Journal of Proteomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

N-terminome and proteogenomic analysis of the *Methylobacterium extorquens* DM4 reference strain for dichloromethane utilization

Sabrina Bibi-Triki[a,1], Gauthier Husson[b,1], Bruno Maucourt[a], Stéphane Vuilleumier[a], Christine Carapito[b,**], Françoise Bringel[a,*]

[a] Université de Strasbourg, CNRS, GMGM UMR 7156, Department of Microbiology, Genomics and the Environment, Strasbourg, France
[b] Université de Strasbourg, CNRS, IPHC UMR 7178, Laboratoire de Spectrométrie de Masse BioOrganique, Strasbourg, France

ARTICLE INFO

Keywords:
Proteogenomics
N-terminome
Gene discovery
Genome annotation
One-carbon metabolism
Dichloromethane

ABSTRACT

Methylobacterium strains can use one-carbon compounds, such as methanol, for methylotrophic growth. In addition to methanol, a few strains also utilize dichloromethane, a major industrial chlorinated solvent pollutant. With a fully assembled and annotated genome, *M. extorquens* DM4 is the reference bacterium for aerobic dichloromethane degradation. The doublet N-terminal oriented proteomics (dN-TOP) strategy was applied to further improve its genome annotation and a differential proteomics approach was performed to compare *M. extorquens* DM4 grown either with methanol or dichloromethane as the sole source of carbon and energy. These approaches led to experimental confirmation of 259 hypothetical proteins, correction of 78 erroneous predicted start codons, discovery of 39 new proteins and annotation of 66 signal peptides, including essential enzymes involved in methylotrophic growth.

Significance: Dichloromethane (methylene chloride, CH_2Cl_2, DCM) is one of the most widely used industrial halogenated solvents and a potential carcinogen. Microbial rehabilitation of worldwide-contaminated sites involves DCM breakdown by bacteria that are able to grow using this pollutant as their sole carbon and energy source. The most-studied methylotrophic DCM degrader is *Methylobacterium extorquens* strain DM4. Proteomic studies of the *Methylobacterium* genus have been performed previously, but genome-wide investigation of N-termini of expressed proteins has not yet been performed. Differential quantitative proteomic analysis also opens new research perspectives to better monitor and understand bacterial growth with DCM.

1. Introduction

Automatic genome annotation pipelines are error-prone. Use of experimental-based protein analysis data, when available, allows for a fine curation of the overall annotation quality [1]. In particular, up to 10% estimated error rates are anticipated for translational start codon localization (see references within [2]). In this context, the development of high-throughput N-terminomic strategies offers a unique opportunity to increase the efficiency of genome curation (reviewed by [3]). Among these, the doublet N-terminal oriented proteomics (dN-TOP) strategy is based on the specific labelling of free N-terminal alpha amine function of proteins with light and heavy trimethoxyphenyl phosphonium (TMPP) [4]. This method detects both free N-terminal peptides with an acute sensitivity and internal peptides in a single experiment [5]. TMPP labelling generally allows identifying protein N-termini (N-ter) representing 5 to 8% of the total identified polypeptides. dN-TOP can curate bacterial genomes, such as *Mycobacterium smegmatis* [6], the arsenite-oxydizing bacterium *Herminiimonas arsenicoxydans* [4] and the human mitochondrial proteome [4,7,8]. N-terminomic approaches can correct erroneously predicted translational start sites, and identify proteins having their initiator methionine (iMet) cleaved or not by the N-terminal methionine excision (NME) process, an ubiquitous process involved in regulation of protein turnover [9]. It also experimentally confirms protein precursors subjected to signal peptide cleavage by signal peptidases [10] of both the Sec-type allowing the translocation of unfolded proteins across the membrane, and that of the Tat-type allowing folded protein translocation [11].

