

“Unknown Genome” Proteomics

A NEW NAD(P)-DEPENDENT EPIMERASE/DEHYDRATASE REVEALED BY N-TERMINAL SEQUENCING, INVERTED PCR, AND HIGH RESOLUTION MASS SPECTROMETRY*

Diliana Dancheva Simeonova‡§, Iuliana Susnea§¶, Adrian Moise¶, Bernhard Schink‡, and Michael Przybylski¶||

We present here a new approach that enabled the identification of a new protein from a bacterial strain with unknown genomic background using a combination of inverted PCR with degenerate primers derived from N-terminal protein sequences and high resolution peptide mass determination of proteolytic digests from two-dimensional electrophoretic separation. Proteins of the sulfate-reducing bacterium *Desulfotignum phosphitoxidans* specifically induced in the presence of phosphite were separated by two-dimensional gel electrophoresis as a series of apparent soluble and membrane-bound isoforms with molecular masses of ~35 kDa. Inverted PCR based on N-terminal sequences and high resolution peptide mass fingerprinting by Fourier transform-ion cyclotron resonance mass spectrometry provided the identification of a new NAD(P) epimerase/dehydratase by specific assignment of peptide masses to a single ORF, excluding other possible ORF candidates. The protein identification was ascertained by chromatographic separation and sequencing of internal proteolytic peptides. Metal ion affinity isolation of tryptic peptides and high resolution mass spectrometry provided the identification of five phosphorylations identified in the domains 23–47 and 91–118 of the protein. In agreement with the phosphorylations identified, direct molecular weight determination of the soluble protein eluted from the two-dimensional gels by mass spectrometry provided a molecular mass of 35,400 Da, which is consistent with an average degree of three phosphorylations. *Molecular & Cellular Proteomics* 8:122–131, 2009.

The sulfate-reducing bacterium *Desulfotignum phosphitoxidans* can utilize phosphite as electron donor for growth and has been shown to induce a specific protein band of ~40 kDa in one-dimensional SDS-PAGE (1, 2). The genome of this bacterium is unknown, and no genetic information has been available concerning the genes and proteins involved in the process of phosphite oxidation. It is commonly appreciated that in the case of unknown genetic background, there is no

direct approach applicable for proper identification of a protein of interest, and protein identification by proteome analysis is normally based on availability of genomic data. Using bottom-up proteomics (3, 4) with amenable databases, identification of proteins is often straightforward but is highly complex or unfeasible in the absence of genomic data (5, 6). In such cases, suitable derivatization approaches (5) and/or “*de novo*” identification (6) is typically required.

The combination of two-dimensional gel electrophoresis (2-DE)¹ and mass spectrometry has become a powerful tool for protein identification if the genetic background is known. A standard procedure is to excise protein spots from the gel followed by in-gel digestion with a specific protease, extraction, and mass spectrometric analysis of the proteolytic peptides (7, 8). One possibility in this “bottom-up” approach involves the application of tandem mass spectrometry to a set of peptides from a protein digest (9), resulting in a series of sequence-specific fragment ions that can be used for protein identification. In the present study we investigated proteins specifically induced by the sulfate-reducing bacterium *D. phosphitoxidans* in the presence of phosphite. The particular features of bacterial genetics, such as the lack of gene splicing, enabled the combination of proteomics and genetics methods comprising (i) initial N-terminal Edman protein sequencing followed by inverted PCR of degenerate primers for N-terminal sequences and (ii) proteome analysis by high resolution mass spectrometry for protein identification from ORF candidates obtained in the first step. FTICR mass spectrometry (10) was shown in this study to be a powerful tool for unequivocal peptide identification. The analytical scheme of this “unknown genome” proteomics approach is shown in Fig. 1. Using 2-DE-isolated soluble and membrane-bound proteins expressed in the presence of phosphite, this combined approach enabled the identification of a new NAD(P)-dependent epimerase/dehydratase from *D. phosphitoxidans*. In addition, direct mass spectrometric molecular weight determinations and identification of affinity-isolated peptides provided the detection and localization of multiple phosphorylation sites.

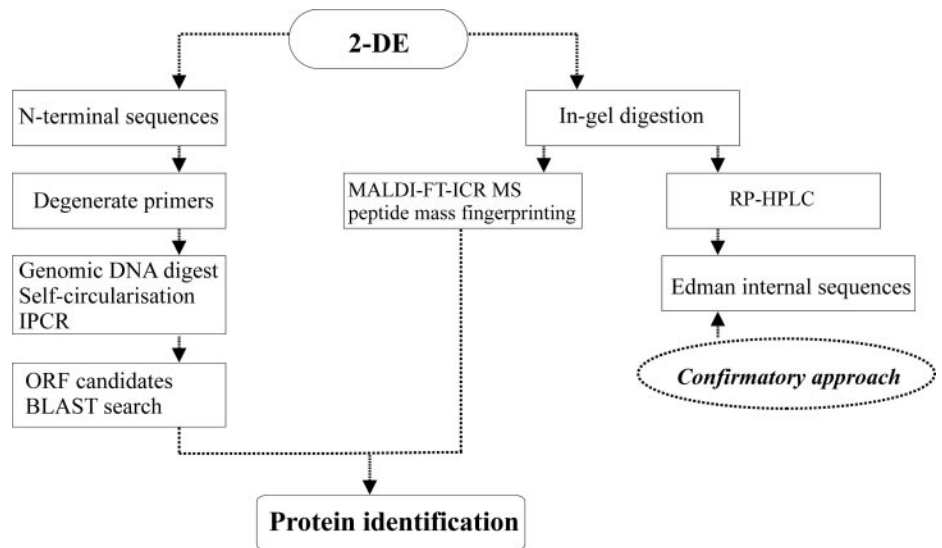
From the ‡Laboratory of Microbial Ecology, Department of Biology and ¶Laboratory of Analytical Chemistry, Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany

Received, June 2, 2008, and in revised form, July 31, 2008

Published, MCP Papers in Press, August 20, 2008, DOI 10.1074/mcp.M800242-MCP200

* The abbreviations used are: 2-DE, two dimensional gel electrophoresis; IPCR, inverted PCR; SF, soluble protein fraction; TEMED, *N,N,N',N'*-tetramethylethylenediamine; SDR, short chain dehydrogenase/reductase.

FIG. 1. Analytical scheme of the combined proteomics-genomics approach used in the identification of a new protein from *D. phosphitoxidans* by (i) N-terminal partial Edman sequencing and RT-PCR and (ii) high resolution FTICR-MS peptide mass determination of possible ORFs. BLAST, Basic Local Alignment Search Tool.



EXPERIMENTAL PROCEDURES

Materials—IPG strips (pH 5–8, 17 cm) were purchased from Bio-Rad. C₁₈ OMIX™ pipette tips were from Varian (Walnut Creek, CA). Carrier ampholytes (Servalyt 5–8) were from Serva (Heidelberg, Germany). Trizma® base, TEMED, Coomassie Brilliant Blue G and R, phosphoric acid 85% (H₃PO₄), ACN, and TFA were from Sigma. The 30% acrylamide, 0.78% N,N'-methylenebisacrylamide solution (37.5:1), DTT, and CHAPS were purchased from Genaxxon Biosciences (Biberach, Germany). Ammonium sulfate ((NH₄)₂SO₄), glycerol, glycine, SDS, TCA, and EDTA were from Roth (Karlsruhe, Germany). Iodoacetamide, methanol (CH₃OH), ammonium hydrogen carbonate (NH₄HCO₃), agarose, and bromphenol blue were from Fluka (Buchs, Switzerland). Trypsin was purchased from Promega (Madison, WI). Urea and thiourea were from Merck. 2,5-Dihydroxybenzoic acid was from Bruker Daltonics (Bremen, Germany). Millipore water (Millipore, Bedford, MA) with a minimum electric resistance at 18.2 megaohms was used for all solutions.

