peptide level. Protein prediction model predicted 69% cytoplasmic, 8% inner membrane, 16% periplasmic, 6% outer membrane and 1% extracellular proteins (Figure 4C).

**3.5. Comparison between the Different Experimental Approaches.** Except for the number of identified proteins and protein localization comparisons, the two replicate injections of the IPG-IEF samples were treated separately in an effort to assess the reproducibility of the method. Besides, one must be aware that in our MudPIT approach peptide spectra were acquired in a data dependent manner (see Material and Methods). A peptide undetected during one run did not necessarily imply that the peptide was absent. Therefore, the word "specifically" used below should not be understood as exclusive.

**3.5.1. Absolute Number of Nonredundant Protein Identified.** Combining the results from all the approaches, we end up with the identification of 932 nonredundant *R. rubrum* proteins. When the 3 approaches mentioned above were compared in term of identified proteins, a relative high overlap between the techniques was observed and therefore only a few proteins remained specific to one particular approach (except for the IPG-IEF-LC approach obviously) (Figure 5). The 21 proteins specific for the 2D-LC approach were all predicted to be either cytoplasmic or periplasmic and included 12 hypothetical proteins (Table 1). On the other hand, of the 77 proteins specific to the SDS-PAGE membrane enriched samples, 21 were predicted membrane proteins (Table 2). In that purpose, we can mention that Rru_A1266 annotated as a photosystem assembly protein could be incorrect since expert annotation using the MaGe platform from GenoScope[22] revealed that

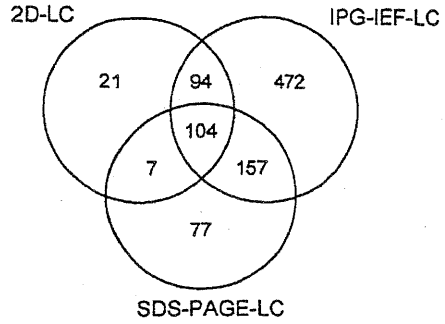



**Figure 5.** Venn diagram showing nonredundant proteins identified in the 2D-LC, the IPG-IEF-LC (pooled results from first and second replicas) and the SDS-PAGE-LC MS/MS approaches.

annotation has no experimental evidence and that the protein would be better annotated as 'conserved hypothetical protein'.

Pooling the data from the different approaches, we could identify 24% (i.e., 932 proteins) of the theoretical annotated proteome of *R. rubrum* cultivated in dark aerobic growth conditions which is comparable to recent studies using a similar mass spectrometer.[23−25] Moreover, our results further supported the observation that the number of proteins expressed in proteobacteria in a stable conditions is around 1000 and that this number remains similar independent of genome size.[26−32]

Therefore, our study constituted the first large-scale proteomic study of this bacterium since Selao et al.[13] identified 44 differentially expressed proteins of 600 protein spots using the 2D-PAGE approach.

**3.5.2. Mass Weight and Isoelectric Point.** In contrast with the other approaches, the overall trend of the molecular weight (MW) distribution (Figure 6A) for the proteins identified during the IEF experiments was very similar to the theoretical proteome, suggesting no real bias in our approach. By comparing MW profiles of the proteins identified in the 2D-LC and the SDS-PAGE-LC experiments, we could see that the 2D-LC and the SDS-PAGE-LC approaches were more favorable to a range of MW of 10−20 and 20−30 kDa, respectively.

Besides, we could see similar isoelectric point (p*I*) distribution for the proteins identified in the IPG-IEF-LC experiments (Figure 6B) compared to the theoretical distribution (Figure 2B), strengthening the hypothesis that the IPG-IEF-LC experiment correctly reflected the proteome coverage.

**3.5.3. Number of Peptides Per Proteins.** The SDS-PAGE experiment significantly increased the number of peptides identified per protein thereby increasing the confidence level of the protein identification. In contrast with the MW and the p*I* distributions, a notable difference appeared between the first and the second replica of the IPG-IEF-LC approach regarding the number of peptides per identified protein (Figure 7).

Indeed, double the amount of identified proteins with more than 10 peptides detected was observed within the second replica compared to the first replica. Moreover, the second analysis detected fewer proteins with only 1 or 2 peptides. This could be related to the preparation of the samples for the IPG-IEF-LC experiment. Indeed, for the first replica, trypsinolysis was stopped using the usual 5% formic acid. However, this leads to the precipitation of the peptides that could not be completely resolubilized using NaOH. Therefore, the inconsistency regarding number of peptides per proteins between the first and the second replica confirmed that many peptides were lost between the 2 injections.

**3.5.4. Exponentially Modified Protein Abundance Index (emPAI).** One way to gain insight on the quantity of a protein in a complex mixture is to calculate its emPAI, which basically represents the number of identified peptides divided by the number of theoretical tryptic peptides.[18]

Median raw emPAI values were, respectively: 0.56, 0.27, 0.39, and 0.43 for the 2D-LC, the IPG-IEF-LC first replica, the IPG-IEF-LC second replica and the SDS-PAGE-LC MS/MS approaches. EmPAI values confirmed the data from the number of peptides per proteins since the emPAI was lower for the IPG-

**Table 1.** The 21 *R. rubrum* S1H Proteins Identified in Dark Aerobic (DAE) Culture Conditions Using Specifically the 2D-LC MS/MS Approach[a]