Methylobacterium constitutes an attractive candidate for methanol-based product biosyntheses [12] and for bioremediation strategies of polluted sites contaminated by chlorinated methanes [13]. Within the genus *Methylobacterium*, the plasmidless *M. extorquens* strain PA1 has emerged as the methylotroph model, notably for growth with the

* Correspondence to: F. Bringel, 28 rue Goethe, 67000 Strasbourg, France.
** Correspondence to: C. Carapito, 25 rue Becquerel, 67087 Strasbourg, France.
E-mail addresses: ccarapito@unistra.fr (C. Carapito), francoise.bringel@unistra.fr (F. Bringel).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jprot.2018.03.012
Received 26 January 2018; Received in revised form 28 February 2018; Accepted 16 March 2018

Table 1
Overview of *M. extorquens* DM4 proteogenomic analysis and discovery of new proteins.

Proteogenomic data	Number of identified CDS (discovered)		
	Chromosome 5.94 Mb, 68% GC	p1METDI 142 Kb, 65% GC	p2METDI 39 Kb, 64% GC
Previously annotated CDS[a]	5958	149	41
Differential proteomics	2079 (9)	46 (8)	11 (0)
dN-TOP[b]	2737 (37)	56 (11)	16 (0)
TMPP-labelled	742 (10)	10 (1)	4 (0)
Internal unlabelled	2700 (27)	55 (10)	16 (0)
Combined proteome coverage[c] [% of total protein identified]	2798 (39) [47%]	63 (16) [42%]	17 (0) [42%]

[a] [17].
[b] 718 proteins were identified by both TMPP-labelled peptides and internal unlabelled peptides by the dN-TOP approach.
[c] Total non-redundant CDS identified using both differential proteomics and dN-TOP approaches.

reduced one-carbon compound methanol (CH_3OH) [14]. Nonetheless, this reference strain is unable to grow with chlorinated methanes, including dichloromethane (CH_2Cl_2) [15]. On the other hand, *M. extorquens* DM4 is the best-characterized dichloromethane-degrading isolate. This strain was isolated from industrial wastewater sludge in Switzerland [16] and its high GC containing genome was manually annotated by an international consortium [17] using the Genoscope platform [18]. Its genome of 6.12 Mb is composed of a chromosome of 5.94 Mb and two plasmids p1METDI of 141 kb and p2METDI of 38 kb. It is composed of 6148 annotated CDS, with 5958 CDS encoded on the chromosome and 149 and 41 annotated CDS encoded on the p1METDI and p2METDI plasmids, respectively (Table 1). Its ability to grow with dichloromethane as sole source of carbon and energy is due to the presence of the dichloromethane dehalogenase-encoding *dcmA* gene located within a catabolic transposon highly conserved in other sequenced genomes of dichloromethane-utilizing strains, including those of different genus [13,19]. Genome-wide random mutagenesis, comparative proteomics determined by 2D-gel analysis [20], comparative transcriptomics [15] and laboratory evolution studies [21] have revealed the molecular adaptive response to dichloromethane utilization.

This study aimed to curate the genome annotation of *M. extorquens* DM4 for proteins produced in growth with dichloromethane and the non-chlorinated reference methylotrophic substrate, methanol. To discover new proteins, the mass spectrometry (MS)-datasets collected using two complementary proteomic approaches (dN-TOP and differential proteomics) were searched with several databases, including a six-frame-translated genome database. In addition to proteins involved in C1 assimilation and dissimulation, we focused on the global response to growth with dichloromethane compared to that with methanol.

