RESEARCH ARTICLE

Protein expression profile of *Gluconacetobacter diazotrophicus* PAL5, a sugarcane endophytic plant growth-promoting bacterium

Leticia M. S. Lery[1, 2], Ana Coelho[1, 3], Wanda M. A. von Kruger[1, 2], Mayla S. M. Gonçalves[1, 3], Marise F. Santos[1, 4], Richard H. Valente[1, 5], Eidy O. Santos[1, 3], Surza L. G. Rocha[1, 5], Jonas Perales[1, 5], Gilberto B. Domont[1, 4], Katia R. S. Teixeira[1, 6] and Paulo M. Bisch[1, 2]

[1] Rio de Janeiro Proteomics Network, Rio de Janeiro, Brazil
[2] Unidade Multidisciplinar de Genômica, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[3] Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[4] Laboratório de Química de Proteínas, Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[5] Laboratório de Toxinologia, Departamento de Fisiologia e Farmacodinâmica- Instituto Oswaldo Cruz- Fundação Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil
[6] Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Seropédica, Brazil

This is the first broad proteomic description of *Gluconacetobacter diazotrophicus*, an endophytic bacterium, responsible for the major fraction of the atmospheric nitrogen fixed in sugarcane in tropical regions. Proteomic coverage of *G. diazotrophicus* PAL5 was obtained by two independent approaches: 2-DE followed by MALDI-TOF or TOF-TOF MS and 1-DE followed by chromatography in a C18 column online coupled to an ESI-Q-TOF or ESI-IT mass spectrometer. The 583 identified proteins were sorted into functional categories and used to describe potential metabolic pathways for nucleotides, amino acids, carbohydrates, lipids, cofactors and energy production, according to the Enzyme Commission of Enzyme Nomenclature (EC) and Kyoto Encyclopedia of genes and genomes (KEGG) databases. The identification of such proteins and their possible insertion in conserved biochemical routes will allow comparisons between *G. diazotrophicus* and other bacterial species. Furthermore, the 88 proteins classified as conserved unknown or unknown constitute a potential target for functional genomic studies, aiming at the understanding of protein function and regulation of gene expression. The knowledge of metabolic fundamentals and coordination of these actions are crucial for the rational, safe and sustainable interference on crops. The entire dataset, including peptide sequence information, is available as Supporting Information and is the major contribution of this work.

Received: September 25, 2007
Revised: December 18, 2007
Accepted: December 19, 2007

Keywords:
Bacterial metabolism / *Gluconacetobacter diazotrophicus* / Sugarcane / Sustainable agriculture

Correspondence: Leticia Miranda Santos Lery, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Av. Brigadeiro Tropowski, Centro de Ciencias da Saúde - Bloco G Lab.G1-043/Ilha do Fundão, Rio de Janeiro, Rio de Janeiro, 21949-900 Brazil
E-mail: leticialery@superig.com.br, llery@biof.ufrj.br
Fax: +55-21-22808193

Abbreviations: ADH, alcohol dehydrogenase; **ALDH,** aldehyde dehydrogenase; **BNF,** biological nitrogen fixation; **EC,** enzyme commission number; **NAD,** nicotinamide adenine dinucleotide; **PPP,** pentose phosphate pathway; **PQQ,** pyrroloquinoline quinone; **TCA,** citric acid cycle

1 Introduction

One of the growth-limiting factors for any organism is nitrogen. Although the most abundant compound in atmosphere, approximately 79%, this large reservoir of free nitrogen is not available for many organisms, including plants [1]. Furthermore, crop production in tropical soils, often assumed to be highly weathered and thus nutrient-depleted, is mainly based on the increasing use of chemical fertilizers. Since nitrogen is the main yield-limiting factor, as much as 30% of the total fertilizers used are nitrogenous chemicals [2]. Apart from costs, fertilizers affect the balance of the global nitrogen cycle, pollute groundwater, and increase the risk of chemical spills and the amount of atmospheric nitrous oxide, a potent "greenhouse" gas [3]. Consequently, decreasing dependence on chemicals input for agricultural production is a need.

Among the environmental-friendly technologies available, great emphasis has been given to the biological nitrogen fixation (BNF) process. The enzymatic reduction of N_2 to ammonia, unique to Bacteria and Archaea [4, 5], makes nitrogen available to living organisms, either directly, by association with N_2-fixing microorganisms (diazotrophs), or indirectly, by conversion of organic nitrogen into NO_3^- and NH_4^+, thus making nitrogen in the soil ready for plant uptake [6].

Besides *Rhizobium* species, which establish a nitrogen-fixing symbiosis in root nodules of legume plants, there is also a major subset of soil free-living bacteria, the endophytic diazotrophs, which fix nitrogen when in association with plant root surfaces or within host tissues [7, 8]. Living as endophytes, the bacteria face less competition for nutrients, interact more intimately with the plant and are protected from changes in the environment [9, 10].

Microbiological survey on cultures of Brazilian sugarcane tissues showed greater diversity of diazotrophic bacteria than in varieties grown in countries where substantial nitrogen input is applied to the soil [11, 12]. Thus, the low-cost high-yield production of sugarcane in Brazil is a reflection of the vast extensions of cropland, the tropical sunny weather plus the nitrogen-fixing activities of these endophytes. This is an example of the optimization of BNF applied to crop production, without the adverse environmental effects of fertilizers [11–13].

Among the endophytic diazotrophic bacteria isolated from sugarcane plants, *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*) was the first described [14]. It was isolated in 1988 from sugarcane juice-rich medium [15] and later characterized as an acid-tolerant N_2-fixing, aerobic, Gram-negative bacterium. It has been found in roots, stems and leaves of sugarcane plants, as well as in Cameroon grass, sweet potato, coffee, tea, banana, ragi, rice and pineapple [16]. *G. diazotrophicus* worldwide distribution is responsible for almost half of the fixed N that is available to host-plants [17].

Other *G. diazotrophicus* contributions to a profitable association with the host-plants include the synthesis of the plant growth-promoting substances auxins and gibberellins [18–20] and of a bacteriocin, whose lysozyme-like activity inhibits growth of the sugarcane pathogen, *Xanthomonas albilineans* [21]. Moreover, *G. diazotrophicus* exhibits antagonistic activity against the phytotoxin produced by the fungus *Colletotrichum falcatum*, the causal agent of red-rot, the most dreaded disease in sugarcane [22].

G. diazotrophicus also secretes a constitutively expressed levansucrase (LsdA), which allows the bacterium to use sucrose, its natural carbon source substrate [23, 24]. Sucrose utilization by *G. diazotrophicus* leads to the accumulation of gluconic acids, which causes acidification of the media [25]. This highlights a secondary biofertilizer activity of *G. diazotrophicus*, since acidification can promote solubilization of zinc and inorganic phosphate compounds in the soil [26]. *G. diazotrophicus* is an endophyte, but has also been found in the rhizosphere [16], therefore, its solubilization activities might work outside and inside the host, turning the elements in the environment promptly available to increase crop productivity.

G. diazotrophicus properties studied so far suggest it is a diazotrophic organism quite distinct from other root-associated bacteria [27]. Therefore, efforts have been made to better understand its physiology, aiming the future development of biotechnological approaches for a more sustainable and environmentally safe crop production [28]. BNF associated to sugarcane crops is a matter of special interest. In Brazil, ethanol fuel produced from sugarcane is cheaper than oil, just as convenient and environmentally superior because it does not increase CO_2 levels. Improvement in agricultural sustainability will require further exploitation of BNF to turn it into the major source of nitrogen for plants. Its advantages include renewability and independence of fossil-based energy, whose costs and supply are affected by international politics and whose manufacture, distribution and use require a complex infrastructure [3].

The characteristics of *G. diazotrophicus* described above emphasize its abilities to directly promote growth and protect the sugarcane plant. However, the large-scale application of selected strains of *G. diazotrophicus* for ethanol production, with regard to a positive energy balance of the Brazilian biofuel program, has been hampered by inconsistent performance in field trials [29–31]. Among possible culprits is the poor knowledge regarding the molecular basis of *G. diazotrophicus*-sugarcane interaction and on the relative contribution of each symbiotic partner to the biology of the intact association.

Over the last years, the Riogene Genome Consortium joined efforts to sequence *G. diazotrophicus* strain PAL5 genomic DNA. Data analysis revealed about 3990 putative ORF (Ferreira, P. C. G. *et al.*, Riogene is a consortium set up to sequence complete genome of *G. diazotrophicus* PAL5. *http://www.riogene.lncc.br/*). Some of these encode well-known functions of *G. diazotrophicus,* such as the establish-

ment of endophytic behavior, host recognition, colonization, plant hormone production and nitrogen-fixing activity (Ferreira, P. C. G. *et al.*, Riogene is a consortium set up to sequence complete genome of *G. diazotrophicus* PAL5. *http://www.riogene.lncc.br/*). However, the majority of the ORF await validation and/or functional characterization. Therefore, current research activities aim at going beyond the realm of genome sequencing, towards proteomic analysis.