Preparation of Protein Samples—Cells grown in 1-liter cultures of *D. phosphitoxidans* in the presence of 10 mM sodium phosphite and/or 10 mM sodium fumarate as electron donors and 10 mM sodium sulfate as electron acceptor were harvested in the late exponential growth phase. Phosphite-induced and non-induced cells were harvested under an anoxic atmosphere (95:5 (v/v) N₂/H₂ (Coy chamber, Ann Arbor, MI)). Cells were washed with anoxic 10 mM Tris-HCl buffer, pH 7.2, containing 0.342 M NaCl and suspended in 3 ml of soluble cytoplasmic extraction reagent containing 50 μl/ml protease inhibitor mixture for bacterial cell extracts (Sigma). Cell-free extracts were prepared anoxically by passing the cells four to five times dropwise through a chilled French pressure cell at 138 megapascals. Unopened cells and cell debris were removed by centrifugation at 27,000 × g for 20 min at 4 °C (Optima™ TL ultracentrifuge, Beckman). Soluble and membrane fractions of proteins were obtained by ultracentrifugation at 57,000 × g for 1 h at 4 °C. Further both protein fractions were treated according to the ProteoPrep Universal Extraction kit (Sigma) and the manufacturer's instructions. Protein content in the preparations was determined spectrophotometrically by the bicinchoinic acid method (BCA protein assay kit, Pierce) with bovine serum albumin as a standard. The soluble and the membrane protein fractions were stored in 200-μl aliquots and separated on 2-DE gels.

Acetone Precipitation—For the 2-DE preparation of samples from the soluble fractions we used acetone precipitation for removal of salts and contaminants. The protein content was precipitated at

–28 °C for 5 h by adding 6 volumes of ice-cold acetone to the sample. After 20 min of centrifugation at 14,926 × g the residual acetone was removed, and the obtained pellet was allowed to dry.

Protein Separation by Two-dimensional Gel Electrophoresis—The samples were applied overnight on 17-cm IPG strips (pH range 5–8) using a passive in-gel rehydration method. Approximately 0.4–0.8 mg of total protein was loaded on one gel. The rehydration solution contained 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 40 mM Tris base, 2% (v/v) Servalyt 5–8, 0.3% DTT, and a trace of bromphenol blue. IEF was carried out using a Multiphor horizontal electrophoresis system (Amersham Biosciences). Rehydrated strips were run in the first dimension for about 23 kV-h at 20 °C. The proteins were focused for 30 min at 150 V, 30 min at 300 V, and 5 h at 3500 V. For the second dimension the IPG strips were equilibrated in 50 mM Tris-HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, a trace of bromphenol blue, 1% (w/v) DTT for 40 min. The second equilibration step used 4.5% (w/v) iodoacetamide instead of DTT for 20 min. In the second separation step (SDS-PAGE) the system used was the Bio-Rad Protean II xi vertical electrophoresis system. 10% SDS gels (1.5 mm thick) were used. Strips were placed on the vertical gels and overlaid with 0.5% agarose in SDS running buffer (25 mM Tris base, 192 mM glycine, 0.1% (w/v) SDS). Electrophoresis was performed in two steps: 25 mA/gel for ~30 min and 40 mA/gel until the dye front reached the anodic end of the gels. After this separation step the proteins were visualized with sensitive colloidal Coomassie staining according to Neuhoff *et al.* (11). The gels were scanned using a GS-710 calibrated imaging densitometer (Bio-Rad). For 2-DE gel comparison PDQuest analysis software (version 6) from Bio-Rad was used.

N-terminal Sequence Determination—For N-terminal sequence determinations 2-DE gels were electroblotted for 2 h onto PVDF membranes (Applied Biosystems) at 50 V using a WEB-M tank blotter (PEQLAB) with a buffer containing 25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3. Gel-blotting papers (200 × 200 mm; Whatman) were used for the blotting sandwich preparation. After transfer, the PVDF membranes were washed with water for 15 min and then with methanol for 5–10 s and incubated in the staining solution (0.1% Coomassie Brilliant Blue R-250 in 40% methanol in water) until protein spots became visible. The membranes were then destained in destaining solution (50% methanol in water, freshly prepared before use) until the background disappeared and the spots were clearly visible. The membranes were air-dried. The spots of interest were

TABLE I

N-terminal sequences determined by Edman analysis of phosphite-induced proteins of *D. phosphitoxidans*

Amino acids not determined unequivocally are shown in parentheses.

Spot ^a	Sequence	Cycles
Sp2	MKEGKVVMTGGTGFI L	20
Sp3	MKEGKVVMTGGTGFI LSQVAEKLIEMGKDVTLF (D)	35
Sp4	MKEGKVVMTGGTGFI LSQVAEKLIEMGKDVTLFDN (D) EQH (N) MK (Y) E (T)	45
Sp5	MKEGKVVMTGGTGFI LSQVAEKLIEMGKDVTLFDN (D)	38

^a Sp2 denotes spot 2, sp3 denotes spot 3, sp4 denotes spot 4, and sp5 denotes spot 5 from two-dimensional gels (Fig. 2, B, C, and D).

TABLE II

Oligonucleotides used in this study

Oligonucleotides in A were developed based on the N-terminal amino acid sequences of phosphate-induced proteins for amplification of self-circularized DNA fragments of *D. phosphitoxidans* genomic DNA. Oligonucleotides in B were used for amplification of EcoRI-digested and self-ligated DNA. *, these are not degenerated primers as compared to RPD2 and FPD2.

Primer	5' → 3'
A	
AFP (24-mer)	5'-GTTGCHGAAAACTKATYGAAATG-3'
ARP (22-mer)	5'-GTRCCRCGGTCATAACAACCTT-3'
BFP (22-mer)	5'-AACTGATYGAAATGGGYAAAGA-3'
BRP (20-mer)	5'-RCCRCGGTCATMACACTT-3'
FPD1 (21-mer)	5'-GCDGARAARCTGATYGARATG-3'
RPD1 (20-mer)	5'-RCCRCRGTCATHACHACYT-3'
RPD2 (20-mer)	5'-RTCYTTNCCCATYTCDATNA-3'
FPD2* (19-mer)	5'-TTGCCGAAAACTGATGA-3'
RPD2* (20-mer)	5'-CCACCGGTCATAACAACCTTTT-3'
B	
fex2 (24-mer)	5'-TTAGAGCCGGAAGGAGGTGATACG-3'
rex2.1 (24-mer)	5'-GTCAAAAAGGCCAACCCATAAAA-3'
fex3.2 (22-mer)	5'-GCGGATACTGAACCCGAATAC-3'
rex3.1 (18-mer)	5'-TCCGGCTGACGCTTCCCTT-3'
fex4 (22-mer)	5'-ATTGGCGGCATCGGCACAGTAA-3'
rex4 (24-mer)	5'-GTCAAAAAGGCCAACCCATAAAA-3'

TABLE III

Degenerate primers for amplification of additional genes, coding for similar proteins, used in this study

Primer	5' → 3'
3intF1 (23-mer)	5'-AAACTKATYGAAATGGGYAAAGA-3'
3intR1 (17-mer)	5'-GCAAAATANGGAATGGT-3'
3152F1 (21-mer)	5'-TCATTTTGTTCAGGGYGATAC-3'
36R1 (17-mer)	5'-GCAAAATANGGAATGGT-3'
3152F2 (18-mer)	5'-AYTTTCATTTTGTTCAGG-3'
38R1 (16-mer)	5'-GGTTTCRTTRTRTCR-3'

completely with 10 units of each of the following restriction endonucleases: EcoRI, BamHI, HindIII (Fermentas International Inc., Burlington, Canada), and MaeIII (Roche Applied Science) in separate reactions in a final volume of 20 μl. The digestion reactions were terminated by heat inactivation of the enzymes where appropriate, and the obtained fragments were self-circularized with T4 DNA ligase (Fermentas GmbH, St. Leon-Rot, Germany). 10 units of T4 ligase were used in a 100-μl total reaction volume. Self-circularization reactions were carried out at 16 °C overnight. IPCRs were performed with self-circularized fragments and primers shown in Table II, part A. The TA cloning kit (Invitrogen) was used in the first round of IPCR with degenerate primers.