| gene identifier | gene name | product name | Mascot score | SC (%) | no. peptides | MW (kDa) | localization prediction |
|---|---|---|---|---|---|---|---|
| Rru_A0089 | - | ATP-dependent DNA helicase RecQ | 133 | 1.11 | 14 | 68.89 | cytoplasmic |
| Rru_A0296 | - | hypothetical protein Rru_A0296 | 183 | 24.77 | 4 | 12.61 | cytoplasmic |
| Rru_A0422 | - | hypothetical protein Rru_A0422 | 217 | 23.44 | 4 | 13.81 | periplasmic |
| Rru_A0497 | - | ABC transporter component | 72 | 5.19 | 8 | 39.34 | cytoplasmic |
| Rru_A1033 | - | hypothetical protein Rru_A1033 | 263 | 13.69 | 1 | 18.62 | cytoplasmic |
| Rru_A1042 | - | hypothetical protein Rru_A1042 | 614 | 45.16 | 13 | 10.23 | cytoplasmic |
| Rru_A1072 | *rpmE* | 50S ribosomal protein L31 | 1561.98 | 30.56 | 26 | 8.11 | periplasmic |
| Rru_A1175 | - | CsbD-like | 884 | 44.62 | 29 | 7.61 | cytoplasmic |
| Rru_A1353 | - | hypothetical protein Rru_A1353 | 332 | 29.59 | 15 | 18.19 | periplasmic |
| Rru_A1354 | - | hypothetical protein Rru_A1354 | 388 | 14.66 | 3 | 12.81 | cytoplasmic |
| Rru_A1446 | - | hypothetical protein Rru_A1446 | 123 | 21.82 | 7 | 18.33 | cytoplasmic |
| Rru_A1665 | *rpmF* | 50S ribosomal protein L32 | 287 | 34.43 | 10 | 7.03 | cytoplasmic |
| Rru_A1867 | - | hypothetical protein Rru_A1867 | 75 | 27.67 | 3 | 16.58 | cytoplasmic |
| Rru_A2061 | - | hypothetical protein Rru_A2061 | 2708 | 49.51 | 51 | 21.44 | periplasmic |
| Rru_A2062 | - | hypothetical protein Rru_A2062 | 258 | 19.20 | 4 | 12.07 | periplasmic |
| Rru_A2208 | - | hypothetical protein Rru_A2208 | 174 | 33.72 | 2 | 9.59 | cytoplasmic |
| Rru_A2309 | - | NADH peroxidase | 51.78 | 4.15 | 2 | 58.87 | cytoplasmic |
| Rru_A2723 | - | rubrerythrin | 127 | 23.95 | 4 | 19.25 | cytoplasmic |
| Rru_A3286 | - | hypothetical protein Rru_A3286 | 252 | 26.17 | 7 | 16.24 | periplasmic |
| Rru_A3521 | - | 50S ribosomal protein L35P | 60 | 16.92 | 1 | 7.41 | periplasmic |
| Rru_B0047 | - | NAD-dependent epimerase/dehydratase | 96 | 1.45 | 10 | 38.71 | cytoplasmic |

[a] 'Rru_A' and 'Rru_B' refer to gene located, respectively, on the chromosome or on the plasmid. SC: sequence coverage. MW: mass weight.

**Table 2.** The 77 *R. rubrum* S1H Proteins Identified in Dark Aerobic (DAE) Culture Conditions Using Specifically the SDS-PAGE-LC MS/MS Approach[a]