In order to achieve broader proteomic coverage of *G. diazotrophicus* PAL5, two independent approaches were used: (i) 2-DE followed by MALDI-TOF or TOF-TOF MS or (ii) 1-DE followed by chromatography in a C18 column online coupled to an ESI-Q-TOF or ESI-IT mass spectrometer. The almost 600 proteins identified make up an extensive database. For many of the identified proteins roles have not been assigned, therefore, functional genomic studies are required to extend our understanding of the bacterium physiology. Direct application of the data generated include comparative proteomics, that could be extremely beneficial not only for exploration of gene expression control by the bacterium under distinct conditions, but also allowing a better understanding of its interaction with sugarcane and other plant-hosts.

2 Materials and methods

2.1 2-DE protein separation and MALDI MS

2.1.1 Cell growth conditions and preparation of whole-cell lysates

G. diazotrophicus, lineage PAL5 (BR11281), was pre-cultured in DYGS medium (2 g/L glucose, 2 g/L yeast extract, 1.5 g/L peptone, 1.3 g/L glutamic acid, 0.5 g/L K_2HPO_4, 0.5 g/L $MgSO_4.7H_2O$, pH 6.0) [32] to exponential phase, when cells were harvested by centrifugation (18 000 × *g*, 10 min, 4°C) and washed twice in PBS buffer [33]. About 2×10^6 cells in 100 μL were inoculated in 2 L of LGI medium (0.2 g/L K_2HPO_4, 0.6 g/L KH_2PO_4, 0.2 g/L $MgSO_4.7H_2O$, 0.02 g/L $CaCl_2.2H_2O$, 0.002 g/L $Na_2MoO_4.2H_2O$, 0.01 g/L $FeCl_3.6H_2O$, 5 mL/L bromothymol blue 0.5% in 0.2M KOH, 10 g/L glucose) [27] containing 10 mM $(NH_4)_2SO_4$ in a reactor (Bioengineering, GAC - capacity for 2.5 L), at 30°C and 400 rpm. The pO_2 in the culture vessel was maintained at atmospheric levels by a constant influx of oxygen at 3 L/min and the pH was kept at 5.5 by inflow of phosphate-citrate buffer 1M pH 5.5. Growth was followed by cell number counts (Newbauer-counting chamber) and optical density at 600 nm (spectrophotometer PE - Model Lambda 11) and cells were collected 48 h after inoculation (exponential phase of growth) (Table 1).

The culture was centrifuged at 10 000 × *g* for 15 min at 4°C and the cells were washed with 100 mM Tris/HCl pH 8.0. For 2-DE analysis, LGI-grown cells were resuspended in 2-D lysis buffer (2% CHAPS, 8 M urea, 0.13 M DTT, 0.5% v/v Pharmalyte 3–10, 1 mM PMSF) and the cell suspension was submitted to 20 cycles of freezing in liquid nitrogen and thawing at 37°C. Lysis of cells was followed by optical microscopy and clear lysates were obtained by centrifugation at 14 000 × *g* for 1 h at 4°C. Protein concentration was estimated by 2D Quant Kit (GE Biosciences).

2.1.2 2-DE, staining and image acquisition

The first dimension of 2-DE was performed using the IPGPhor IEF unit according to the manufacturer recommendations (GE Biosciences). A sample containing 2 mg of protein was submitted to IEF on an 18-cm Immobiline DryStrip pH gradient 4–7 (GE Biosciences). After focusing, proteins on the strips were reduced and alkylated, as recommended [34]. Second dimension electrophoresis was performed on a 12% polyacrylamide/SDS gel on a *Daltsix* Unit (GE Biosciences) according to manufacturer's instructions. Proteins were stained with CBB R-250, after which, the gels were scanned (ImageScanner LabScan v5.0) and analyzed with the ImageMaster 2D Platinum v5.0 software (GE Biosciences). Values of p*I* and M_r values of the proteins of interest were determined considering a linear distribution of pH and the Rainbow broad range molecular weight protein markers (GE Biosciences), respectively.

2.1.3 In-gel digestion

Excision and processing of protein spots were performed as previously described [35]. Digestions were done with sequencing grade-modified porcine trypsin (Promega, Madison, WI) at 37°C for 16 h.

Table 1. Growth conditions, protein separation and identification methods

Growth condition	Phase of growth	Sample	Separation method	Identification method
LGI	Exponential	Whole-lysate	2-DE	MALDI-TOF and MALDI TOF/TOF
DYGS	Late exponential	Whole-lysate	1-DE	LC-Q-TOF
MS/10	Stationary	Membrane enriched fraction	1-DE	LC-Ion-Trap
MS/10	Stationary	Whole-lysate	1-DE	LC-Ion-Trap
MS/10	Stationary	Whole-lysate	IEF	LC-Ion-Trap

2.1.4 MALDI-TOF/TOF-TOF analysis

Of the tryptic peptides, 0.5 μL was mixed with 0.5 μL of a saturated solution of CHCA matrix in 50% ACN, 1% TFA. The mixture was spotted onto a MALDI sample plate and allowed to crystallize at room temperature.

MALDI-TOF analysis was performed in a Voyager DE PRO BioSpectrometer Workstation (Applied Biosystem, Foster City, CA, USA), operated in reflectron-delayed extraction mode with high resolution for the 800–4000-Da range.

MALDI-TOF-TOF peptide sequencing was performed by precursor ion fragmentation, using N_2 gas in the collision cell at 2.8×10^{-6} torr, in a 4700 Explorer Proteomics Analyzer (Applied Biosystem). Trypsin autolysis peptides masses 842.5 and 2211.1 and calibration mixture 1 or 2 (Sequazyme Peptide Mass Standard kit, PerSeptive Biosystems, Foster City, CA, USA) were used, respectively, as internal and external standards in both MS and MS/MS procedures.

Protein identification was carried out against the *G. diazotrophicus* database (Ferreira, P. C. G. *et al.*, Riogene is a consortium set up to sequence complete genome of *G. diazotrophicus* PAL5. *http://www.riogene.lncc.br/*) using the MASCOT software (Matrix Science, London, UK) [36] and the following parameters: cys-carbamidomethylation as fixed modification; methionine oxidation as variable modification; one missed trypsin cleavage, monoisotopic masses, peptide and ion tolerances at 0.5 and 0.3 Da. Under these conditions, a probability based score >51 was considered significant ($p < 0.05$). The other criteria for identification were a minimum of 20% of protein coverage and four peptides with hits in the database (for PMF data) or two peptide sequence tags (for MS/MS).

2.2 1-DE, peptide chromatography and ESI MS

2.2.1 Cell growth conditions and sample preparation

G. diazotrophicus was grown in DYGS medium in Erlenmeyers flasks, at 30°C and 150 rpm for 72 h. Cells were collected and washed as described above and resuspended in SDS-PAGE Laemmli buffer and heated for 10 min at 100°C. A clear lysate was obtained after centrifugation at $14\,000 \times g$ for 1 h at 4°C.

Alternatively, the bacterium was grown in a tenfold diluted Murashige and Skoog medium (MS/10) [37], without hormones and sucrose. The cells were harvested after 9 days, washed and resuspended in 8.1 M urea, 4% CHAPS, 1% DTT, 2% Pharmalyte, 3–10, 8 mM PMSF, 0.5% Triton X-100. Cells were disrupted by 20 cycles of freeze-thawing as in Section 2.1, sonication and vortexing. A clear lysate was obtained after centrifugation at $57\,000 \times g$ for 2.5 h at 5°C. The supernatant was colleted and stored at –80°C until used. Membrane proteins in the pellet of the latter centrifugation were prepared as previously described [38, 39] and resuspended in 1% CHAPS, 2% NP-40, 5% Triton X-100, 1% Pharmalyte, 8 M urea.

Protein concentration was estimated by 2D Quant Kit (GE Biosciences) or the BCA protein assay [40].

2.2.2 1-DE and IEF

Two independent, but complementary protein separation approaches were used: SDS-PAGE or IEF. SDS-PAGE was performed on a 12% polyacrylamide gel in a mini Protean II system (Bio-Rad).

IEF was carried out on Immobiline DryStrip (IPG, 7 cm, linear pH gradient 3–10, GE Biosciences). Proteins (80 μg) were completed to a volume of 130 μL resulting in a final concentration of 8.1 M urea, 2% CHAPS, 0.28% DTT, 1% IPG buffer 3–10, and 0.002% w/v Bromophenol Blue. The IPGphor IEF System was used to perform an active rehydratation step for 12 h at 30 V. IEF conditions were according to the manufacturer's instructions.

2.2.3 In-gel digestion

Excision and processing of protein spots from SDS-PAGE gels were performed as previously described [35].