Total RNAs were isolated from cultures of *D. phosphitoxidans* grown to late logarithmic phase ($A_{578} \sim 0.28-0.30$) in minimal medium containing 10 mM sulfate plus either 10 mM fumarate or 10 mM phosphite. Total RNA isolations were carried out with the RNeasy minikit (Qiagen, Valencia, CA) according to the manufacturer's instructions. For the removal of contaminating genomic DNA, the RNA preparations were on-column digested with DNase I, (DNase I, RNase-free set, Qiagen). The DNase I-treated RNA was used as a template in one-step reversed transcription assays using SuperScript II reverse transcriptase and Platinum Taq DNA polymerase (Invitrogen) according to the manufacturer's protocol. The RNA concentration in each preparation was assessed spectrophotometrically. The positive control with only genomic DNA and the negative control containing only RNA without reverse transcriptase were run under identical PCR amplification conditions. Gene-specific oligonucleotide probes for amplification of junction region ORF3-ORF4 used O34F (5'-TTTCTCGGCCAATTAATACTCTCC-3') and O34R (5'-AGCTTTTGGGTTTCTTCATACAT-3') in phosphite-induced and non-induced cells.

Proteolytic Digestion—Spots were excised manually from the gel and subjected to in-gel digestion with trypsin according to Mortz *et al.* (12). The excised gel pieces were washed with deionized water for 15 min, dehydrated by addition of 3:2 ACN/deionized water for 30 min at 25 °C, and dried in a SpeedVac centrifuge (30 min). They were destained by addition of 50 mM NH₄HCO₃ (15 min), dehydrated with 3:2 ACN/deionized water (15 min), and dried in a SpeedVac centrifuge (30 min). Freshly prepared trypsin solution (12.5 ng/μl trypsin in 50 mM NH₄HCO₃) was added and incubated at 4 °C (on ice) for 45 min and

excised and fully destained, dried, and kept at 4 °C in Eppendorf tubes until sequencing. Prior to sequencing, the destained protein spots were wetted with 100% methanol and applied into the sequencing cartridge. Sequence determinations were performed on an Applied Biosystems Model 494 Procise Sequencer attached to a Model 140C Microgradient System, a 785A Programmable Absorbance Detector, and a 610A Data Analysis System. All solvents and reagents used were of highest analytical grade purity (Applied Biosystems). For sequencing of both blotted and lyophilized samples, the corresponding standard pulsed liquid methods were used.

Design of Degenerate Primers and PCR—Primer sets (shown in Table II A) were developed from the N-terminal protein sequences using the following codon usage tables: *Desulfobacter vibrioformis* and *Desulfobacula toluolica* as closest relatives of *D. phosphitoxidans*. Additionally the *Escherichia coli* reversed translation codon table and best reverse translate, worst reverse translate, and degenerate codon tables of *E. coli* (DNASTar software, version 5.01) were used. For amplification of the EcoRI self-ligated fragment harboring the gene coding for a putative NAD(P)-dependent epimerase/dehydratase, primers shown in Table II part B were used. For the localization and identification of additional loci coding for similar proteins or isoforms of the putative NAD(P)-dependent epimerase/dehydratase, the degenerate oligonucleotide pairs shown in Table III were used. 1 μg of chromosomal DNA of *D. phosphitoxidans* was digested

then for 12 h at 37 °C in 50 mM NH₄HCO₃. After removal of the supernatant fraction, peptide extraction was performed with a solution of 3:2 ACN, 0.1% TFA in deionized water at room temperature (three steps of 1 h each). For Lys-C digestion the excised gel pieces were washed first with deionized water for 10 min and then with a solution of 25 mM Tris-HCl, pH 8.5, with 1 mM EDTA for 30 min. The gel spots were destained by addition of 25 mM Tris-HCl, pH 8.5, with 1 mM EDTA, 50% ACN (10 min). This last step was repeated until the Coomassie dye was completely removed. The gel pieces were dehydrated for 10 min by the addition of 100% ACN and dried in the SpeedVac centrifuge (1 h). The freshly prepared Lys-C solution (10 ng/μl Lys-C in 25 mM Tris-HCl, pH 8.5, with 1 mM EDTA) was added and incubated at 4 °C (on ice) for 30 min followed by overnight incubation (16–20 h) at 37 °C in 25 mM Tris-HCl, pH 8.5, with 1 mM EDTA. After transfer of the supernatant fraction into tubes containing 100 μl of deionized water and 5 μl of a solution of 50% acetonitrile, 5% TFA, peptide extraction was performed at room temperature (three steps of 10 min each). The eluates (supernatant and elution fractions that were collected in the same tube) were lyophilized to dryness.

HPLC Isolation of Peptides—Lys-C peptides obtained by in-gel digestion of protein spots were separated by analytical HPLC on a Bio-Rad system using a Vydac C₄ column (250 × 4.6-mm inner diameter, 5-μm silica, 300-Å pore size) (Vydac, Hesperia, CA). The samples were dissolved in 200 μl of 0.1% TFA (aqueous solution), and the peptides were separated using a linear gradient elution (0 min, 0% B; 5 min, 0% B; 105 min, 100% B; 110 min, 100% B; 115 min, 0% B; 120 min, 0% B) with eluent A (0.1% TFA in water) and eluent B (0.1% TFA in acetonitrile/water (80:20, v/v)). The flow rate was 1 ml/min, and the peaks were detected at 220-nm wavelength.

ZipTip Cleanup Procedure—C₁₈ OMIX pipette tips were used for purification of the protein digests. The ZipTip procedure was carried out in five steps: wetting (50% ACN in deionized water), equilibration (1% TFA) of the ZipTip pipette tip, binding of peptides and proteins to the pipette tip, washing (0.1% TFA), and elution (50% ACN in 0.1% TFA).

Mass Spectrometry—MALDI-FTICR mass spectrometric analysis of the in-gel digested proteins was performed with a Bruker APEX II FTICR instrument equipped with an actively shielded 7-tesla superconducting magnet, a cylindrical infinity ICR analyzer cell, and an external Scout 100 fully automated X-Y target stage MALDI source with pulsed collision gas (Bruker Daltonics). The pulsed nitrogen laser was operated at 337 nm, and ions were directly desorbed into a hexapole ion guide situated 1 mm from the laser target (13). Ions generated by 20 laser shots were accumulated in the hexapole for 0.5–1 s at 10 V and extracted at –10 V into the analyzer cell. A 100 mg/ml solution of 2,5-dihydroxybenzoic acid in ACN, 0.1% TFA in water (2:1) was used as matrix. 0.5 μl of matrix solution and 1 μl of sample solution were mixed on the stainless steel MALDI target and allowed to dry. Mass spectra were obtained by acquisition of 64 scans. External calibration was carried out using the monoisotopic masses of singly protonated ion signals of bovine insulin (5730.609 Da), bovine insulin B-chain oxidized (3494.651 Da), human neurotensin (1672.917 Da), human angiotensin I (1296.685 Da), human bradykinin (1060.569 Da), and human angiotensin II (1046.542 Da). Acquisition and processing of spectra were performed with XMASS software version 6.1.2 (Bruker Daltonics). For peptide mass fingerprinting monoisotopic masses of all singly charged ions from the MALDI-FTICR mass spectra (generated by XMASS, version 6.1.2) were directly used for database search using a MASCOT peptide mass fingerprinting search engine (Matrix Science) in combination with ProFound engine. Search and acceptance criteria were as follows: 10–50-ppm mass error tolerance, one missed cleavage site permitted, methionine oxidation as variable modification, other proteobacteria as taxonomy (277,231 sequences), and 3 as the minimum number for

matched peptides for protein identification. The database used was National Center for Biotechnology Information non-redundant (NCBI/nr) (July 12, 2006, 3,783,042 sequence entries, 1,304,471,729 residues), a compilation of several databases including Swiss-Prot, Protein Information Resource, Protein Research Foundation, Protein Data Bank, and GenBank™ coding sequence translations.