| gene identifier | product name | Mascot score | SC (%) | no. peptides | MW (kDa) | localization prediction |
|---|---|---|---|---|---|---|
| Rru_A0020 | glutathione peroxidase | 76.09 | 16.07 | 2 | 18.51 | periplasmic |
| Rru_A0081 | chemotaxis sensory transducer | 72.05 | 8.01 | 4 | 39.79 | cytoplasmic |
| Rru_A0096 | secretion protein HlyD | 175.16 | 14.64 | 3 | 34.14 | cytoplasmic |
| Rru_A0122 | hypothetical protein Rru_A0122 | 215.24 | 38.33 | 9 | 12.73 | periplasmic |
| Rru_A0139 | MarR family transcriptional regulator | 70.79 | 11.97 | 1 | 12.96 | cytoplasmic |
| Rru_A0213 | hypothetical protein Rru_A0213 | 239.17 | 12.78 | 5 | 46.85 | periplasmic |
| Rru_A0231 | hypothetical protein Rru_A0231 | 62.77 | 13.37 | 1 | 21.70 | periplasmic |
| Rru_A0291 | hypothetical protein Rru_A0291 | 130.19 | 13.51 | 2 | 12.36 | inner membrane |
| Rru_A0333 | Ferritin and Dps | 102.64 | 13.29 | 1 | 17.65 | cytoplasmic |
| Rru_A0364 | ubiquinol-cytochrome c reductase, iron–sulfur subunit | 286.21 | 32.24 | 22 | 19.79 | periplasmic |
| Rru_A0392 | molybdopterin binding domain-containing protein | 579.47 | 41.18 | 9 | 26.57 | inner membrane |
| Rru_A0434 | organic solvent tolerance protein OstA-like | 56.82 | 4.85 | 2 | 94.76 | outer membrane |
| Rru_A0496 | cyclic nucleotide-binding domain-containing protein | 158.58 | 16.28 | 2 | 13.80 | cytoplasmic |
| Rru_A0568 | chemotaxis sensory transducer | 184.71 | 15.71 | 9 | 69.44 | cytoplasmic |
| Rru_A0576 | methyl-accepting chemotaxis sensory transducer | 107.07 | 3.49 | 5 | 71.53 | cytoplasmic |
| Rru_A0636 | hypothetical protein Rru_A0636 | 77.15 | 9.80 | 2 | 11.53 | cytoplasmic |
| Rru_A0637 | hypothetical protein Rru_A0637 | 520.34 | 14.84 | 11 | 12.85 | periplasmic |
| Rru_A0755 | chemotaxis sensory transducer | 1066.32 | 31.04 | 22 | 60.75 | inner membrane |
| Rru_A0782 | hypothetical protein Rru_A0782 | 100.50 | 9.82 | 1 | 12.48 | cytoplasmic |
| Rru_A0823 | cupin region | 73.70 | 5.65 | 2 | 34.05 | cytoplasmic |
| Rru_A1073 | hypothetical protein Rru_A1073 | 327.35 | 23.20 | 8 | 41.30 | outer membrane |
| Rru_A1074 | MotA/TolQ/ExbB proton channel family protein | 422.15 | 20.81 | 9 | 46.94 | inner membrane |
| Rru_A1091 | MotA/TolQ/ExbB proton channel | 344.53 | 8.77 | 5 | 33.08 | inner membrane |
| Rru_A1109 | chemotaxis sensory transducer | 123.21 | 4.95 | 5 | 59.74 | inner membrane |
| Rru_A1130 | nitrogen regulatory protein P–II | 56.54 | 9.82 | 1 | 12.28 | cytoplasmic |
| Rru_A1197 | chemotaxis sensory transducer | 315.78 | 22.53 | 11 | 72.94 | inner membrane |
| Rru_A1218 | phage shock protein A, PspA | 144.37 | 20.35 | 3 | 25.26 | cytoplasmic |
| Rru_A1266 | photosystem I assembly BtpA | 143.65 | 28.46 | 5 | 28.56 | inner membrane |
| Rru_A1293 | methyl-accepting chemotaxis sensory transducer | 831.34 | 36.12 | 22 | 60.06 | inner membrane |
| Rru_A1311 | MerR family transcriptional regulator | 178.32 | 9.09 | 3 | 15.12 | cytoplasmic |
| Rru_A1402 | CheW protein | 93.19 | 7.26 | 1 | 19.26 | cytoplasmic |
| Rru_A1403 | methyl-accepting chemotaxis sensory transducer | 501.09 | 17.58 | 15 | 67.09 | inner membrane |
| Rru_A1408 | chemotaxis sensory transducer | 159.19 | 5.63 | 6 | 60.04 | inner membrane |
| Rru_A1438 | sigma-24 (FecI) | 58.89 | 4.62 | 2 | 19.29 | cytoplasmic |
| Rru_A1535 | plasmid maintenance system antidote protein | 60.47 | 12.20 | 1 | 13.65 | periplasmic |
| Rru_A1554 | methyl-accepting chemotaxis sensory transducer | 154.50 | 13.74 | 5 | 52.27 | cytoplasmic |
| Rru_A1563 | NADH dehydrogenase subunit I | 143.73 | 11.11 | 3 | 19.09 | cytoplasmic |
| Rru_A1668 | integration host factor, alpha subunit | 207.35 | 33.33 | 5 | 11.44 | cytoplasmic |
| Rru_A1733 | hypothetical protein Rru_A1733 | 53.93 | 3.17 | 1 | 40.27 | periplasmic |
| Rru_A1843 | OmpA/MotB | 61.77 | 10.56 | 2 | 37.09 | periplasmic |
| Rru_A1888 | protein translocase subunit secG | 492.35 | 63.01 | 10 | 14.27 | periplasmic |
| Rru_A1985 | hypothetical protein Rru_A1985 | 262.56 | 41.26 | 6 | 15.59 | outer membrane |
| Rru_A2082 | hemerythrin HHE cation binding region | 85.83 | 6.22 | 2 | 23.81 | cytoplasmic |
| Rru_A2154 | biopolymer transport protein ExbD/TolR | 86.15 | 22.58 | 2 | 16.11 | cytoplasmic |
| Rru_A2218 | short chain dehydrogenase | 74.86 | 17.16 | 4 | 28.28 | cytoplasmic |
| Rru_A2375 | hypothetical protein Rru_A2375 | 239.29 | 37.21 | 5 | 9.19 | periplasmic |
| Rru_A2424 | gamma-glutamyltranspeptidase | 165.40 | 22.00 | 7 | 40.85 | periplasmic |
| Rru_A2438 | hypothetical protein Rru_A2438 | 72.92 | 7.95 | 2 | 15.93 | periplasmic |
| Rru_A2485 | Alpha,alpha-trehalose-phosphate synthase (UDP-forming) | 53.52 | 1.96 | 2 | 52.32 | cytoplasmic |
| Rru_A2546 | chemotaxis sensory transducer | 150.12 | 5.51 | 5 | 73.16 | inner membrane |
| Rru_A2605 | phosphoheptose isomerase | 88.00 | 15.96 | 2 | 19.76 | cytoplasmic |
| Rru_A2608 | hypothetical protein Rru_A2608 | 151.26 | 18.95 | 3 | 10.75 | cytoplasmic |
| Rru_A2760 | hypothetical protein Rru_A2760 | 394.39 | 28.00 | 5 | 7.68 | periplasmic |
| Rru_A2774 | chemotaxis sensory transducer | 1658.00 | 32.24 | 31 | 77.36 | inner membrane |
| Rru_A2839 | hypothetical protein Rru_A2839 | 53.73 | 4.40 | 2 | 27.14 | cytoplasmic |
| Rru_A2894 | ABC transporter component | 108.83 | 7.00 | 1 | 32.03 | inner membrane |
| Rru_A2910 | hypothetical protein Rru_A2910 | 201.16 | 26.81 | 6 | 33.10 | outer membrane |
| Rru_A2911 | ABC-type uncharacterized transport system auxiliary component-like | 350.28 | 36.45 | 8 | 21.95 | cytoplasmic |
| Rru_A2968 | hypothetical protein Rru_A2968 | 116.25 | 13.51 | 1 | 20.08 | periplasmic |
| Rru_A3038 | hypothetical protein Rru_A3038 | 229.73 | 20.45 | 3 | 14.43 | cytoplasmic |
| Rru_A3117 | polysaccharide export protein | 71.56 | 6.70 | 1 | 22.51 | periplasmic |
| Rru_A3153 | virulence factor protein | 89.79 | 3.77 | 2 | 99.35 | cytoplasmic |
| Rru_A3154 | hypothetical protein Rru_A3154 | 265.63 | 10.37 | 8 | 111.29 | cytoplasmic |
| Rru_A3155 | hypothetical protein Rru_A3155 | 127.37 | 14.62 | 5 | 55.93 | periplasmic |

**Table 2.** Continued

| gene identifier | product name | Mascot score | SC (%) | no. peptides | MW (kDa) | localization prediction |
|---|---|---|---|---|---|---|
| Rru_A3193 | glutathione *S*-transferase-like protein | 67.06 | 8.56 | 1 | 24.62 | inner membrane |
| Rru_A3253 | hypothetical protein Rru_A3253 | 187.67 | 18.56 | 3 | 38.72 | cytoplasmic |
| Rru_A3299 | glycoside hydrolase family protein | 74.27 | 3.21 | 1 | 76.16 | cytoplasmic |
| Rru_A3360 | electron transport protein SCO1/SenC | 57.47 | 10.00 | 1 | 22.41 | periplasmic |
| Rru_A3409 | hypothetical protein Rru_A3409 | 82.16 | 8.22 | 2 | 31.63 | cytoplasmic |
| Rru_A3414 | ABC transporter component | 86.87 | 18.82 | 2 | 27.00 | cytoplasmic |
| Rru_A3437 | aminoglycoside phosphotransferase | 59.39 | 9.44 | 1 | 39.31 | cytoplasmic |
| Rru_A3474 | bacterioferritin | 77.01 | 12.42 | 2 | 18.57 | cytoplasmic |
| Rru_A3593 | phosphoribosyl-ATP pyrophosphatase | 70.83 | 22.22 | 3 | 12.42 | cytoplasmic |
| Rru_A3649 | 2′,5′ RNA ligase | 89.37 | 9.05 | 1 | 21.69 | cytoplasmic |
| Rru_A3665 | chemotaxis sensory transducer | 1142.60 | 30.51 | 18 | 59.61 | inner membrane |
| Rru_A3699 | acriflavin resistance protein | 221.41 | 8.65 | 6 | 109.79 | inner membrane |
| Rru_B0016 | Phage integrase | 56.47 | 3.38 | 3 | 23.76 | cytoplasmic |

[a] 'Rru_A' and 'Rru_B' refer to gene located, respectively, on the chromosome or on the plasmid. SC: sequence coverage. MW: mass weight.