IPG strips were cut in 0.7-cm pieces and proteins were reduced with 350 μL of 25 mM NH_4HCO_3, pH 8.0 and 50% ACN, containing 10 mg/mL DTT. Proteins were then alkylated with 350 μL of a 30 μg/mL iodoacetamide (IAA) solution. Each piece was shrunk by dehydration in ACN and dried in a vacuum centrifuge. Trypsin digestions were performed as previously described [41]. Peptides were extracted in 80 μL of 50% ACN and 5% formic acid (FA). Samples were concentrated by vacuum centrifugation to approximately 10 μL and stored at −20°C until MS analysis.

2.2.4 ESI-Q-TOF analysis

LC-MSMS was performed in a Q-TOF Ultima API mass spectrometer (Micromass, Manchester, UK) to analyze tryptic peptides from proteins separated by SDS-PAGE. A nanoflow ESI source was used with a lockspray source for lock-mass measurements during all the chromatographic run. Samples of approximately 0.5 μg of digested proteins were desalted online using a Waters Opti-Pak C18 trap column. The mixture of trapped peptides was then separated by elution with a water/ACN 0.1% formic acid gradient through a Nanoease C18 (75-μm id) capillary column. Data were acquired in data-dependent mode (DDA), and multiple-charged peptides ions (+2 and +3) were automatically mass selected and dissociated in MS/MS experiments. Typical LC and ESI conditions were a 200 nL/min flow, a 3 kV nanoflow capillary voltage, a 100°C block temperature and a 100 V cone voltage.

The MS/MS spectra were processed using the ProteinLynx 2.0 software (Waters, Milford) and the pkl file generated was searched against the database using the MASCOT software (MatrixScience, London, UK).

2.2.5 C18 chromatography coupled to ESI-IT MS

Digested protein samples were resolved on a 15 cm × 300 μm id (300 Å; 3 μm) ProteCol C18 column (SGE, Australia). Mobile phase A was 0.1% v/v formic acid in water, and B was 0.1% v/v formic acid in ACN. Eluting peptides in the column effluent were directly electrosprayed into a LCQ Deca XP Plus IT spectrometer (Thermo Finnigan, USA). The voltage applied to the electrospray source was 3.30 kV, while the inlet capillary was held at 31.0 V and 180°C. Ten arbitrary units (range 0–100) of sheath gas flow were used. Spectra of eluting peptides were acquired in a data-dependent fashion by first acquiring a full MS scan from *m/z* 400 to 2000 followed by consecutive MS/MS scans of the five most intense ions of the previous full MS scan.

The MS/MS spectra were extracted and the charge state was deconvoluted and deisotoped by BioWorks version 3.3. All MS/MS samples were analyzed using SEQUEST (ThermoFinnigan; version 27, rev. 12) and X!Tandem (http://www.thegpm.org version 2007.01.01.1) against the *G. diazotrophicus* database. Fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 1.4 Da were used in the searches. Iodoacetamide derivative of cysteine was specified in SEQUEST and X!Tandem as a fixed modification. Oxidation of methionine, tryptophan oxidation to formylkynurenin and acrylamide adduct of cysteine were specified in SEQUEST and X!Tandem as variable modifications.

The Scaffold (version Scaffold-01_06_18, Proteome Software, Portland, OR) software was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm [42]. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least two identified peptides. Protein probability scores were assigned by the Protein Prophet algorithm [43].

3 Results and discussion

3.1. Protein database assembly and expression profile of *G. diazotrophicus*

3.1.1 2-DE analysis followed by MALDI MS

The whole-cell lysate proteome profile of *G. diazotrophicus* grown in LGI, was obtained by 2-DE. Three 2-DE gels of *G. diazotrophicus* were run that showed the same protein profile. One of them was selected as the reference (Fig. 1A). A set of 368 protein spots spread over the pH range 4 to 7 was clearly visualized on the gel after CBB staining. The vast majority of proteins (*ca* 80%) possessed M_r in the range 22 to 97 kDa (Fig. 1A); only 7% were bigger than 97 kDa and 13% smaller than 22 kDa.

All 368 protein spots were trypsin digested and analyzed by MALDI MS. From these, 205 proteins (*ca* 52%) were identified at $p \leq 0.05$. Fig. 1A shows the successfully identified spots numbered c1 to c368, as in the Supporting Information, where number of peptides matched, scores, experimental p*I* and M_r, peptide sequences, sequence coverage and charge state are also listed.

Some spots contained two proteins; therefore, the 205 proteins derived, in fact, from 190 spots. Multiple protein identifications, in some cases, result from streaky spots and/or diffusion of proteins in close proximity.

Several other proteins were isoforms, probably derived from PTM of the same primary sequence. The majority of these isoforms presented little or no observable M_r modification, but differed in charge. Thus, the 205 proteins were, ultimately, the products of 161 distinct genes (*i.e.* independent accession numbers, Supporting Information Table S1).

Assuming that the *G. diazotrophicus* genome encodes about 3990 proteins, the 161 proteins identified correspond to 4% of the proteomic potential of the bacterium. To raise the ratio of identified/predicted proteins we employed an alternative gel-based analysis: 1-DE protein separation, in gel protein trypsinization, separation of the tryptic peptides generated and MS/MS analysis (1-DE/LC/MS/MS). Moreover, to maximize the chance of detecting as many distinct proteins as possible by 1-DE/LC/MS/MS, we analyzed lysates from cells cultured under different conditions and used diverse protein extraction methods, as described in Section 2.2.

3.1.2 ESI-Q-TOF analysis

G. diazotrophicus was grown in DYGS, the cells were lysed in Laemmli buffer and soluble proteins were separated by 1-DE. The resulting gel was sliced into ten pieces of similar sizes (Fig. 1B) and the proteins were in-gel digested with trypsin.

The tryptic peptide mixtures were separated as described in Section 2.2, according to hydrophobicity. This methodology permitted the separation of proteins and peptides, respectively, according to M_r and hydrophobicity.

A set of 13 872 MS/MS spectra were acquired, allowing the identification of 537 proteins from distinct genes at $p<0.05$ (Supporting Information Table S1). Since the identity of 127 from the 537 had been previously determined by the 2-DE/MALDI strategy, overall, we managed to identify 574 different proteins (Supporting Information Table S1).

3.1.3 ESI-IT analysis

Subcellular fractions of *G. diazotrophicus* grown in MS/10 medium were analyzed on ESI-IT. Soluble and membrane proteins of *G. diazotrophicus* were separated by SDS-PAGE and the whole-cell lysate proteins were separated by IEF on Immobiline DryStrips. Each gel was cut in ten pieces and proteins were in-gel trypsin digested. Tryptic peptides were resolved on a C18 column and directly electrosprayed into an IT mass spectrometer. In all, 96 proteins were identified,