MALDI-TOF mass spectrometric analysis of the Lys-C digest was carried out with a Bruker Biflex™ linear TOF mass spectrometer (Bruker Daltonics) equipped with a nitrogen UV laser (337 nm), a dual channel plate detector, a 26-sample Scout source, a video system, and an XMASS data system for spectra acquisition and instrument control. A saturated solution of α-cyano-4-hydroxycinnamic acid in ACN, 0.1% trifluoroacetic acid in water (2:1, v/v) was used as the matrix. Aliquots of 0.8 μl of the sample solution and the saturated matrix solution were mixed on the stainless steel MALDI target and allowed to dry. Acquisition of spectra was carried out at an acceleration voltage of 20 kV and a detector voltage of 1.5 kV. Molecular weight analysis of intact proteins was performed by MALDI-TOF-MS by excision of spots from 2-DE gels, destaining as described under “Proteolytic Digestion,” and SpeedVac drying. Each gel spot was crushed in an Eppendorf cup and incubated with 20–50 μl of an organic solvent mixture consisting of 50% formic acid, 25% acetonitrile, 15% isopropanol, 10% water (v/v/v/v) (14). After a 20-min incubation in an ultrasonic bath at room temperature followed by centrifugation, 1 μl of the supernatant was placed on the MALDI target.

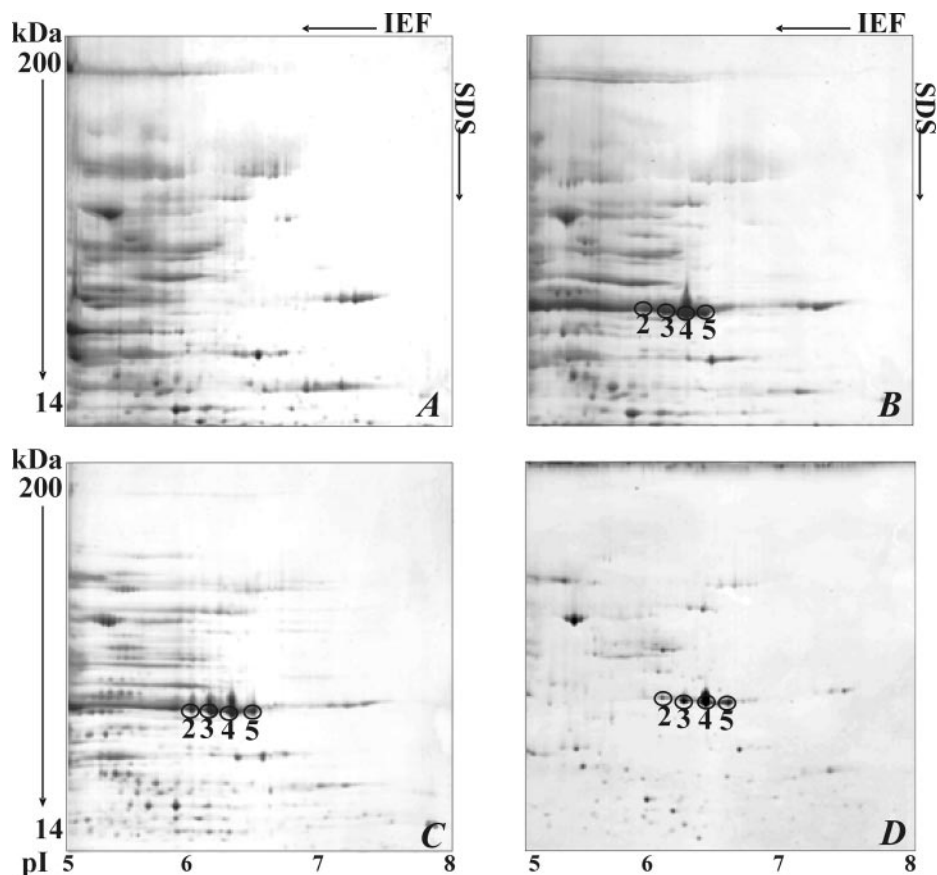
Affinity Isolation of Peptides—For analysis of phosphorylations, in-gel tryptic protein digest mixtures were purified by IMAC as described previously (15). The digestion mixtures were applied to Zip-Tip_{MC} tips (Millipore) with Ga(III) IMAC (200 mM gallium nitrate) in a solution of 0.1% acetic acid with 10% ACN at conditions suggested by the manufacturer. Affinity-bound peptides were eluted with 0.3 M ammonium hydroxide solution (2 μl) and spotted directly onto the MALDI target. Monoisotopic masses obtained by MALDI-FTICR-MS for IMAC affinity-isolated peptides were compared with the corresponding mixtures from the in-gel tryptic digestion experiments.

RESULTS

N-terminal Protein Sequencing and Design of Degenerate Primers—After two-dimensional gel electrophoretic separation of proteins of *D. phosphitoxidans*, four protein spots with a molecular mass around 40 kDa were detected that were expressed only in the presence of phosphite. To compare and assign the specific phosphite-dependent protein expression pattern, cells were grown in minimal medium supplemented with fumarate or phosphite as electron donor and sulfate or CO₂ as terminal electron acceptor in the following combinations: (a) system I, fumarate/sulfate; (b) system II, phosphite/CO₂; (c) system III, phosphite/sulfate; and (d) system IV, fumarate/phosphite/sulfate (Fig. 2). For all growth conditions, proteins were separated in two fractions, membrane protein fraction and soluble protein fraction (SF). Gels were scanned and analyzed using the PDQuest software (Bio-Rad). The comparison of the 2-DE maps of the four different growth systems for the soluble fractions is shown in Fig. 2. Spots 2, 3, 4, and 5 were clearly assigned to be expressed only in the presence of phosphite.

N-terminal Edman sequence determinations were performed for the spots expressed in the presence of phosphite. Protein spots were electroblotted onto PVDF membranes using a wet transfer procedure. Membranes were stained with 0.1% Co-

FIG. 2. 2-DE of phosphite-induced proteins in the soluble fraction from *D. phosphitoxidans* cells grown with the following electron donor/acceptor pairs: fumarate/sulfate (A), phosphite/CO₂ (B), phosphite/sulfate (C), and fumarate/phosphite/sulfate (D). The encircled spots (2, 3, 4, and 5) were found to be expressed only in the presence of phosphite. The following pI values were assigned to the proteins: spot 2, 6.0; spot 3, 6.2; spot 4, 6.3; spot 5, 6.5.



massie Brilliant Blue R-250 in 40% aqueous methanol, and protein spots were excised from the membranes and after destaining were subjected to automated Edman sequencing. N-terminal sequences of spots 2, 3, 4, and 5 from the soluble fraction, system II (Table I), provided definite sequence determinations for ~40 cycles, except for spot 2, which yielded only 17 residues because of its low concentration in 2-DE. All protein spots provided identical N-terminal sequences.