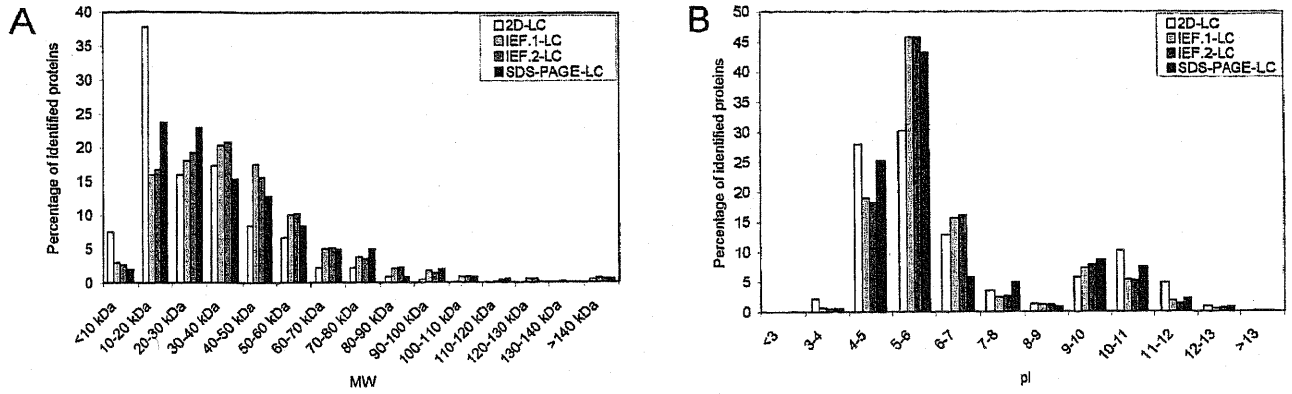



**Figure 6.** Distribution of *R. rubrum* S1H identified proteins from the 2D-LC, the IPG-IEF-LC first replica (IEF.1-LC) and the second replica (IEF.2-LC), and the SDS-PAGE-LC MS/MS based on mass weight (MW) (A) and isoelectric point (p*I*) (B).

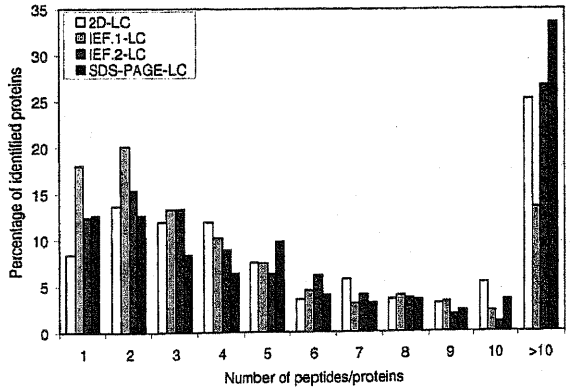



**Figure 7.** Distribution of *R. rubrum S1H* identified proteins from the 2D-LC, the IPG-IEF-LC first replica (IEF.1-LC) and the second replica (IEF.2-LC), and the SDS-PAGE-LC MS/MS approaches based on the number of peptides detected per protein.

IEF-LC first replica where more proteins identified with 1 or 2 peptides were found.

Besides, the most abundant proteins such as the chaperonins CsbD-like (Rru_A1175) and GroEL (Rru_A0587), the F0F1 ATP synthase subunit beta (Rru_A1226), the cytochrome c (Rru_A1020) and a ribosomal protein (Rru_A2684) were present in the cytoplasmic and periplasmic fractions. They represented, respectively, about 7.5%, 5.5%, 6.0%, 2.8% and 2.8 % of their respective total emPAI (Figure 8). We could also mention a hypothetical protein (Rru_A2202) and a stress protein (Rru_A0894) as the most abundant proteins in the inner membrane and outer membrane fractions, respectively. The expert annotation using the MaGe platform[22] revealed Rru_A2202 to be the protease co-factor HflC. It could be surprising to find the OmpA protein (Rru_A3328) predicted to be periplasmic, but the localization score prediction (26.7% cytoplasmic, 13.3% inner membrane, 20.0% outer membrane and 40.0% periplasmic) indicated some membrane domains. On the other hand, protein annotated with the term metabolism like the thiamine biosynthesis (Rru_A2008), the pyruvate kinase (Rru_A2465) and the NADH peroxidase (Rru_A2309) related proteins represented less than 0.02% of the total emPAI and therefore could be seen as low abundant.

Considering emPAI related COG distribution from all data, the 'Function unknown' (S) class contained the most abundant protein whatever the technique employed, namely, the chaperonin CsbD-like (Figure 9). The second most abundant protein was retrieved in the 'Energy and metabolism' (C) class, namely, the F0F1 ATP synthase subunit beta (Rru_A1226), while the third one was the GroEL chaperone (Rru_A0587) from the 'Posttranslational modification, protein turnover, chaperones' (O) class. The top 5 was rounded out with Rru_A2690 and Rru_A2702, both being elongation factors from the 'Translation, ribosomal structure and biogenesis' (J) class. In that purpose, the classes showing the most abundant proteins logically appeared to be the O and the J classes containing, respectively, the chaperones and the ribosomal proteins. We could also mention the relative high abundance of Rru_A3283, a granule associated protein also named phasin most commonly associated with polyhydroxybutyrate inclusions.