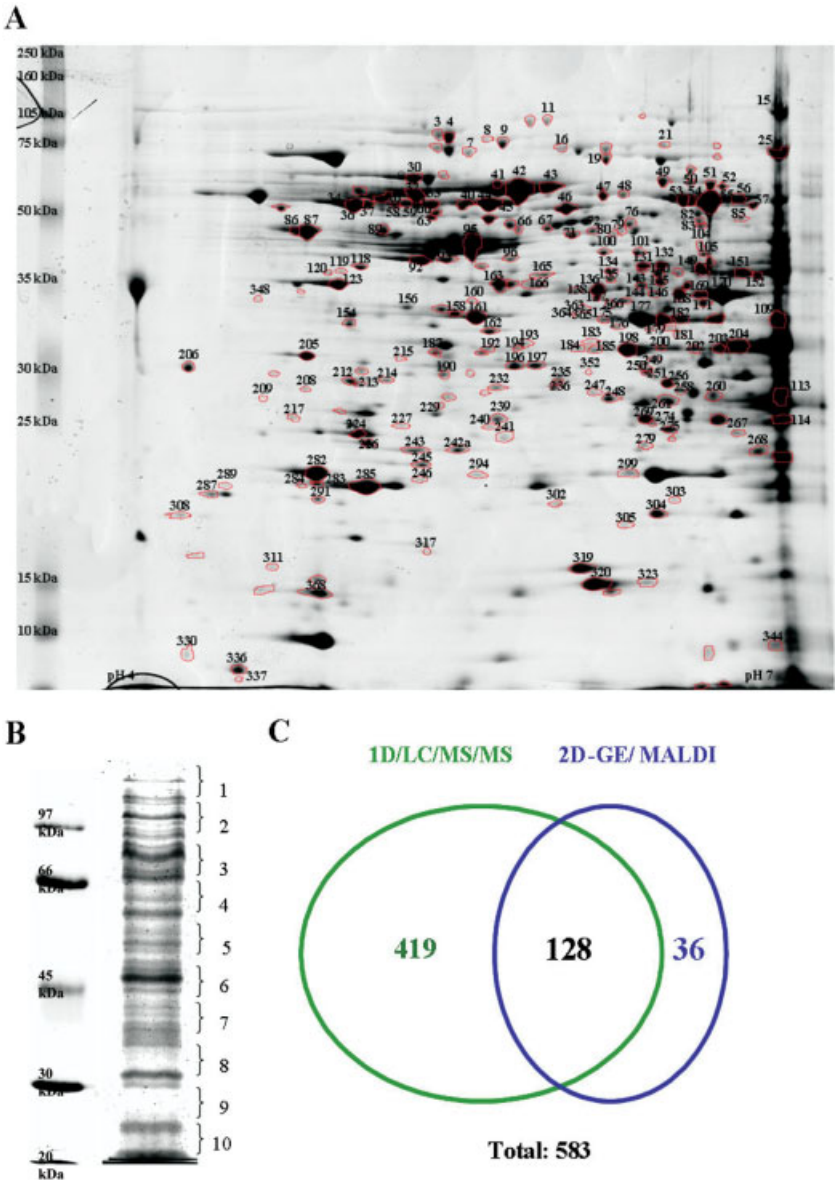

Figure 1. Methods of *G. diazotrophicus* protein expression profile construction and analysis. (A) 2-DE electrophoresis map of whole-cell lysate from *G. diazotrophicus*, showing the 368 visualized spots. 205 proteins in the 190 numbered spots were identified by MS. (B) The 1-DE gel of whole-cell lysate proteins from DYGS grown *G. diazotrophicus*, sliced in ten fragments used for in-gel protein tryptic digestion. (C) Venn diagram with the sets of proteins identified by each method: 161 distinct proteins were identified by 2-DE, 547 by 1-DE/LC/MS/MS, with an intersection of 128 proteins of a total of 583 proteins.

with confidence level >95% and a minimum of two peptides identified per protein. Statistics and reliability analyses were performed in Scaffold proteome software.

Since 87 of the 96 proteins had been identified by MALDI or LC-Q-TOF, only nine novel proteins were added to the proteome database of *G. diazotrophicus*. As a whole, 583 *G. diazotrophicus* proteins were successfully identified in this work (36 exclusively by 2-DE/MALDI, 128 by both methods and 419 exclusively in the 1-DE/LC/MS/MS approach; Fig. 1C, Supporting Information Table S1). This represents 15% of the estimated proteome of the bacterium relative to genome encoded proteins. Only 4% of the proteins were identified by 2-DE and MALDI MS. Further protein identification was only possible by the use of 1-DE/LC/MS/MS, confirming the robustness of these procedures.

3.2 Localization, functional classification and special categories of proteins

3.2.1 Protein cellular localization sites

The majority of the proteins described here are soluble; however, some membrane-associated proteins were also identified. These were outer membrane lipoproteins (spots c15, c109, c152, c203, c204, c25, c368), GTP-binding protein TypA/BipA (spot c5), a dipeptidase M19 (spot c156) and several ATP synthase subunits (spots c330, c34, c36). Membrane-associated proteins identified by 1-DE/LC/MSMS comprised an UDP-N-acetylenolpyruvoylglucosamine reductase (cell wall formation [44]), an outer membrane lipoprotein, Omp16, Porin B precursor (of the outer mem-

brane efflux protein), two putative surface antigens and a YaeT precursor. In *Escherichia coli*, *yaeT* gene codes for an essential function in targeting and folding of outer membrane proteins [45, 46].

Export processes proteins included a polysaccharide export protein, CtrD (capsule polysaccharide export ATP-binding protein), HlyD (membrane fusion protein of type-I secretion system [47]) and type II/IV secretion system protein, SecB (protein export chaperone [48]).

Identification of additional membrane proteins from *G. diazothrophicus* is of special interest, since in the endophyte stage the bacteria are in intimate contact with plant cells surfaces, nutrients and other microorganisms within the plant apoplast compartment. Moreover, *in vitro*, drastic ultrastructural alterations have been observed in many N-fixing endophytes, including *G. diazothrophicus* [16] as a consequence of protein pattern changes, including membrane remodeling. Therefore, a proteomic analysis of membrane envelope proteins of *G. diazothrophicus* under distinct physiological states would certainly lay a foundation for a better understanding of membrane functions in the cell.

3.2.2 Functional classification of proteins

The 583 proteins were grouped according to the Sanger classification system for bacterial proteins [49] (Fig. 2). The major group included 220 proteins (38% of those identified) with functions in the central/intermediary/miscellaneous metabolism. The second major group comprised 84 proteins with roles in information transfer (transcription, translation, and modification of nucleic acids). Other well-represented groups were surface (54) and energy metabolism (46) proteins.

A group of 50 conserved unknown, as well as 27 unknown proteins were also identified. An interesting set comprehended 32 proteins with roles in adaptation to atypical conditions, folding of molecules and pathogenicity. Twenty-seven regulatory proteins were found, even though these are usually low-abundant in bacterial cells. Other categories represented were those involved in degradation of large (39) and small molecules (4).

3.2.3 Proteins involved in transport

Twenty-four of the proteins identified, were transporters or members of transport systems. Among these were several ABC transport system components, D-ribose, D-xylose and fucose periplasmic binding proteins, ATP-binding proteins of sugar and vitamin B1 uptake systems and a member of an uncharacterized transport process. Other transporter system protein was SufC, a versatile ATPase that can associate either with other Suf proteins to form a Fe–S cluster-assembling machinery or with membrane proteins to form an Fe–S ABC exporter complex [50].

In addition, we managed to identify PhoU, a protein involved in the transport of inorganic phosphate under limiting conditions in many bacterial species [51]. However, PhoU in *E. coli* also plays role as a global negative regulator, beyond its role in phosphate metabolism, shutting down cellular metabolism to ensure metabolically quiescent bacteria survival under various stress conditions [52].

Other transport-related proteins identified were ExbB, of the biopolymers transport system, and TolB plus TonB. TolB is a component of a multiprotein translocation complex, with roles in the recycling of peptidoglycan and in the covalent linking of peptidoglycan and lipoproteins [53]. TonB, on the other hand, is the transducer of the proton-motive force

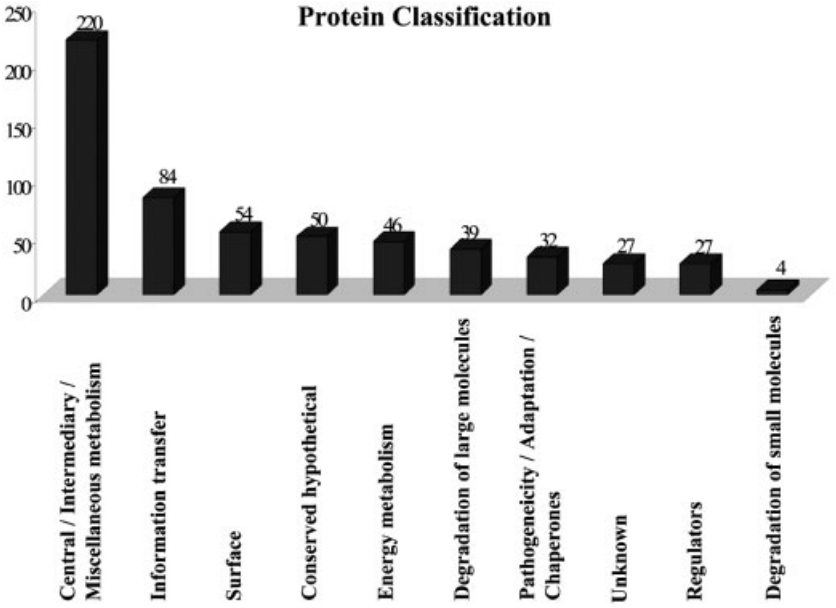

Figure 2. Classification of the *G. diazotrophicus* proteins identified according to the functional categories of the Sanger prokaryotic scheme. Absolute values per categories are shown.

energy for active transport of iron-siderophore complexes and vitamin B12 across the outer membrane [54]. Interestingly, in *E. coli*, both TonB and TolB are involved in the translocation of colicins and in this process TonB functions in association with protein ExbB [55, 56], mentioned above. Mechanisms of iron acquisition, such as the one involving TonB are quite important for endophytic bacteria, since free iron level is low in plants and iron sequestering is important for host-bacteria relationships [54].

3.2.4 Nitrogen fixation-related proteins

Nitrogenase is an enzymatic complex that catalyzes the reduction of dinitrogen to ammonia coupled to the hydrolysis of ATP [57]. It consists of two component metalloproteins, the iron protein (NifH) and molybdenum-iron protein (NifD, NifK) [58]. The major cluster of *nif/fix* and associated genes of *G. diazotrophicus* have been characterized [59, 60]. Their arrangement on the genome of the bacterium is quite similar to that of the *nif/fix* cluster in *Azospirillum brasilense*, however, the individual gene products are more similar to those in species of *Rhizobiaceae* or in *Rhodobacter capsulatus* [60].

We managed to detect, in some of the samples, the three nitrogenase proteins NifH, NifD and NifK, as well as NifA and NtrX. NifA is the transcriptional activator of the *nif*HDK operon and NtrX is the response regulator of the two-component system NtrY/NtrX involved in the activation of nitrogen assimilatory genes such as *gln*A [61].