Identification of the Gene Coding for a Putative NAD(P)-dependent Epimerase/Dehydratase—Degenerate primers were designed on the basis of the first 35 residues obtained for proteins from the soluble fraction of system II that were specifically expressed in the presence of phosphite (see Table I) using the codon usage tables described under “Experimental Procedures.” The amplification of digested and self-circularized DNA fragments of *D. phosphitoxidans* genomic DNA with primer pair AFP/RPD2* (*, not degenerate primer) gave a single amplification product. The amplicon of 851 nucleotides was obtained after IPCR with BamHI-digested and self-ligated fragments. The PCR product was cloned into a pCR2.1 vector, transformed in *E. coli* INV α F' cells, and sequenced. On its right end, the fragment contained 24 nucleotides coding for the sequence MKEGKVVG. Further digest of the genomic DNA with endonuclease EcoRI, self-circularization, and IPCR with primer pair fex2/rex2.1 resulted in an amplicon

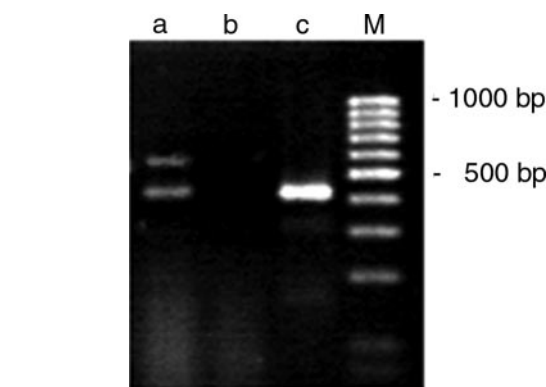


FIG. 3. RT-PCR of total RNA prepared from *D. phosphitoxidans* grown with 10 mM phosphite. Lane a shows the complete RT reaction, lane b is a negative control in which no reverse transcriptase was added into the reaction mixture, and lane c represents the PCR-positive control in which chromosomal DNA was used as a template. Lane M, 100-bp molecular weight marker.

of 3179 nucleotides in length. This product was amplified and completely sequenced with primer pairs fex3.2/rex3.1 and fex4/rex4. All nucleotide sequences formed one contig with 98% match of consensus sequence in which the locus coding for a putative NAD(P)-dependent epimerase/dehydratase was identified. The gene is 951 bp long, coding for a protein of 317 amino acids with a calculated molecular mass of 35,212 Da

ORF 4: putative Epimerase/Dehydratase-MW (Da): 35 212; pI 5.7

MKEGKVMTG GTGFILSQVA EKLIEMGKDV TLFDNNEQHN MYEETQKLE
 EKDNFHFVQG DTRDKAAVED VIKDAETVYH FAALMGTSRR FKQEVETVEV
 NVIGTLNLQ ASLDTGKYY VHPPRPPLSV MLTPYIISKG AQTQFTQMFH
KVYGLPTIGL NIQNCYGARE RAILNPNTYR PGEGRKFMAS AIIAALKNEP
 IPVFGDGEQS SDWVHIDDIV EALVLAPCDA AVGQVMDFGV GESITINKIA
 QIVIEMTKSK SKIEHLPMT GEAKVHTKAD NAPAKYLGW EPKIDLREGL
 KRTIPYFAKS LGIESPV

ORF 5: putative UhpC, sugar-phosphate sensor, MFS-type permease

MW (Da): 50 653; pI 9.5

SLIFIHSFKV GDKSKKEGLC EMKDATAKDK YPEFEGGLT DSEQVSYDRF
 RFWKVWILVS IWYSFYLR LNWMCMPPM IKDLGITKTQ AGVGATVLFW
 SYAFGLISG KLGDTYGARV MNTIGGIGTV ILNIIVASMA RINIMLIPWG
 FNGFIQQQAY APTNMITQW YPKAKRGLAT GIFATSMGVA SLEFWLITGT
 VAAHYGWRAA FTYPLLFCTL PLTILFFILA RSKPKDAGYP EYKMTNTI
 SSKAEELRDD QISGFKAWGL LFGNWKVCL AVASFMLYMG RYGLLTWVPL
 YYAETAGINL KKIPIATIAL PLGMAVGPII AGWISDRFFK AKRYQILTIY
 MLAFTTIMII LASAGLKTG LPLSFFLLAI GGFVGLSVG LVFTTACDFG
 GRRMAGTAVG SVNFFNYMGA GTQGFVIGMI LDATKNWGIV FGLLSGCAIL
 GIILVNIVRE

ORF 8: putative anaerobic Mg-protoporphyrin IX monomethylester Cyclase

MW (Da): 54 780; pI 7.8

RLLSNQFGDH SLMKVTLINP PWQFYSPSKL YPLGVSYLAS HLIHCGFDSI
 DILDNLFEIS DTQDVLKKS EIVESRNANI LGITCWTVHL LFCIEFVRVY
 KRKHPDVKII MGGVHASSQP EEIMKLCVPD IVVRGEGEET IVSLLRAIEK
 NNDLNEVHGI SYRKDKERH TPDRPLIKNL DEIPYPAYHL LEPIEKYQPL
 NRKYVSLIA SRGCPYKCIF CSANRLWKYQ RRRNPKNILG EIKWLKDNID
 VGFRLRFEDD LTNKRDWALE FFVSLKELSI HFDCLTRIXK VDSELLAMR
 DAGCIGVYHG IESASPRLLK LLRKGFPKWW NIEYIKKTI AEEVSLGLVPT
 VSAMIGIPTE TKKEIQLTFD LVCELKRLGA KTQLWLMPY PDTEAVSLYK
 EELIMDRWK ELKQFDVFST VPREAYSHLI KKYSYLVPDN WIFRNDINDE
 REMKMLYLSG ASRILGELEF V

FIG. 4. High resolution mass spectrometric peptide assignments and identification of partial sequences of a new NAD-dependent epimerase/dehydratase by comparison of different ORF candidates. Proteolytic peptides of spot 3 from SFII determined by MALDI-FTICR-MS (see Fig. 5) were compared with the calculated masses for fragments of translation products for ORF4, ORF5, and ORF8. Identified sequences and cleavage sites are shown in bold and underlined. MFS, major facilitator family.

and pI of 5.7 (Figs. 3 and 4). The genomic DNA regions upstream and downstream of the gene coding for the putative NAD(P)-dependent epimerase/dehydratase were amplified by combination of IPCR and nested PCR (data not shown). The translated sequences of the ORFs coded in these regions were compared with the MALDI-FTICR mass spectrometric determination of peptides and provided the unambiguous identification of a single protein (ORF4; see Figs. 4 and 5).

The protein sequence was checked for similarity and conserved domains with BlastP and revealed highest similarity to

the UDP-glucose 4-epimerase gb|AAM0406.1 of *Methanosarcina acetivorans* C2A with 31% identity of residues and 53% similarity. Verification of the sequence for putative signal peptides with the SignalP software (version 3.0) did not yield a result. The nucleotide sequence of the new gene was assigned as a putative epimerase/dehydratase and deposited in the GenBank database under accession number ABU54327.