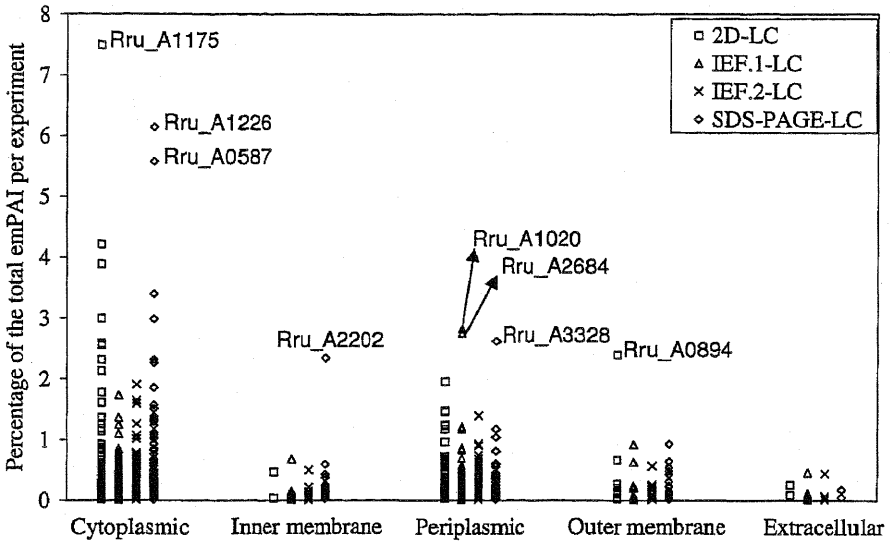



**Figure 8.** emPAI distribution of *R. rubrum* S1H identified proteins from the 2D-LC, the IPG-IEF-LC first replica (IEF.1-LC) and the second replica (IEF.2-LC), and the SDS-PAGE-LC MS/MS experiments in regard to protein localization prediction.

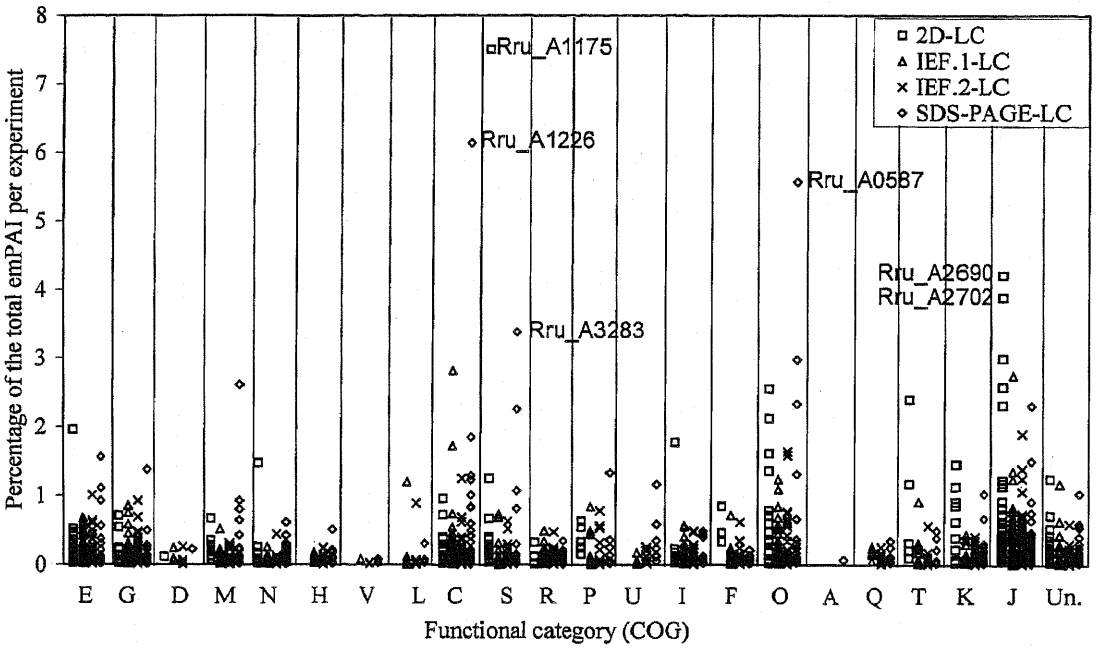



**Figure 9.** emPAI distribution of *R. rubrum* S1H identified proteins from the 2D-LC, the IPG-IEF-LC first replica (IEF.1-LC) and the second replica (IEF.2-LC), and the SDS-PAGE-LC MS/MS experiments in regard to the COG classification. COG nomenclature: E, Amino acid transport and metabolism; G, Carbohydrate transport and metabolism; D, Cell division and chromosome partitioning; M, Cell envelope biogenesis, outer membrane; N, Cell motility and secretion; H, Coenzyme metabolism; V, Defense mechanisms; L, DNA replication, recombination and repair; C, Energy production and conversion; S, Function unknown; R, General function prediction only; P, Inorganic ion transport and metabolism; U, Intracellular trafficking, secretion, and vesicular transport; I, Lipid metabolism; F, Nucleotide transport and metabolism; O, Post-translational modification, protein turnover, chaperones; Q, Secondary metabolites biosynthesis, transport and catabolism; T, Signal transduction mechanisms; K, Transcription; J, Translation, ribosomal structure and biogenesis; Un, Unclassified.

Finally, considering the 'Cell envelope biogenesis, outer membrane' (M) class, the SDS-PAGE-LC MS/MS approach seemed to be the most appropriate approach to study membrane proteins (Figure 9).

The emPAI approach permitted to estimate the absolute protein amount of *R. rubrum* S1H grown in dark aerobic culture conditions and put forward common highly abundant proteins like those involved in protein synthesis and energy metabolism.[24,25] On the other hand, our study revealed the high abundance of chaperone proteins already mentioned for *Methylobacterium extorquens* AM1, another α-proteobacterium.[33,34] Besides, the IPG-IEF-LC approach appeared to be more sensitive than the 2D-LC approach, as it was able to detect the low level constitutive expression of the antenna and photoreaction center proteins (Rru_A0617 and Rru_A2974) expressed in chemoheterotrophic culture conditions, as demonstrated by Bérard et al.[35] (Supplementary Tables 2 and 3 in Supporting Information).

Because the emPAI index is easily calculated from the information output of a database search engine like Mascot, this information should be kept in the reporting of all proteomic results. Hence, the emPAI approach could be used to

**Table 3.** *Rhodospirillaceae* Group Specific Proteins Detected during This Study[a]

| gene number | product name | Mascot score | SC (%) | no. peptides | MW (kDa) |
|---|---|---|---|---|---|
| Rru_A1756 | hypothetical protein Rru_A1756 | 118.87 | 9.35 | 2 | 15.78 |
| Rru_A2112 | hypothetical protein Rru_A2112 | 197.76 | 20.68 | 3 | 26.22 |
| Rru_A2510 | hypothetical protein Rru_A2510 | 191.62 | 31.52 | 3 | 20.58 |
| Rru_A3662 | hypothetical protein Rru_A3662 | 351.86 | 29.41 | 5 | 12.62 |
| Rru_A3739 | hypothetical protein Rru_A3739 | 165.96 | 14.54 | 4 | 58.03 |

[a] SC: sequence coverage. MW: mass weight.