3.2.5 Proteins involved in adaptation and protection responses

Among the proteins involved in cellular adaptation and protection responses we found an universal stress-related protein, five chaperones, three proteins involved in iron metabolism, two antioxidant proteins, two cold-shock proteins, an organic solvent tolerance protein, a multidrug resistance protein EmrA, a bacteriocin and a paraquat-inducible protein PqiB.

The chaperones identified were DnaK, DnaJ, HtpG, a chaperone-binding protein and SurA. The ensemble of chaperones comprises systems for ATP-dependent folding and assembly of newly synthesized proteins, for the translocation of unfolded proteins across membranes and refolding and degradation of misfolded and aggregated proteins [62].

One of the proteins that provide folding assistance during protein biosynthesis is DnaK. The specificity of the folding processes depends on the activity of co-chaperones, such as DnaJ, functioning together with DnaK and GrpE [63]. SurA, on its turn, is involved in the folding of extracytoplasmic proteins, but has also a protective role, since is required for *E. coli* survival during the stationary growth phase [64]. HtpG, on the other hand, a homologue of the mammalian chaperone HSP90, is not essential for *E. coli* survival [65], but has a crucial role in thermal stress management in cyanobacteria [66].

Chaperone and associated proteins that allow bacterial cells to cope with alterations of protein conformation are usually expressed in response to an increase in ambient temperature. Cold-shock response, to the contrary, comprises predominantly proteins such as helicases, nucleases, and ribosome-associated components [67]. Two putative cold-shock response proteins namely, DEAD box protein A and protein CspE, were identified in *G. diazotrophicus*. Interestingly, CspE in *E. coli* is not cold induced, but is constitutively expressed at 37°C and it is functionally important, since it regulates the expression of two proteins involved in the stress response. Therefore, the expression of DEAD box protein A and of CspE in *G. diazotrophicus* might be an indication that they function in the bacterial adaptation to other external stress factors besides protecting against cold distress. However, there is also evidence that they play important roles in bacterial cell physiology under normal conditions. Experimental data suggest they bind mRNA molecules to regulate ribosomal translation and control the rate of mRNA degradation and termination of transcription [68].

Two proteins involved in cellular protection are superoxide dismutase (SOD) and peroxiredoxin, which take part in cell defenses against reactive oxygen species (ROS) [69].

3.2.6 Regulatory proteins

Thirty-four (*ca.* 6%) of the 583 proteins have predicted regulatory activities, contrasting with the notion that proteomic analysis is unable to detect low-abundance regulatory proteins. These proteins are likely to regulate global networks and mutagenesis of the corresponding genes should yield important insights into the control of overall cellular metabolism. Some of the regulatory proteins identified were BolA-like protein, which in *E. coli* is essential for a shift in cell morphology and leads to increased transcription of the cell wall synthetic genes *dacA* (PBP5), *dacC* (PBP6) and *ampC* [70]; N utilization substance protein B, involved in the transcription termination process [71]; ROS/MucR transcriptional regulator protein, which in *E. coli* regulates diverse functions, including biosynthesis of the exopolysaccharides succinoglycan and galactoglucan [72]; transcriptional regulatory protein ChvI, a member of a two-component system that regulates succinoglycan synthesis [73]; and a putative transcriptional regulator LysR.

Other regulators identified include transcription antitermination protein NusG, transcriptional regulator of the IclR family/regucalcin, a transcriptional regulator of CarD family, polyphosphate kinase, inosine-guanosine kinase, ribokinase, ribulokinase, phosphoglycerate kinase, uridylate kinase, PrkA serine protein kinase and a putative serine protease, the three nitrogenase proteins NifH, NifD and NifK, as well as NifA and NtrX, a response regulator of the NtrY/NtrX two-component system.

3.3 Metabolic pathways in *G. diazotrophicus*

Enzyme Commission numbers (EC) could be assigned to 234 out of the 583 proteins. These EC numbers were used to search the Kyoto Encyclopedia of genes and genomes (KEGG) database for metabolic routes. As a great number of proteins can be shared by various metabolic pathways, identification of a few proteins in a pathway may be the consequence of this interconnected metabolic network. Therefore, a pathway was deemed potentially active in *G. diazotrophicus* when a minimum of three different representative enzymatic activities were identified. Furthermore, at least one of these proteins should be exclusive of the referred pathway.

By applying these criteria to the set of proteins identified in this work, 33 pathways were represented in *G. diazotrophicus* protein database (Table 2). In several cases, a complete or near-complete set of enzymes of a pathway was identified.

Amino acid metabolism was represented by 12 pathways, namely, glutamate metabolism (12 proteins), glycine, serine and threonine (11), methionine (9), aspartate and alanine (8), phenylalanine, tyrosine and tryptophan (7), glutathione (7), valine, leucine and isoleucine (6), lysine (6), arginine and proline (6), tyrosine (5), cysteine (3) and urea cycle and metabolism of amino groups (4).

Nine pathways for carbohydrates metabolism were characterized. The most represented pathways were pentose

Table 2. Metabolic pathways represented on the proteome of *G. diazotrophicus*, according to KEGG[a)]

	Metabolic pathways	Nº Ecs
Amino acids metabolism (12)	1- Glutamate metabolism	12
	2- Glycine, serine and threonine metabolism	11
	3- Methionine metabolism	9
	5- Alanine and aspartate metabolism	8
	4- Phenylalanine, tyrosine and tryptophan metabolism	7
	6- Glutathione metabolism	7
	7- Valine, leucine and isoleucine metabolism	6
	8- Lysine biosynthesis	6
	9- Arginine and proline metabolism	5
	10- Tyrosine metabolism	5
	11- Urea cycle and metabolism of amino groups	4
	12- Cysteine metabolism	3
Carbohydrate metabolism (9)	14- Pentose phosphate pathway	14
	13- Glycolysis/Gluconeogenesis	13
	15- Pyruvate metabolism	9
	16- Citrate cycle (TCA cycle)	8
	17- Fructose and mannose metabolism	5
	18- Butanoate metabolism	5
	19- Glucose and glucoronate interconversions	4
	20- Starch and sucrose metabolism	4
	21- Glycoxylate and dicarboxylate metabolism	4
Energy metabolism (3)	22- Carbon fixation	13
	23- Methane metabolism	7
	24- Nitrogen metabolism	4
Metabolism of cofactors and vitamins(2)	25- Panthotenate and CoA biosynthesis	8
	26- Phorphyrin and chlorophyll metabolism	4
Lipid metabolism (2)	27- Fatty acids biosynthesis	4
	28- Biosynthesis of steroids	4
Biosynthesis of secondary metabolites (1)	29- Streptomycin metabolism	6
Nucleotide metabolism (2)	30- Purine metabolism	20
	31- Pyrimidine metabolism	17
Genetoic information processing (1)	32-Aminoacyl-tRNA biosynthesis	11
Environmental information processing (1)	33- Two-component system	5

a) A pathway was deemed potentially active in *G. diazotrophicus* when a minimum of three different enzymatic activities were identified and, at least one of them should be exclusive of that pathway. Numbers of pathways *per* metabolic group (*n*) and of proteins (Nº ECs) *per* pathway are shown.

phosphate (PPP, 14 proteins), glycolysis/gluconeogenesis (13; see Ssection 3.3.1), pyruvate metabolism (9) and citric acid cycle (TCA, 8). In addition, 13 proteins for carbon fixation, seven for the methane metabolism and four of the nitrogen metabolism were among those of the energy metabolism group.

Nucleotide pathways comprehended 20 proteins for purines and 17 for pyrimidines metabolisms. Other metabolic routes defined in this work were two for cofactors biosynthesis, two for lipid metabolism, one for secondary metabolite biosynthesis, one for genetic information processing and one for environmental information processing.

3.3.1 Glucose utilization: direct oxidation, pentose phosphate pathway, and other routes for carbohydrates utilization

It has been previously described that the key step in glucose utilization by *G. diazotrophicus* is its periplasmic oxidation to gluconate catalyzed by a the periplasmic pyrroloquinoline quinone (PQQ)-linked glucose dehydrogenase [74]. Accordingly, some enzymes for the utilization of gluconate were identified in this work: gluconate 5-dehydrogenase, 6-phosphogluconate dehydrogenase, a key enzyme of the PPP (see below), and 2-ketogluconate reductase.

Other intracellular mechanisms for carbohydrate utilization rely on NAD-glucose dehydrogenase [75, 76] and include the PPP and Embden-Meyerhof pathways. A third one, is the Entner-Doudoroff pathway, with two key enzymes (EC 4.2.1.12 and 4.1.2.14), which have not been detected in the samples of *G. diazotrophicus* analyzed. This pathway, however, is present in species of the *Acetobacteriaceae* family, ex. *Acetobacter xylinum* [77] and among the *Gluconobacter* species, ex. *G. oxydans*, whose genome has been sequenced recently [78]. Therefore, the presence of the Entner-Doudoroff pathway in *G. diazotrophicus* cannot be discarded and further investigation is required.