Results from total RNA RT-PCR obtained from cells grown with and without phosphite are shown in Fig. 3. The oligonucleotides used were specific for the junction between the

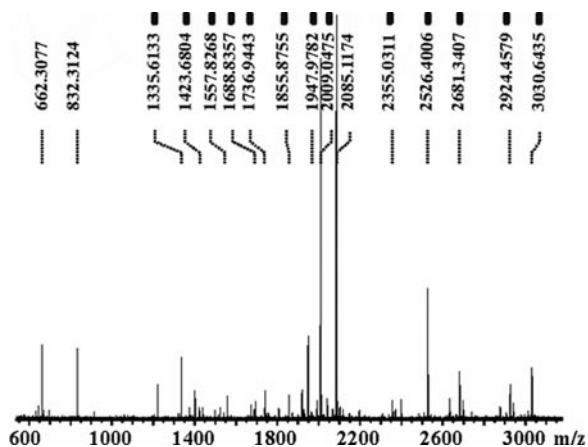


FIG. 5. MALDI-FTICR mass spectrum of the tryptic in-gel digested peptide mixture from spot 3 from system II (phosphite/ CO_2), soluble fraction. Monoisotopic masses of singly charged ions corresponding to the identified peptide sequences are shown.

gene coding for the putative NAD(P)-dependent epimerase/dehydratase and the previous ORF. Primers were designed to yield an amplification product of ~ 420 bp of which ~ 250 bp were in the intergenic region between the prior ORF and the gene of interest. This result is in agreement with the calculated length of the amplification product of 399 bp, suggesting that the mRNA spans the junction between two genes.

To detect similar loci in the genome of *D. phosphitoxidans* that might be responsible for the synthesis of putative isoforms (similar proteins) to the one coding for an NAD(P)-dependent epimerase/dehydratase, two approaches were used. (i) Degenerate oligonucleotides (Table III) based on the N-terminal sequences were used as forward primers together with primers based on internal peptide sequences obtained from Edman degradation. (ii) In addition, degenerate primers were developed from the internal peptide sequences obtained from protein spot 3 of the soluble fraction. In all cases a single amplicon per reaction was obtained that after sequencing resulted in nucleotide sequences of different lengths but complete identity with the nucleotide sequence of the gene coding for an NAD(P)-dependent epimerase/dehydratase. An exception was found for the amplified products oligonucleotides 3152F2 (fraction 15, SF) and 38R1 (fraction 8, SF). With this pair one amplification product of 626 bp was obtained that was sequenced and yielded 47 of 115 amino acids identity (40%) with the C terminus of gb|EAX56319.1 (anthranilate/*para*-aminobenzoate synthase component I-like (*Candidatus Desulfococcus oleovorans* Hxd3)). This locus in *Candidatus Desulfococcus oleovorans* Hxd3 codes for a product of 741 amino acids. A further ClustalW alignment on the nucleotide and amino acid levels of the partial product obtained with 3152F2 and 38R1 and the product of the NAD(P)-epimerase/dehydratase did not show positive results. We assume that this product was formed because of mispriming of degenerate oligonucleotides.

High Resolution Mass Spectrometric Identification of a Single Protein, a New NAD(P)-dependent Epimerase/Dehydratase—The monoisotopic masses of all singly charged peptide ions from the MALDI-FTICR mass spectra of the tryptic in-gel digests of spots 2, 3, 4, and 5 (Fig. 2) from the soluble protein fraction were used directly for database search. The NCBI nr database (October 1, 2007) was used as protein database, and MASCOT peptide mass fingerprinting and ProFound were used as search engines using the following search parameters: 10–50-ppm mass tolerance, one missed cleavage site permitted, Met oxidation as variable modification, other proteobacteria as taxonomy, and a minimum number of 3 matched peptides required for protein identification. Database searches did not provide a conclusive protein identification. This result was expected because there are no genomic data available for *D. phosphitoxidans*. However, unequivocal protein identification was obtained by high resolution peptide mass analyses using MALDI-FTICR-MS and by comparison of monoisotopic masses of peptide ions for each protein spot (spots 2–5 from the soluble fraction of system II) with the calculated masses for fragment ions of the ORFs obtained from genomic DNA amplification (Figs. 4 and 5). Fig. 4 illustrates examples of the protein identification for spot 3 SFII by comparison of fragment ions for ORFs 4, 5, and 8, which provided peptide assignments only for ORF4 and thus unequivocal identification of this protein, an NAD(P)-dependent epimerase/dehydratase. No peptide mass could be assigned for ORF8, and one peptide mass was assigned for ORF5; in contrast peptides for ORF4 covered 75% of the complete amino acid sequence. These results ascertained that the protein spots were products of the gene coding for an NAD(P)-dependent epimerase/dehydratase. The MALDI-FTICR mass spectrum acquired for the tryptic digestion mixture of spot 3 from SFII is shown in Fig. 5. In the mass range 1300–3100 Da, 14 tryptic peptides (13 displayed in the mass spectrum) were identified with high mass accuracies (Δm , 1.2–11.5 ppm; considered mass threshold, 20 ppm) as tryptic peptide fragments of the new protein (Table IV).

Protein Structure Confirmation by Sequence Determination of Internal Peptides—The identification of a specific NAD(P)-dependent epimerase/dehydratase was confirmed by isolation and sequence determination of internal Lys-C peptide fragments. As an example, the HPLC analysis of Lys-C proteolytic peptides for protein spot 3 from the 2-DE gel of SFII (Fig. 2) is shown in Fig. 6. Twenty fractions eluted from the HPLC column were subjected to Edman sequence determination (see Table V). The internal peptide sequences obtained by Edman sequence were aligned with the translated amino acid sequence of the ORF4 gene (Table V) and unequivocally confirmed the identification of the NAD(P)-dependent epimerase/dehydratase sequence.

Identification and Localization of Phosphorylations—To identify possible phosphorylation(s) of the new protein, high resolution MALDI-FTICR-MS determinations upon metal ion

TABLE IV

MALDI-FTICR-MS analysis of tryptic peptides from spot 3 of SFII

The experimental (exp) *m/z* values of tryptic fragments and calculated (calc) *m/z* values from the ORF4 translation product are compared together with mass accuracies (ppm).

Identified peptide	<i>m/z</i> _{calc}	<i>m/z</i> _{exp}	Δm
			ppm
Asp ⁵³ -Arg ⁶³	1335.6076	1335.6133	4.3
Gly ¹⁴⁰ -Lys ¹⁵¹	1423.6787	1423.6804	1.2
Ala ¹⁷² -Arg ¹⁸⁵	1557.8132	1557.8268	8.7
Ala ²⁷⁹ -Lys ²⁹³	1688.8279	1688.8357	4.6
Val ⁶ -Lys ²²	1736.9251	1736.9443	11
Asp ⁷⁴ -Arg ⁹⁰	1855.8643	1855.8755	6
Leu ⁴⁸ -Arg ⁶³	1947.9559	1947.9782	11.5
Val ¹⁵² -Arg ¹⁶⁹	2009.0273	2009.0475	10.1
Asp ²⁹ -Lys ⁴⁷	2355.0194	2355.0311	5
Tyr ¹¹⁹ -Lys ¹³⁹	2526.3908	2526.4006	3.9
Ala ⁶⁶ -Arg ⁹⁰	2681.3239	2681.3407	6.3
Asp ⁶⁴ -Arg ⁹⁰	2924.4458	2924.4579	4.1
Leu ²³ -Lys ⁴⁷	3026.3870	3026.3767	3.4
Phe ⁹¹ -Lys ¹¹⁸	3030.6357	3030.6435	2.6