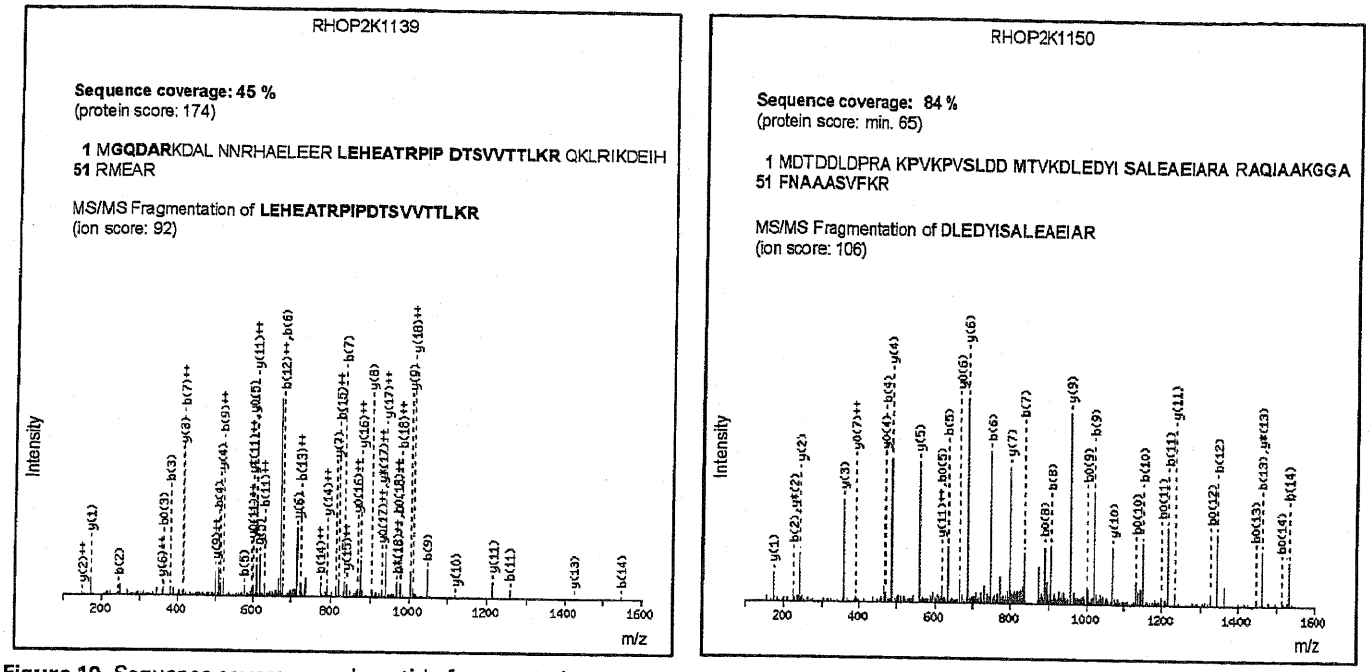



**Figure 10.** Sequence coverage and peptide fragmentation spectra of protein identified from the MaGe CDS models. Matched peptides are shown in bold in the sequence coverage. For RHOP2K1150, the 84% protein sequence coverage resulted from 2 analyses.

quantify protein abundance in cells from various growth conditions.[18] However, one must be aware that this approach is biased by the prefractionation method that could change the actual protein concentration due to sample preparation and peptides ionization. Moreover, Hendrickson et al.[36] mentioned that the emPAI approach is not as sensitive as stable isotope labeling for detecting small changes (i.e., in the order of 2-fold) in protein abundance.

**3.6. Identification of Proteins Specific for the *Rhodospirillaceae* Family.** By means of phylogenomics, Gupta and Mok[37] have described signature proteins for the α-proteobacteria and its main groups including the Rhodospirillales group. Within the latter group, they found 14 proteins unique for the *Rhodospirillaceae* family (*Rhodospirillum* and *Magnetospirillum*). During our study, we could identify 5 of these *Rhodospirillaceae* specific proteins: Rru_A1756, Rru_A2112, Rru_A2510, Rru_A3662 and Rru_A3739 (Table 3). All these 5 proteins were identified thanks to the IPG-IEF-LC MS/MS approach (Supplementary Tables 2 and 3 in Supporting Information) while Rru_A2112 and Rru_A2510 were also identified after SDS-PAGE-LC MS/MS (Supplementary Table 4). Beside, Rru_A3662 and Rru_A3739 were also detected with the 2D-LC MS/MS approach (Supplementary Table 1).

Concerning their abundance based on emPAI measurement, only Rru_A1756 (0.48) and Rru_A3662 (0.61) were above the emPAI median value for the IPG-IEF-LC first replica (0.39). While Rru_A2510 (0.35) was closed to the median value; Rru_A3739 (0.25) and especially Rru_A2112 (0.12) were below that value.

These 5 proteins were previously all annotated as 'hypothetical protein'. From now, they could be called 'conserved protein of unknown functions' since homologues in both the *Rhodospirillum* and *Magnetospirillum* genus have been found.[37] Only a putative secreted fate for Rru_A2112 and Rru_A3739 proteins could be deduced from the presence of predicted signal peptide cleavage sites in the amino acid sequences (SignalP 3.0 server: http://www.cbs.dtu.dk/services/SignalP/).

**3.7. Identification of New *R. rubrum* S1 Proteins.** Magnifying Genomes (MaGe) is a microbial genome annotation system based on a relational database containing information on bacterial genomes, as well as a Web interface to achieve genome annotation projects.[22] *R. rubrum* S1 is part of the MaGe 'MagnetoScope' project (https://www.genoscope.cns.fr/agc/mage/wwwpkgdb/MageHome/index.php?webpage=mage). The MaGe system used a CoDing sequence (CDS) model algorithm which is different from NCBI and allowed them to identify 258 extra CDS (MaGe Web site). The present study permitted to confirm the translation of two of these 'newly predicted CDS', namely, RHOP2K1139 and RHOP2K1150 both detected in the 2D-LC experiment, while RHOP2K1139 was also detected in the IPG-IEF-LC MS/MS approach. Reliable sequence coverage and fragmentation spectra lead to a Mascot score above 50 (Figure 10).

The emPAI value for RHOP2K1139 was 1.39 which was almost triple the median in the 2D-LC experiment (0.56). Beside, the RHOP2K1150 emPAI value was exactly equal to the median for the same experiment. The relative abundance of RHOP2K1139 was even more pronounced in the IPG-IEF-LC

experiments where the same emPAI value (i.e., 1.39) was found for the first replica (median = 0.27) and second replica (median = 0.39).