2. Material and methods

2.1. Growth, harvesting of bacteria and protein extraction

M. extorquens DM4 was grown aerobically at 30 °C on a rotary shaker at 120 rpm (Multitron, Infors HT) in 1 L Erlenmeyer flasks containing 220 mL of M3Cl0 medium, as published in [22] except for the substitution of $CuCl_2$ by $CuSO_4$ in the trace element solution. The carbon and energy source dichloromethane or methanol was added at 10 mM in the sterile M3Cl0, as previously described [15]. Erlenmeyer used for growth with dichloromethane was sealed with gas-tight screw caps (Supelco mininert valve, Sigma-Aldrich). When cultures reached mid-exponential growth phase (O.D. at 600 nm of 0.15), bacteria were centrifuged 10 min at 4 °C at 6000*g* (Heraeus Multifuge X1R, Thermo Fischer Scientific) and pellets were stored at −20 °C until use. Cell pellets were re-suspended in 800 µL cold DNase, RNase, protease-free water (Sigma-Aldrich) containing anti-protease cocktail (Complete Protease Inhibitor Cocktail, Roche). Glass beads (700 mg; 0.1 mm, ThermoFischer Scientific) were added prior to mechanical disruption (RETSCH MM2 Pulverizer Mixer Mill, Hamilton Instrument). Three shaking cycles at maximum speed for 1 min interspaced by on ice incubation for 1 min, were performed. Cell lysates were centrifuged at 11,228*g* at 4 °C. Proteins were precipitated overnight at −20 °C by mixing 600 µL of supernatant with 4 volumes of ice-cold acetone, harvested by centrifugation at 14,000*g* for 30 min at 4 °C (Heraeus Multifuge X1R, Thermo Fischer Scientific) and re-suspended in derivatization buffer (50 mM Tris-HCl, 6 M urea, 2 M thiourea, 1% SDS, pH 8.2). Protein concentration was determined by RC DC Protein Assay kit (Bio-Rad) following manufacturer's instructions.

2.2. N-terminal protein derivatization

A total of 100 µg of proteins were reduced during 1 h under shaking in presence of 0.9 µL of 200 mM tributylphosphine. Proteins were then alkylated by incubating for 1 h in the dark while shaking in presence of 50 mM iodoacetamide (2.8 µL and 1.2 µL of 700 mM iodoacetamide for dichloromethane and methanol growth conditions, respectively). 4 µL of an equimolar solution of light and heavy TMPP (trimethoxyphenyl phosphonium, 100 mM, 80% acetonitrile, 200:1 reagent:protein molar ratio, 1 h shaking) were added. Light- and heavy-labelled TMPP induce mass shifts on peptides of 572.18 Da and 581.21 Da, respectively. The derivatization reaction was stopped by adding 100 mM hydroxylamine (2.8 µL and 1.3 µL of 16.3 M hydroxylamine for dichloromethane and methanol growth conditions, respectively).

2.3. SDS-PAGE

100 µg of TMPP-labelled proteins were loaded on SDS-PAGE gel and separated in 8 bands for the dN-TOP analysis, while 30 µg of unlabelled proteins were loaded on SDS-PAGE gel in technical triplicates and stacked in a single band for the differential proteomic analysis. Then, gel bands of proteins were excised, cut in small pieces, in-gel reduced (10 mM dithiothreitol in 25 mM NH_4HCO_3), alkylated (55 mM iodoacetamide in 25 mM NH_4HCO_3), and digested overnight using modified porcine trypsin (Promega) at 37 °C.

2.4. NanoLC-MS/MS analysis

NanoLC-MS/MS analyses were performed on a nanoAcquity UPLC device (Waters) coupled to a Q-Exactive Plus mass spectrometer (Thermo Scientific, Bremen, Germany). Peptides were separated on an Acquity UPLC BEH130 C18 column (250 mm × 75 µm with 1.7 µm diameter particles) and a Symmetry C18 precolumn (20 mm × 180 µm with 5 µm diameter particles; Waters). The solvent system consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Samples were loaded on the enrichment column over 3 min at 5 µL/min with 1% solvent B. Peptides were eluted at 450 nL/min with the following gradient of solvent B: for the dN-TOP analysis, from 1 to 8% over 2 min, 8 to 40% over 58 min, and 40 to 90% over 1 min; for differential proteomic analysis, from 1 to 20% over 68 min, 20 to 35% over 17 min, and 35 to 90% over 1 min. The MS capillary voltage was set to 1.8 kV at 250 °C. The system operated in a data-dependent acquisition mode with automatic switching between MS (mass range 300–1800 m/z with a resolving power of 70,000 at 200 m/z, automatic gain control fixed at 3×10^6 ions, and a maximum injection time set at 50 ms) and MS/MS (first fixed mass 100 m/z with a resolving power of 17,500 at 200 m/z, automatic gain control fixed at 1×10^5, and the maximal injection time set to 100 ms) modes. The 10 most abundant peptides (intensity threshold 2×10^5) were selected on