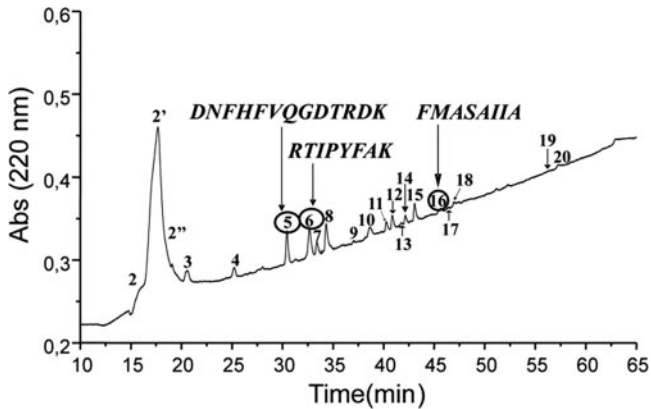


FIG. 6. HPLC separation of peptide fragments after in-gel digestion of spot 3 from SFII with Lys-C protease. Sequences identified by Edman analysis are shown for fractions 5, 6, and 16. Abs, absorbance.

affinity enrichment (IMAC) of tryptic peptides was performed. Analyses of tryptic digestion mixtures from protein spots 3 and 4 of SFII with and without IMAC enrichment revealed the identification of two phosphorylated peptides in spots 3 and 4 from SFII. In both spots four phosphorylation sites were identified in the peptide fragment Phe⁹¹-Lys¹¹⁸, corresponding to phosphorylations at the residues Thr⁹⁷, Thr¹⁰⁵, Ser¹¹², and Thr¹¹⁵. In addition, a single phosphorylation was identified for spot 3 in the peptide Leu²³-Lys⁴⁷; this sequence contains three possible phosphorylation sites (Thr³¹, Tyr⁴², and Thr⁴⁵). The increasing number of phosphorylations identified corresponded with the more acidic pI values for the spots separated by 2-DE (see Fig. 2). Thus, in protein spot 3 (pI 6.2) of the soluble fraction of system II, five phosphorylations were identified, whereas four phosphorylations were found in spot 4 (pI 6.3) (Table VI). The pI values for the phosphorylated proteins were significantly higher compared with the pI calculated for the unphosphorylated protein (5.8); the structural basis for this difference is unclear at present.

The identification of phosphorylations was further ascertained by molecular weight determinations of the intact proteins using MALDI-TOF-MS. MALDI-MS was performed by elution of proteins from 2-DE gels and placing 1- μ l aliquots on the target as illustrated in Fig. 7 by the spectrum of protein spot 4 from SFII. A molecular mass of 35,400 Da (average mass of 1+ to 5+ charged ions) was determined that by comparison with the unmodified molecular mass of 35,212 Da is consistent with an approximate (average) degree of three phosphorylations.

DISCUSSION

Anaerobic phosphite oxidation by bacteria has been discovered only recently, and *D. phosphitoxidans* is the only bacterium known thus far to oxidize phosphite to phosphate and gain energy from the oxidation process. We present here the first protein and its gene involved in this process. The

TABLE V

Alignment of internal tryptic peptide sequences from spot 3 determined by Edman sequencing with the protein product coded by the NAD-dependent epimerase/dehydratase gene

Amino acid residues not unequivocally determined are indicated by “X.” Sp, spot.

Gene product, ORF4:	¹ MKEGKVVMTGGTGFI LSQVAEKLIEMGKDVQLFDNNETHNMYEETQKLLLEEKDNFHFVQGDTRDK ⁶⁵
SFII, Sp 3:	VVMTGGTGFI DVQLFDNNET LLEEKDNFHFVQGDTRDK
Gene product, ORF4:	⁶⁶ AAVEDVIKDAETVYHFAALMGTSSRFKQEVETVEVNVIGTLNVLQASLDTGVKYYVHPPRPLSVWLTPY ¹³⁵
SFII, Sp 3:	DAETVYHFAALMGTSS
Gene product, ORF4:	¹³⁶ IISKGAQTQFTQMFHKVYGLPTIGLNIQNCYGARERAILNPNTYRPGGRKFMASAIIAALKNEPIPVFG ²⁰⁵
SFII, Sp 3:	GAQTQXXXMFHK FMASAIIA
Gene product, ORF4:	²⁰⁶ DGEQSSDWVHIDDIVEALVLA PCDAAVGQVMDFGVGESITINKIAQIVIEMTKSKSKIEHLPMTGEAKV ²⁷⁵
SFII, Sp 3:	IAQVXEMTK SKIEHLPMTGEAK
Gene product, ORF4:	²⁷⁶ HTKADNAPAKEYLGWEPKIDLREGLKRTIPYFAKSLGIESPV ³¹⁷
SFII, Sp 3:	EYLGXEPK RTIPYFAKSLGIESPV

NAD(P)-dependent epimerase/dehydratase identified is a new protein that on the basis of its conserved domains, amino acid sequence, and conservation degree of catalytic residues is assigned as a member of the short chain dehydrogenase/reductase (SDR) family of proteins. Peptide mass fingerprinting using known databases provided no identification consistent with the presence of a new protein. In the “unknown genome proteomics” approach used (Fig. 1), proteomics and genetics techniques were successfully combined using high resolution peptide mass determinations by FTICR-MS to discriminate between different ORFs from putative genes found upstream and downstream of the gene coding for an NAD(P)-dependent epimerase/dehydratase. The MALDI-FTICR-MS data identified the only correct protein out of a series of possible products and were ascertained by Edman sequence determination of internal peptides and in complete agreement with the translated gene product. In addition, direct analysis of the new protein by MALDI-MS upon elution from 2-DE identified a molecular mass of 35,400 Da, which is in agreement with the calculated mass of the amino acid sequence and a degree of approximately three phosphorylations; a total

of five phosphorylations were identified by mass spectrometric analysis of IMAC-isolated peptides.

The first epimerase studied, the UDP-glucose 4-epimerase of *E. coli* (GALE), belongs to the superfamily of SDRs (16). This protein is a homodimer with a molecular mass of 37.3 kDa and an N-terminal nucleotide cofactor binding domain. The SDRs are a diverse group of enzymes with highly divergent sequences (15–30% sequence identity) (16), most of which are dimers or tetramers (17). The SDR superfamily members contain two characteristic signature sequences, a YXXXK motif (where X can be any residue) and a GXGXXG motif (Rossmann fold) that is usually found near the cofactor binding pocket. The proteins that adopt Rossmann fold binding of nucleotide cofactors such as NAD(P) or FAD are known to function as oxidoreductases (18). Both signature sequences were found in the new NAD(P)-dependent epimerase/dehydratase identified. A Rossmann fold, GTGFIL carrying one non-conserved substitution at the last residue, is located at the N-terminal part of the protein (residues 11–16). The second characteristic sequence found, ¹³⁵YIISK¹³⁹, fully corresponds to the YXXXK motif.