No function has yet been found for these two proteins, but their corresponding genes could also be detected in related bacteria like *Magnetospirillum magneticum* AMB-1 and *Rhodobacter sphaeroides* 2.4.1 (MaGe Web iste) where they were annotated as 'conserved hypothetical protein'.

Eventually, all the MS and MS/MS data acquired so far could be used in the future to look for new Open Reading Frames, namely, the proteogenomic approach, a cost-effective means to add value to genome annotation.[29,38,39]

## 4. Conclusions

We presented here the first shotgun proteomic analysis of the life-support α-proteobacterium *R. rubrum* S1H cultivated in standard chemoheterotrophic culture conditions. Using a combination of gel-free and gel-based peptides fractionation methods prior to LC-MS/MS analysis, we could identify 932 unique proteins covering 24% of *R. rubrum* theoretical annotated proteome including 5 family signature proteins and 2 'newly predicted CDS' from the MaGe gene annotation platform. Abundance data were obtained using the emPAI approach and revealed common highly abundant proteins like those involved in protein synthesis and energy metabolism. On the other hand, our study revealed the high abundance of chaperone proteins already mentioned for the α-proteobacterium *M. extorquens* AM1 under succinate growth.

This work paves the way for further shotgun proteomic investigations on *R. rubrum* S1H, for example, in photoheterotrophic culture conditions, as well as on the other MELiSSA strains, with the final goal of developing a reliable and a sustainable life support system for extended manned space missions.

**Acknowledgment.** We thank the European Space Agency (ESA-PRODEX) and the Belgian Federal Office for Science, Technology and Culture (DWTC/SSTC) for the financial support (PRODEX agreements). In addition, this research was partly funded by the Belgian Nuclear Research Center (SCK•CEN) through the PhD AWM grant of F. Mastroleo. R. Wattiez is a Research Associate of FNRS.

**Supporting Information Available:** The complete lists of proteins identified using the 2D-LC, the IEF-LC (first and second replicas) and the SDS-PAGE-LC MS/MS approaches. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Mergeay, M.; Verstraete, W.; Dubertret, G.; Lefort-Tran, M.; Chipaux, C.; Binot R. A. 'MELiSSA' —A Micro-Organisms-Based Model for 'CELSS' Development. *Proceedings at the 3rd European Symposium on Space Thermal Control & Life Support Systems*; Noordwijk: The Netherlands, 1988; pp 65–68.

(2) Hendrickx, L.; De Wever, H.; Hermans, V.; Mastroleo, F.; Morin, N.; Wilmotte, A.; Janssen, P.; Mergeay, M. Microbial ecology of the closed artificial ecosystem MELiSSA (Micro-Ecological Life Support System Alternative): Reinventing and compartimentalizing the Earth's food and oxygen regeneration system for long-haul space exploration missions. *Res. Microbiol.* **2006**, *157*, 77–86.

(3) Farges, B.; Poughon, L.; Creuly, C.; Cornet, J. F.; Dussap, C. G.; Lasseur, C. Dynamic aspects and controllability of the MELiSSA project: a bioregenerative system to provide life support in space. *Appl. Biochem. Biotechnol.* **2008**, *151*, 686–699.

(4) Hendrickx, L.; Mergeay, M. From the deep sea to the stars: human life support through minimal communities. *Curr. Opin. Microbiol.* **2007**, *10*, 231–237.

(5) Favier-Teodorescu, L.; Cornet, J. F.; Dussap, C. G. Modelling continuous cultures of *Rhodospirillum rubrum* in a photobioreactor under light limited conditions. *Biotechnol. Lett.* **2003**, *25*, 359–364.

(6) Montràs, A.; Pycke, B.; Boon, N.; Gòdia, F.; Mergeay, M.; Hendrickx, L.; Pérez, J. Distribution of *Nitrosomonas europaea* and *Nitrobacter winogradskyi* in an autotrophic nitrifying biofilm reactor as depicted by molecular analyses and mathematical modelling. *Water Res.* **2008**, *42*, 1700–1714.

(7) Rossi, N.; Derouiniot-Chaplain, M.; Jaouen, P.; Legentilhomme, P.; Petit, I. *Arthrospira platensis* harvesting with membranes: fouling phenomenon with limiting and critical flux. *Bioresour. Technol.* **2008**, *99*, 6162–6167.

(8) Christophe, G.; Guiavarch, E.; Creuly, C.; Dussap, C. G. Growth monitoring of *Fibrobacter succinogenes* by pressure measurement. *Bioprocess Biosyst. Eng.* **2009**, *32*, 123–128.

(9) Cañas, B; Piñeiro, C; Calvo, E.; López-Ferrer, D.; Gallardo, J. M. Trends in sample preparation for classical and second generation proteomics. *J. Chromatogr., A.* **2007**, *1153*, 235–258.

(10) Cargile, B. J.; Bundy, J. L.; Freeman, T. W.; Stephenson, J. L., Jr. Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J. Proteome Res.* **2004**, *3*, 112–119.

(11) Cargile, B. J.; Talley, D. L.; Stephenson, J. L., Jr. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis* **2004**, *25*, 936–945.

(12) Gan, C. S.; Reardon, K. F.; Wright, P. C. Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of *Synechocystis* sp. PCC 6803. *Proteomics* **2005**, *5*, 2468–2478.

(13) Selao, T. T.; Nordlund, S.; Norén, A. Comparative proteomic studies in *Rhodospirillum rubrum* grown under different nitrogen conditions. *J. Proteome Res.* **2008**, *7*, 3267–3275.

(14) Sistrom, W. R. A requirement for sodium in the growth medium of *Rhodopseudomonas sphaeroides*. *J. Gen. Microbiol.* **1960**, *22*, 77–85.

(15) Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005**, *2*, 667–675.

(16) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.

(17) Wang, J.; Sung, W. K.; Krishnan, A.; Li, K. B. Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines. *BMC Bioinf.* **2005**, *6*, 174.

(18) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265–1272.

(19) Bjellqvist, B.; Hughes, G. J.; Pasquali, C.; Paquet, N.; Ravier, F.; Sanchez, J.-C.; Frutiger, S.; Hochstrasser, D. F. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **1993**, *14*, 1023–1031.

(20) Schwartz, R.; Ting, C. S.; King, J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res.* **2001**, *11*, 703–709.