The PPP has been described as the main pathway for the catabolism of sugars among *Acetobacteriaceae* [79, 80]. The first stage of the PPP, the oxidative stage, is a rate-limiting step redox reaction catalyzed by glucose-6-phosphate-dehydrogenase (EC 1.1.1.49) and by the enzymes EC 3.1.1.31, EC 5.1.3.1, and EC 5.3.1.6 (Supporting Information Table S1). Feeding this pathway, the gluconate produced by the periplasmic oxidation of glucose, enters directly into PPP as 6-phospho-gluconate [76]. The second stage of PPP is non-oxidative and catalyzed by the enzymes EC 4.1.2.9, EC 2.2.1.1 and EC 2.2.1.2, also described herein (Supporting Information Table S1).

One important enzyme, 6-phosphofructokinase (EC 2.7.1.11), for the commitment step in the glycolysis pathway was not identified, in agreement with previous studies [76, 79]. The set of enzymes of glycolysis/gluconeogenesis pathway identified was: 5.4.2.2, 3.1.3.11, 4.1.2.13, 5.4.2.4, 5.4.2.1, 4.2.1.11 and 2.7.1.40. The detection of proteins that turns glycerate-1,3 biphosphate into pyruvate plus the presence of a typical catabolic enzyme, the pyruvate kinase (EC 2.7.1.40) might indicate some catabolic role for the lower part of the Embden-Meyerhof pathway. However, the identification of fructose-bisphosphate aldolase (EC 4.1.2.13) and, most importantly, of the fructose-1,6-bisphosphatase (EC 3.1.3.11) strongly suggests that this pathway also functions in gluconeogenesis in *G. diazotrophicus*. The data presented here complement previous work since, instead of enzyme activities, the proteins themselves were detected and unquestionably identified.

Several TCA cycle enzymes were identified: ATP citrate synthase (EC 2.3.3.1), aconitate hydratase (EC 4.2.1.3), isocitrate dehydrogenase (EC 1.1.1.42), dihydrolipoamide succinyltransferase component E2 of 2-oxoglutarate dehydrogenase complex (EC 1.2.4.2), succinyl-CoA synthetase beta chain (EC 6.2.1.5), two subunits of succinate dehydrogenase (flavoprotein and iron-sulfur proteins), and fumarate hydratase (EC 4.2.1.2). Some subunits of pyruvate dehydrogenase, used to produce acetyl-CoA, that fuels the TCA cycle, were also found. However, the two main enzymes of the glyoxylate by-pass, isocitrate lyase and malate synthase, were not found, in agreement with a previous study [76]. A strain of *G. diazotrophicus*, PAL3, has a complete TCA cycle [76], and it is assumed that this could be true for *Gluconacetobacter* in general. In contrast, *Gluconobacter*, another genus of the *Acetobacteriaceae* family, has an incomplete TCA cycle [78, 81].

3.3.2 Oxidative phosphorylation and complexes of electron transport chain

Respiration is an important feature of *G. diazotrophicus* physiology, which relies on a high respiratory capacity to protect its nitrogenase from O_2 [25, 27]. In the *Acetobacteraceae* family a large number of membrane dehydrogenases (quinoproteins and flavoproteins) yield electrons to the respiratory chain [25, 82]. The general structure of the aerobic respiratory system of *G. diazotrophicus* includes various dehydrogenases, cytochrome c as electron carrier, ubiquinone Q10 as the main electron collector in the inner membrane, and terminal ubiquinol oxidases [83].

Members of the respiratory electron transport chain were detected as a large group of approximately 30 proteins. The presence of six components of the NADH:ubiquinone oxireductase, namely, subunits B, C/D, E, F, G and I, suggests high abundance of an enzyme composed of many polypeptides, identified as type 1 NADH dehydrogenases (NDH-1) [84]. Various subunits of alcohol dehydrogenases (ADH) were found in *G. diazotrophicus*: one is highly similar to the ADH from *G. europaeus* and *G. polyoxogenes* [85]; two smaller ones, classified as GroES-like ADH, presented similarity to GOX0313 and GOX0314 from *G. oxydans* and other two were zinc-binding proteins, one similar to the putative oxidoreductase GOX2594, also from *G. oxydans* and, the other, a homologue of YP_159089 from *Azoarcus* sp. EbN1. Of the aldehyde dehydrogenases (ALDH) identified, one is similar

to that found in *G. europaeus* (accession number CAA69955) [86] and the other is smaller and more strongly related to the ALDH from *Xanthobacter autotrophicus*, locus AAC13641 [87].

Four subunits of an electron-transfer flavoprotein were identified: FixA, FixB, FixC and EtfB. The roles of these proteins in *G. diazotrophicus* are not known, however, in *R. meliloti*, *fixABC* genes are located in a single operon, are co-transcriptionally activated by the NifA with the nitrogenase genes *nifH*, *D*, *K*, and are essential for nitrogen fixation [88].

We also managed to identify two NADH:flavin oxireductases of the OYE family, COG1902, one quinone oxireductase, COG0604, one nitroreductase, COG0778, an additional PQQ-containing dehydrogenase, similar to *Gluconobacter oxydans* GOX1969, and one nitropropane deoxygenase, like GOX1587.

Membrane ubiquinone is the general electron acceptor for all bacterial membrane-bound dehydrogenases [25, 84]. The ubiquinones, and in particular Q10, are active electron carriers receiving electrons from the outer and inner faces of the membrane [89]. *G. diazotrophicus* in addition to Q10 its major ubiquinone, has, in smaller proportions, the ubiquinone Q9 and PQQ. Ubiquinone synthesis proceeds in three stages, the formation of the ring, the addition of a hydrophobic polyisoprenoid chain, and modifications of the ring [90]. Among the enzymes that work in these pathways we did identify a geranyltranstransferase of the prenyl chain synthesis stage (EC 2.5.1.10) and ring modifying enzymes, UbiE (a methyltransferase), UbiG (a SAM O-methyltransferase) and UbiH (a hydroxylase).

It has been shown that *G. diazotrophicus*, an oxidase negative species, uses a ubiquinol oxidase, instead of cytochrome c oxidase, at the terminal end of the respiratory chain [25], and two ubiquinol oxidases were described in the bacterium. One is the cytochrome ba-type KCN-resistance and the other [25], the cytochrome bb/bd-type KCN-sensitive [84]. In this work, the subunit II of the cytochrome ba-type ubiquinol oxidase was identified. It has high similarity in amino acid sequence to the corresponding subunit of the cytochrome a1 (ba)-type ubiquinol oxidases of both, *Acetobacter pasterianus* and *A. aceti* [91].

4 Concluding remarks

This is the first broad proteomic description of *G. diazotrophicus*, an endophytic bacterium, responsible for the major fraction of the nitrogen fixed in sugarcane in tropical regions.

The 208 proteins identified by 2-DE/MALDI are the starting point for the establishment of a proteome reference map for *G. diazotrophicus*. If we add those to the set identified by 1-DE/LC/MS/MS, we end up with 583 distinct proteins that were sorted into their functional categories and were used to describe metabolic pathways potentially functional in this bacterium. Accordingly, many metabolic pathways related to nucleotides, amino acids, carbohydrates, lipids, cofactors and energy metabolism have been described. The identification of such conserved biochemical pathway will help comparisons between *G. diazotrophicus* and other bacterial species. Further proteomic studies, however, will be required to highlight species-specific pathways to document differences.

From the 583 proteins in the database, nearly 13% were classified as conserved unknown or unknown proteins. These yet unexplored data constitute a potential target for functional genomic studies. Therefore, the proteomic database presented herein lays the foundation for future studies on the physiology and metabolism of *G. diazotrophicus*. The knowledge of metabolic fundamentals and coordination of these actions are crucial for the rational, safe and sustainable interference on crops. The entire dataset, including peptide sequence information, is available as Supporting Information and is the major contribution of this work.

We thank Dr. Fabio Lopes Olivares, from Universidade Estadual do Norte Fluminense Darcy Ribeiro for helpful discussion of the manuscript. We are grateful to LNLS/MCT for the use of the 4700 Proteomics Analyzer and Q-TOF Ultima API mass spectrometer. We thank FAPERJ, CNPq and CAPES for financial support.

The authors have declared no conflict of interest.

5 References

[1] Kim, J., Rees, D. C., Nitrogenase and biological nitrogen fixation. *Biochemistry* 1994, *33*, 389–397.

[2] Muthukumarasamy, R., Revathi, G., Seshadri, S., Lakshminarasimhan, C., *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. *Curr. Sci.* 2002, *83*, 137–145.

[3] Hardy, R. W. F., Biological Nitrogen Fixation: Research Challenges – A Review of Research Grants Funded by the U.S. Agency for International Development. *National Academy Press* 1994, *1*.

[4] Sawada, H., Kuykendall, L. D., Young, J. M., Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. *J. Gen. Appl. Microbiol.* 2003, *49*, 155–179.