A multiple alignment (ClustalW (1.83)) at the amino acid level of the translated product and 20 randomly selected UDP-glucose 4-epimerases from NCBI (December 6, 2007), including the sequence of *gaE* (P09147) from the Swiss-Prot/TrEMBL database, revealed that two of three amino acids assigned to be important in the catalytic mechanism of the enzyme are conserved in the new protein. The Lys¹⁵³, Tyr¹⁴⁹, and Ser¹²⁴/Thr¹²⁴ residues were identified as catalytic residues of UDP-glucose 4-epimerase by site-directed mutagenesis (19, 20) (the numbering refers to the GALE monomer from *E. coli*). The residues Tyr¹⁴⁹ and Lys¹⁵³ were conserved in all sequences examined, including the new protein identified. Tyr¹⁴⁹ functions as a proton acceptor in GALE. Both residues Tyr¹⁴⁹ and Lys¹⁵³ were found to be conserved also among

TABLE VI
Identification of phosphorylated peptides by IMAC/MALDI-FTICR-MS for protein spots 3 and 4 from SFII

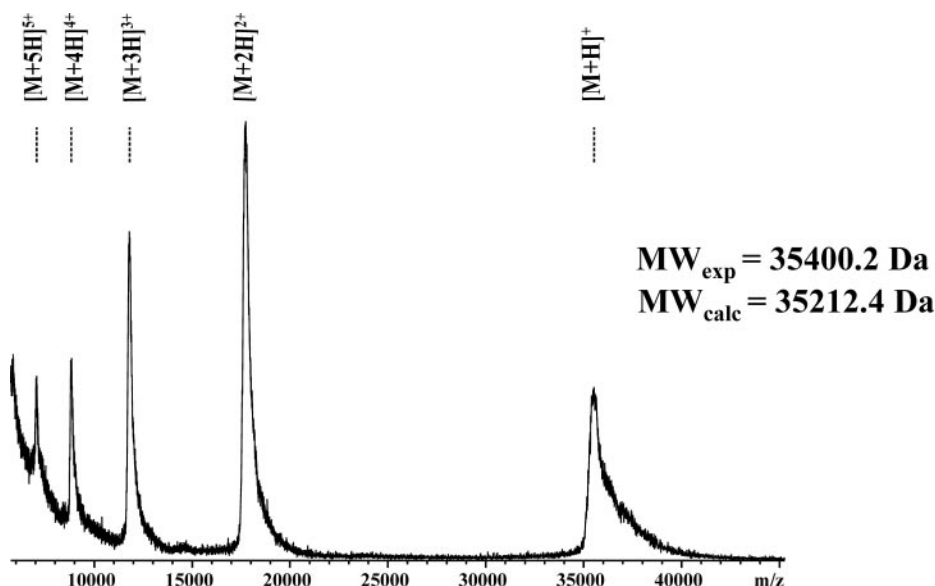
See legend to Fig. 2 for pI values of phosphorylated proteins.

Protein spot no.	IMAC-isolated peptide <i>m/z</i>		$\Delta m/P^a$	Sequence ^b
	Non-phosphorylated	Phosphorylated		
3 SFII	3026.3870	3106.3545	0.4/1	23–47
	3030.6292	3350.5021	2.3/4	91–118
4 SFII	3030.6348	3350.5029	1.0/4	91–118

^a Number of phosphorylations found.

^b ²³LIEMGKDVTLFDNNEQHNM⁴⁷YEETQK⁴⁷, one phosphorylation (P) (Thr³¹ or Tyr⁴² or Thr⁴⁵); ⁹¹FKQEVETIVEVNVIGITLNLVQLASLDIGVK¹¹⁸, four phosphorylations (Thr⁹⁷, Thr¹⁰⁵, Ser¹¹², and Thr¹¹⁵).

FIG. 7. Molecular mass determination by MALDI-TOF-MS of protein spot 4 from SFII. The average isotope molecular mass obtained was 35,400.2 Da (calculated molecular mass, 35,212.4 Da; see phosphorylations in Table VI). *calc*, calculated; *exp*, experimental.



members of the SDR superfamily. The third residue, Ser¹²⁴/Thr¹²⁴, involved in the catalytic function of UDP-glucose 4-epimerases, is not conserved, in contrast to all other aligned sequences, in the new protein, which contains an Arg residue at this position. Functionally Ser¹²⁴/Thr¹²⁴ are involved in substrate binding in all known UDP-glucose 4-epimerases. In summary, the new protein can be assigned to the SDR family of proteins based on its specific pattern sequences and the functions in which such sequences are involved. This protein is assigned as an NAD(P)-dependent epimerase/dehydratase according to its amino acid sequence. Further studies at the biochemical and protein structure level will elucidate the functional mode and the catalytic mechanism of this new enzyme and enable its detailed classification.

Acknowledgment—We gratefully acknowledge the expert assistance of Dr. Marilena Manea with the HPLC separation of the Lys-C digests.

* This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG), Bonn, Germany (Grant SI 1300/1-1, Bacterial Anaerobic Phosphite Oxidation) and in part by DFG Grants PR175-12/1 and PR175-13/1. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) ABU54327.

§ Both authors contributed equally to this work.

|| To whom correspondence should be addressed. Tel.: 49-7531-882249; Fax: 49-7531-883097; E-mail: Michael.Przybylski@uni-konstanz.de.

REFERENCES

1. Schink, B., and Friedrich, M. (2000) Phosphite oxidation by sulphate reduction. *Nature* **406**, 37
2. Schink, B., Thiemann, V., Laue, H., and Friedrich, M. W. (2002) Desulfotignum phosphitoxidans sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch. Microbiol.* **177**, 381–391
3. Hortin, G. L., Jortani, S. A., Ritchie, J. C., Jr., Valdes, R., Jr., and Chan, D. W. (2006) Proteomics: a new diagnostic frontier. *Clin. Chem.* **52**, 1218–1222
4. Resing, K. A., and Ahn, N. G. (2005) Proteomics strategies for protein identification. *FEBS Lett.* **579**, 885–889
5. Leon, I. R., Neves-Ferreira, A. G., Valente, R. H., Mota, E. M., Lenzi, H. L., and Perales, J. (2007) Improved protein identification efficiency by mass

- spectrometry using N-terminal chemical derivatization of peptides from *Angiostrongylus costaricensis*, a nematode with unknown genome. *J. Mass Spectrom.* **42**, 1363–1374
6. Waridel, P., Frank, A., Thomas, H., Surendranath, V., Sunyaev, S., Pevzner, P., and Shevchenko, A. (2007) Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. *Proteomics* **7**, 2318–2329
7. Bai, Y., Galetskiy, D., Damoc, E., Paschen, C., Liu, Z., Griese, M., Liu, S., and Przybylski, M. (2004) High resolution mass spectrometric alveolar proteomics: identification of surfactant protein SP-A and SP-D modifications in proteinosis and cystic fibrosis patients. *Proteomics* **4**, 2300–2309
8. Damoc, E., Youhnovski, N., Crettaz, D., Tissot, J. D., and Przybylski, M. (2003) High resolution proteome analysis of cryoglobulins using Fourier transform-ion cyclotron resonance mass spectrometry. *Proteomics* **3**, 1425–1433
9. Wysocki, V. H., Resing, K. A., Zhang, Q., and Cheng, G. (2005) Mass spectrometry of peptides and proteins. *Methods* **35**, 211–222
10. Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **17**, 1–35
11. Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255–262
12. Mortz, E., Vorm, O., Mann, M., and Roepstorff, P. (1994) Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol. Mass Spectrom.* **23**, 249–261
13. Baykut, G., Jertz, R., and Witt, M. (2000) Matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation. *Rapid. Commun. Mass Spectrom.* **14**, 1238–1247
14. Ehring, H., Stromberg, S., Tjernberg, A., and Noren, B. (1997) Matrix-assisted laser desorption/ionization mass spectrometry of proteins extracted directly from sodium dodecyl sulphate-polyacrylamide gels. *Rapid. Commun. Mass Spectrom.* **11**, 1867–1873
15. Posewitz, M. C., and Tempst, P. (1999) Immobilized gallium (III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71**, 2883–2892
16. Allard, S. T., Giraud, M. F., and Naismith, J. H. (2001) Epimerases: structure, function and mechanism. *CMLS Cell. Mol. Life Sci.* **58**, 1650–1665
17. Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D. (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**, 6003–6013
18. Kleiger, G., and Eisenberg, D. (2002) GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C(alpha)-H...O hydrogen bonds and van der Waals interactions. *J. Mol. Biol.* **323**, 69–76
19. Swanson, B. A., and Frey, P. A. (1993) Identification of lysine 153 as a functionally important residue in UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry* **32**, 13231–13236
20. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. *Biochemistry* **35**, 5137–5144