(21) Kihara, D.; Kanehisa, M. Tandem clusters of membrane proteins in complete genome sequences. *Genome Res.* **2000**, *10*, 731–743.

(22) Vallenet, D.; Labarre, L.; Rouy, Z.; Barbe, V.; Bocs, S.; Cruveiller, S.; Lajus, A.; Pascal, G.; Scarpelli, C.; Médigue, C. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* **2006**, *34*, 53–65.

(23) Pan, C.; Oda, Y.; Lankford, P. K.; Zhang, B.; Samatova, N. F.; Pelletier, D. A.; Harwood, C. S.; Hettich, R. L. Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Mol. Cell. Proteomics* **2008**, *7*, 938–948.

(24) Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F. U.; Kerner, M. J.; Frishman, D. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **2008**, *9*, 102.

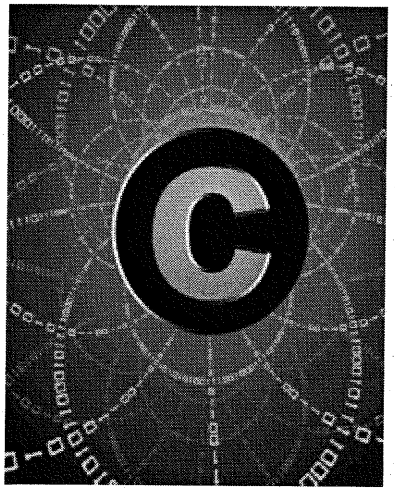
(25) Assiddiq, B. F.; Snijders, A. P.; Chong, P. K.; Wright, P. C.; Dickman, M. J. Identification and characterization of *Sulfolobus solfataricus* P2 proteome using multidimensional liquid phase protein separations. *J. Proteome Res.* **2008**, *7*, 2253–2261.

(26) Rohmer, L.; Guina, T.; Chen, J.; Gallis, B.; Taylor, G. K.; Shaffer, S. A.; Miller, S. I.; Brittnacher, M. J.; Goodlett, D. R. Determination

and comparison of the *Francisella tularensis* subsp.novicida U112 proteome to other bacterial proteomes. *J. Proteome Res.* **2008**, *7*, 2016–2024.
(27) Kolker, E.; Purvine, S.; Galperin, M. Y.; Stolyar, S.; Goodlett, D. R.; Nesvizhskii, A. I.; Keller, A.; Xie, T.; Eng, J. K.; Yi, E.; Hood, L.; Picone, A. F.; Cherny, T.; Tjaden, B. C.; Siegel, A. F.; Reilly, T. J.; Makarova, K. S.; Palsson, B. O.; Smith, A. L. Initial proteome analysis of model microorganism *Haemophilus influenzae* strain Rd KW20. *J. Bacteriol.* **2003**, *185*, 4593–4602.
(28) Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E., Jr.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.; Soupene, E.; Hunt, D. F. Toward a protein profile of Escherichia coli: comparison to its transcription profile. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9232–9237.
(29) Jaffe, J. D.; Berg, H. C.; Church, G. M. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **2004**, *4*, 59–77.
(30) Elias, D. A.; Monroe, M. E.; Smith, R. D.; Fredrickson, J. K.; Lipton, M. S. Confirmation of the expression of a large set of conserved hypothetical proteins in *Shewanella oneidensis* MR-1. *J. Microbiol. Methods* **2006**, *66*, 223–233.
(31) VerBerkmoes, N. C.; Shah, M. B.; Lankford, P. K.; Pelletier, D. A.; Strader, M. B.; Tabb, D. L.; McDonald, W. H.; Barton, J. W.; Hurst, G. B.; Hauser, L.; Davison, B. H.; Beatty, J. T.; Harwood, C. S.; Tabita, F. R.; Hettich, R. L.; Larimer, F. W. Determination and comparison of the baseline proteomes of the versatile microbe *Rhodopseudomonas palustris* under its major metabolic states. *J. Proteome Res.* **2006**, *5*, 287–298.
(32) Zhang, W.; Gritsenko, M. A.; Moore, R. J.; Culley, D. E.; Nie, L.; Petritis, K.; Strittmatter, E. F.; Camp, D. G., II; Smith, R. D.; Brockman, F. J. A proteomic view of *Desulfovibrio vulgaris* metabolism as determined by liquid chromatography coupled with tandem mass spectrometry. *Proteomics* **2006**, *6*, 4286–4299.
(33) Laukel, M.; Rossignol, M.; Borderies, G.; Völker, U.; Vorholt, J. A. Comparison of the proteome of *Methylobacterium extorquens* AM1 grown under methylotrophic and nonmethylotrophic conditions. *Proteomics* **2004**, *4*, 1247–1264.
(34) Bosch, G.; Skovran, E.; Xia, Q.; Wang, T.; Taub, F.; Miller, J. A.; Lidstrom, M. E.; Hackett, M. Comprehensive proteomics of *Methylobacterium extorquens* AM1 metabolism under single carbon and nonmethylotrophic conditions. *Proteomics* **2008**, *8*, 3494–3505.
(35) Bérard, J.; Bélanger, G.; Gingras, G. Mapping of the *puh* messenger RNAs from *Rhodospirillum rubrum*. Evidence for tandem promoters. *J. Biol. Chem.* **1989**, *18*, 10897–10903.
(36) Hendrickson, E. L.; Xia, Q.; Wang, T.; Leigh, J. A.; Hackett, M. Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *Analyst* **2006**, *131*, 1335–1341.
(37) Gupta, R. S.; Mok, A. Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. *BMC Microbiol.* **2007**, *7*, 106.
(38) Fermin, D.; Allen, B. B.; Blackwell, T. W.; Menon, R.; Adamski, M.; Xu, Y.; Ulintz, P.; Omenn, G. S.; States, D. J. Novel gene and gene model detection using a whole genome open reading frame analysis in proteomics. *Genome Biol.* **2006**, *7*, R35.
(39) Ansong, C.; Purvine, S. O.; Adkins, J. N.; Lipton, M. S.; Smith, R. D. Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Briefings Funct. Genomic Proteomic* **2008**, *7*, 50–62.

PR900007D

# subito copyright regulations



Copies of articles ordered through subito and utilized by the users are subject to copyright regulations. By registering with subito, the user commits to observing these regulations, most notably that the copies are for personal use only and not to be disclosed to third parties.

Should delivery be made by e-mail or FTP the copy may only be printed once, and the file must be permanently deleted afterwards.