[5] Young, J. P. W., Phylogenetic classification of nitrogen-fixing organisms. In: Stacy, G., Burris, R. H., Evans, H. J. (Eds.). *Biological Nitrogen Fixation*. Chapman e Hall, New York NY 1992, pp. 43–86.

[6] Ferguson, S. J., Nitrogen cycle enzymology. *Curr. Opin. Chem. Biol.* 1998, *2*, 182–193.

[7] Döbereiner, J., History and new perspectives of diazotrophs in association with non-leguminous plants. *Symbiosis* 1992, *13*, 1–13.

[8] Azevedo, J. L., Maccheroni Junior, W., Pereira, J. O., Araújo, W. L., Endophytic microorganisms: a review on insect control and recent advances on tropical plants. *Electronic J. Biotechnol.* 2000, *3*, 40–65.

[9] James, E. K., Nitrogen fixation in endophytic and associative symbiosis. *Field Crops Res.* 2000, *65*, 197–209.

[10] Reinhold-Hurek, B., Hurek, T., Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 1998, *6*, 139–144.

[11] Lima, E., Boddey, R. M., Döbereiner, J., Quantification of biological nitrogen fixation associated with sugar cane using a 15N-aided nitrogen balance. *Soil Biol. Biochem.* 1987, *19*, 165–170.

[12] Boddey, R. M., Biological nitrogen fixation in sugar cane: a key to energetically viable biofuel production. *Crit. Rev. Plant Sci.* 1995, *14*, 263–279.

[13] Boddey, R. M., Sá, J. C. M., Alves, B. J., Urquiaga, S., The contribution of biological nitrogen fixation for sustainable agricultural systems in tropics. *Soil Biol. Biochem.* 1997, *29*, 787–799.

[14] Reis, V. M., Olivares, F. L., Döbereiner, J., Improved methodology for isolation of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat. *World J. Microbiol. Biotechnol.* 1994, *10*, 401–405.

[15] Cavalcante, V. A., Döbereiner, J., A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. *Plant Soil* 1988, *108*, 23–31.

[16] Muthukumarasamy, R., Revathi, G., Seshadri, S., Lakshminarasimhan, C., *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. *Curr. Sci.* 2002, *83*, 137–145.

[17] Boddey, R. M., Urquiaga, S., Reis, V., Döbereiner, J., Biological nitrogen fixation associated with sugar cane. *Plant Soil* 1991, *137*, 111–117.

[18] Lee, S., Sevilla, M., Meletzus, D., Gunapala, N., Kennedy, C., in: Stacey, G., Keen, N. (Eds.), *Plant Microbe Interactions*, APS Press 2000, *5*, 196–204.

[19] Bastian, F., Cohen, A., Piccoli, P., Luna, V., Baraldi, R., Production of indole-3-acetic acid and gibberellins A1 y A3 by Acetobacter diazotrophicus and Herbaspirillum seropedicae in chemically-defined culture media. *Plant Growth Reg.* 1998, *24*, 7–11.

[20] Fuentes-Ramirez, L. E., Jimenez-Salgado, T., Abarca-Ocampo, I. R., Caballero-Mellado, J., *Acetobacter diazotrophicus*, and indolacetic acid producing bacterium isolated from sugarcane cultivars of Mexico. *Plant Soil* 2003, *154*, 145–150.

[21] Pinon, D., Casas, M., Blanch, M., Fontaniella, B. *et al.*, *Gluconacetobacter diazotrophicus*, a sugar cane endosymbiont, produces a bacteriocin against *Xanthomonas albilineans*, a sugar cane pathogen. *Res. Microbiol.* 2002, *153*, 345–351.

[22] Muthukumarasamy, R., Rebathi, G., Vadivelu, M., Antagonistic potential of N2-fixing Acetobacter diazotrophicus against Colletotrichum falcatum Went., a causal of red-rot of sugarcane. *Curr. Sci.* 2000, *78*, 1063–1065.

[23] Arrieta, J. G., Sotolongo, M., Menendez, C., Alfonso, D. *et al.*, A type II protein secretory pathway required for levansucrase secretion by *Gluconacetobacter diazotrophicus*. *J Bacteriol.* 2004, *186*, 5031–5039.

[24] Hernandez, L., Sotolongo, M., Rosabal, Y., Menendez, C. *et al.*, Structural levansucrase gene (lsdA) constitutes a functional locus conserved in the species *Gluconacetobacter diazotrophicus*. *Arch. Microbiol.* 2000, *174*, 120–124.

[25] Flores-Encarnacion, M., Contreras-Zentella, M., Soto-Urzua, L., Aguilar, G. R. *et al.*, The respiratory system and diazotrophic activity of *Acetobacter diazotrophicus* PAL5. *J. Bacteriol.* 1999, *181*, 6987–6995.

[26] Saravanan, V. S., Madhaiyan, M., Thangaraju, M., Solubilization of zinc compounds by the diazotrophic, plant growth promoting bacterium *Gluconacetobacter diazotrophicus*. *Chemosphere* 2007, *66*, 1794–1798.

[27] Stephan, M. P., Oliveira, M., Teixeira, K. R. S., Martinez-Drets, G., Döbereiner, J., Physiology and dinitrogen fixation of *Acetobacter diazotrophicus*. *FEMS Microbiol. Lett.* 1991, *77*, 67–72.

[28] Sevilla, M., Burris, R. H., Gunapala, N., Kennedy, C., Comparison of benefit to sugarcane plant growth and 15N2 incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus* wild-type and Nif-mutants strains. *Mol. Plant Microbe Interact.* 2001, *14*, 358–366.

[29] Boddey, R. M., Urquiaga, S., Alves, J. R. B., Reis, V., Endophytic nitrogen fixation in sugarcane: present knowledge and future applications. *Plant Soil* 2003, *252*, 139–149.

[30] Boddey, R. M., de Oliveira, O. C., Urquiaga, S., Reis, V. M. *et al.*, Biological nitrogen fixation associated with sugar cane and rice: contributions and prospects for improvement. *Plant Soil* 1995, *90*, 195–209.

[31] Baldani, J. I., Baldani, V. L., History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *An Acad Bras Cienc* 2005, *77*, 549–579.

[32] Rodrigues Neto, J., Malavolta Jr., V. A., Victor, O., Meio simples para isolamento e cultivo de *Xanthomonas campestris pv. citri* Tipo B. *Summa Phytopathol.* 1986, *12*, 16.

[33] Sambrook, J., Fritsch, E. F., Manniatis, T., *Molecular cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory Press, New York 1989, 2nd edn.

[34] Berkelman, T., Stenstedt, T., *2-D Electrophoresis using Immobilized pH Gradients: Principles and Methods*. Amersham Biosciences Piscataway, NJ 1998.

[35] Hellman, U., Wernstedt, C., Gonez, J., Heldin, C. H., Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* 1995, *224*, 451–455.

[36] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

[37] Murashige, T., Skoog, F., A revised medium for rapid growth and biossays with tobacco tissue cultures. *Physiol. Plant* 1962, *15*, 473–497.

[38] Blonder, J., Goshe, M. B., Moore, R. J., Pasa-Tolic, L., *et al.*, Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography-tandem Mass Spectrometry. *J. Proteome Res.* 2002, *1*, 351–360.

[39] Fujiki, Y., Hubbard, A. L., Fowler, S., Lazarow, P. B., Isolation of Intracellular Membranes by Means of Sodium Carbonate Treatment : Application to Endoplasmic Reticulum. *J. Cell Biol.* 1982, *93*, 97–102.

[40] Brown, R. E., Jarvis, K. L., Hyland, K. J., Protein measurement using Bicinchoninic acid: elimination of interfering substances. *Anal. Biochem.* 1989, *180*, 136–139.

[41] Coelho, A., de Oliveira Santos, E., Faria, M. L., de Carvalho, D. P. *et al.*, A proteome reference map for *Vibrio cholerae* El Tor. *Proteomics* 2004, *4*, 1491–1504.

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

[43] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[44] Pucci, M. J., Discotto, L. F., Dougherty, T. J., Cloning and identification of the *Escherichia coli* murB DNA sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine reductase. *J. Bacteriol.* 1992, *174*, 1690–1693.

[45] Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G. *et al.*, YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol. Microbiol.* 2006, *61*, 151–164.

[46] Werner, J., Misra, R., YaeT (Omp85) affects the assembly of lipid-dependent and lipid-independent outer membrane proteins of *Escherichia coli*. *Mol. Microbiol*. 2005, *57*, 1450–1459.

[47] Gilson, L., Mahanty, H. K., Kolter, R., Genetic analysis of an MDR-like export system: the secretion of colicin V. *Embo J* 1990, *9*, 3875–3884.

[48] Driessen, A. J., SecB, a molecular chaperone with two faces. *Trends Microbiol.* 2001, *9*, 193–196.

[49] Twells, R., Thomson, N., Parkhill, J., Carver, T., *Working with pathogen genomes*. The first Joint Pasteur Institute/Wellcome Trust course on genomics in South America. *Institut Pasteur* 2006.

[50] Nachin, L., El Hassouni, M., Loiseau, L., Expert, D., Barras, F., SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.* 2001, *39*, 960–972.

[51] Amemura, M., Makino, K., Shinagawa, H., Kobayashi, A., Nakata, A., Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. *J. Mol. Biol*. 1985, *184*, 241–250.

[52] Li, Y., Zhang, Y., PhoU is a persistence switch involved in persister formation and tolerance to multiple antibiotics and stresses in *Escherichia coli*. *Antimicrobial Agents Chemother.* 2007, *51*, 2092–2099.

[53] Abergel, C., Bouveret, E., Claverie, J. M., Brown, K. *et al.*, Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 A resolution. *Structure* 1999, *7*, 1291–1300.

[54] Simpson, W., Olczak, T., Genco, C. A., Characterization and expression of HmuR, a TonB-dependent hemoglobin receptor of *Porphyromonas gingivalis*. *J. Bacteriol.* 2000, *182*, 5737–5748.

[55] Duche, D., Letellier, L., Geli, V., Benedetti, H., Baty, D., Quantification of group A colicin import sites. *J. Bacteriol.* 1995, *177*, 4935–4939.

[56] Geli, V., Benedetti, H., Insertion of proteins into membranes. A survey. *Subcell. Biochem.* 1994, *22*, 21–69.

[57] Rees, D. C., Howard, J. B., Nitrogenase: standing at the crossroads. *Curr. Opin. Chem. Biol.* 2000, *4*, 559–566.

[58] Kim, J., Rees, D., Nitrogenase and biological nitrogen fixation. *Biochemistry* 1994, *33*, 389–397.

[59] Bellone, C. H., De Bellone, S. D. V. C., Pedraza, R. O., Monzon, M. A., Cell colonization and infection thread formation in sugar cane roots by *Acetobacter diazotrophicus*. *Soil Biology & Biochemistry* 1997, *29*, 965–967.

[60] Lee, S., Reth, A., Meletzus, D., Sevilla, M., Kennedy, C., Characterization of a major cluster of nif, fix, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus*. *J. Bacteriol.* 2000, *182*, 7088–7091.

[61] Pawlowski, K., Klosse, U., de Bruijn, F. J., Characterization of a novel *Azorhizobium caulinodans* ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. *Mol. Gen. Genet.* 1991, *231*, 124–138.

[62] Hartl, F. U., Molecular chaperones in cellular protein folding. *Nature* 1996, *381*, 571–579.

[63] Hendrick, J. P., Hartl, F. U., Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* 1993, *62*, 349–384.

[64] Lazar, S. W., Kolter, R., SurA assists the folding of *Escherichia coli* outer membrane proteins. *J. Bacteriol.* 1996, *178*, 1770–1773.

[65] Bardwell, J. C., Craig, E. A., Ancient heat shock gene is dispensable. *J. Bacteriol.* 1988, *170*, 2977–2983.

[66] Tanaka, N., Nakamoto, H., HtpG is essential for the thermal stress management in cyanobacteria. *FEBS Lett.* 1999, *458*, 117–123.

[67] Weber, M. H., Marahiel, M. A., Bacterial cold shock responses. *Sci. Progr.* 2003, *86*, 9–75.

[68] Ermolenko, D. N., Makhatadze, G. I., Bacterial cold-shock proteins. *Cell. Mol. Life Sci.* 2002, *59*, 1902–1913.

[69] Wood, Z. A., Schroder, E., Robin Harris, J., Poole, L. B., Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem. Sci.* 2003, *28*, 32–40.

[70] Santos, J. M., Lobo, M., Matos, A. P., De Pedro, M. A., Arraiano, C. M., The gene bolA regulates dacA (PBP5), dacC (PBP6) and ampC (AmpC), promoting normal morphology in *Escherichia coli*. *Mol. Microbiol*. 2002, *45*, 1729–1740.

[71] Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T. *et al.*, The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 1998, *396*, 133–140.

[72] Keller, M., Roxlau, A., Weng, W. M., Schmidt, M. *et al.*, Molecular analysis of the *Rhizobium meliloti* mucR gene regulating the biosynthesis of the exopolysaccharides succinoglycan and galactoglucan. *Mol- Plant Microbe Interact-* 1995, *8*, 267–277.

[73] Cheng, H. P., Walker, G. C., Succinoglycan production by *Rhizobium meliloti* is regulated through the ExoS-ChvI two-component regulatory system. *J. Bacteriol.* 1998, *180*, 20–26.

[74] Attwood, M. M., van Dijken, J. P., Pronk, J. T., Glucose metabolism and gluconic acid production by *Acetobacter diazotrophicus*. *J. Ferment. Bioeng.* 1991, *72*, 101–105.

[75] Luna, M. F., Bernardelli, C. E., Galar, M. L., Boiardi, J. L., Glucose metabolism in batch and continuous cultures of *Gluconacetobacter diazotrophicus* PAL 3. *Curr Microbiol* 2006, *52*, 163–168.

[76] Alvarez, B., Martínez-Drets, G., Metabolic characterization of *Acetobacter diazotrophicus*. *Can. J. Microbiol.* 1995, *41*, 918–924.

[77] Kersters, K., De Ley, J., The occurrence of the Entner-Doudoroff pathway in bacteria. *Antonie van Leeuwenhoek* 1968, *34*, 393–408.

[78] Prust, C., Hoffmeister, M., Liesegang, H., Wiezer, A. *et al.*, Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat. Biotechnol.* 2005, *23*, 195–200.

[79] De Ley, J., Gillis, M., Swings, J., Family VI. Acetobacteraceae. In: *Bergey's Manual of Systematic Bacteriology*, Williams & Williams, Baltimore, MD 1984, pp. 2268–2285.

[80] White, G. A., Wang, C. H., The dissimilation of glucose and gluconate by *Acetobacter xylinum*. 1. The origin and the fate of triose phosphate. *Biochem. J.* 1964, *90*, 408–423.

[81] Greenfield, S., Claus, G. W., Nonfunctional tricarboxylic acid cycle and the mechanism of glutamate biosynthesis in *Acetobacter suboxydans*. *J. Bacteriol.* 1972, *112*, 1295–1301.

[82] Sievers, M., Swings, J., Family II. *Acetobacteraceae*, in: Garrity, G. M. (Ed.), *Bergey's manual of systematic bacteriology*. 2005, *2*, 41–50.

[83] Sievers, M., Swings, J., Genus VIII. *Gluconacetobacter*, in: Garrity, G. M, (Ed.), *Bergey's Manual of Systematic Bacteriology*, Springer-Verlag, New York 2005, pp. 72–77.

[84] Gonzalez, B., Martinez, S., Chavez, J. L., Lee, S. *et al.*, Respiratory system of *Gluconacetobacter diazotrophicus* PAL5 Evidence for a cyanide-sensitive cytochrome bb and cyanide-resistant cytochrome ba quinol oxidases. *Biochim. Biophys. Acta* 2006, *1757*, 1614–1622.

[85] Tamaki, T., Fukaya, M., Takemura, H., Tayama, K. *et al.*, Cloning and sequencing of the gene cluster encoding two subunits of membrane-bound alcohol dehydrogenase from Acetobacter polyoxogenes. *Biochim. Biophys. Acta* 1991, *1088*, 292–300.

[86] Thurner, C., Vela, C., Thony-Meyer, L., Meile, L., Teuber, M., Biochemical and genetic characterization of the acetaldehyde dehydrogenase complex from *Acetobacter europaeus*. *Arch. Microbiol.* 1997, *168*, 81–91.

[87] Bergeron, H., Labbe, D., Turmel, C., Lau, P. C., Cloning, sequence and expression of a linear plasmid-based and a chromosomal homolog of chloroacetaldehyde dehydrogenase-encoding genes in *Xanthobacter autotrophicus* GJ10. *Gene* 1998, *207*, 9–18.

[88] Earl, C. D., Ronson, C. W., Ausubel, F. M., Genetic and structural analysis of the *Rhizobium meliloti* fixA, fixB, fixC, and fixX genes. *J. Bacteriol.* 1987, *169*, 1127–1136.

[89] Yamada, Y., Kondo, K., *Gluconoacetobacter*, a new subgenus comprising the acetate-oxidizing acetic acid bacteria with ubiquinone-10 in the genus *Acetobacter*. *J. Gen. Appl. Microbiol.* 1984, *30*, 297–303.

[90] Szkopinska, A., Ubiquinone. Biosynthesis of quinone ring and its isoprenoid side chain. Intracellular localization. *Acta Biochimica Polonica* 2000, *47*, 469–480.

[91] Fukaya, M., Tayama, K., Tamaki, T., Ebisuya, H. *et al.*, Characterization of a cytochrome a1 that functions as a ubiquinol oxidase in *Acetobacter aceti*. *J. Bacteriol.* 1993, *175*, 4307–4314.

