Identification of the Missing Links in Prokaryotic Pentose Oxidation Pathways

**EVIDENCE FOR ENZYME RECRUITMENT**

Received for publication, June 9, 2006; in revised form, July 17, 2006. Published, JBC Papers in Press, July 17, 2006, DOI 10.1074/jbc.M605549200

Stan J. J. Brouns‡, Jasper Walther‡, Ambrosius P. L. Snijders§, Harmen J. G. van de Werken‡, Hanneke L. D. M. Willemen‡, Petra Worm‡, Marjon G. J. de Vos‡, Anders Andersson¶, Magnus Lundgren‡, Hortense F. M. Mazon**, Robert H. H. van den Heuvel**, Peter Nilsson**, Laurent Salmon**, Willem M. de Vos‡, Phillip C. Wright**, Rolf Bernander‡, and John van der Oost‡

From the ‡Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Hesselink van Sichelenweg 4, 6703 CT Wageningen, the Netherlands, the §Biological and Environmental Systems Group, Department of Chemical and Process Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom, the ¶Department of Biotechnology, Royal Institute of Technology (KTH), SE-106 91 Stockholm, Sweden, the †Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala, Sweden, the **Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands, and the ‡‡Laboratoire de Chimie Bioorganique et Bioinorganique, CNRS-UMR 8182, Institut de Chimie Moléculaire et des Matériaux d’Orsay (ICMMO), Université de Paris-Sud, Bâtiment 420, F-91405 Orsay, France

The pentose metabolism of Archaea is largely unknown. Here, we have employed an integrated genomics approach including DNA microarray and proteomics analyses to elucidate the catabolic pathway for D-arabinose in *Sulfolobus solfataricus*. During growth on this sugar, a small set of genes appeared to be differentially expressed compared with growth on D-glucose. These genes were heterologously over-expressed in *Escherichia coli*, and the recombinant proteins were purified and biochemically studied. This showed that D-arabinose is oxidized to 2-oxoglutarate by the consecutive action of a number of previously uncharacterized enzymes, including a D-arabinose dehydrogenase, a D-arabinonate dehydratase, a novel 2-keto-3-deoxy-D-arabinonate dehydratase, and a 2,5-dioxopentanoate dehydrogenase. Promoter analysis of these genes revealed a palindromic sequence upstream of the TATA box, which is likely to be involved in their concerted transcriptional control. Integration of the obtained genomic data with genomic context analysis strongly suggests the occurrence of pentose oxidation pathways in both Archaea and Bacteria, and predicts the involvement of additional enzyme components. Moreover, it revealed striking genetic similarities between the catabolic pathways for pentoses, hexaric acids, and hydroxyproline degradation, which support the theory of metabolic pathway genesis by enzyme recruitment.

Pentose sugars are a ubiquitous class of carbohydrates with diverse biological functions. Ribose and deoxyribose are major constituents of nucleic acids, whereas arabinose and xylose are building blocks of several plant cell wall polysaccharides. Many prokaryotes, as well as yeasts and fungi, are able to degrade these polysaccharides, and use the released five-carbon sugars as a sole carbon and energy source. At present, three main catabolic pathways have been described for pentoses. The first is present in Bacteria and uses isomerases, kinases, and epimerases to convert D- and L-arabinose (Ara) and D-xylose (Xyl) into dihydroxyacetone phosphate (DHAP) (Fig. 1A), which is further metabolized by the enzymes of the phosphoketolase or pentose phosphate pathway. The genes encoding the pentose-converting enzymes are often located in gene clusters in bacterial genomes, for example, the *araBAD* operon for L-Ara (1), the *xylAB* operon for D-Xyl (2), and the *darK-fucP* gene cluster for D-Ara (3). The second catabolic pathway for pentoses converts D-Xyl into D-xylulose 5-phosphate as well, but the conversions are catalyzed by reductases and dehydrogenases instead of isomerases and epimerases (Fig. 1B). This pathway is commonly found in yeasts, fungi, mammals, and plants, but also in some bacteria (4–6). In a third pathway, pentoses such as L-Ara, D-Xyl, D-ribose, and D-Ara are metabolized non-phosphorylatively to either 2-oxoglutarate (2-OG)4 or to pyruvate and glycolaldehyde (Fig. 1C). The conversion to 2-OG, which is a tricarboxylic acid cycle intermediate, proceeds via the subsequent action of a pentose dehydrogenase, a pentonolactonase, a pentonic acid

---

*This work was supported by a grant from the European Union in the framework of the SCREEN project contract QLK3-CT-2000-00649 and a dedicated functional genomics grant from the Swedish Research Council (to R. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**The on-line version of this article (available at http://www.jbc.org) contains Supplemental Material which can be found at:**

http://www.jbc.org/cgi/content/full/M605549200/DC1

**The abbreviations used are:** 2-OG, 2-oxoglutarate; HSCFE, heat-stable cell-free extract; CFE, cell-free extract; Ara, arabinose; KDA, 2-keto-3-deoxy-arabinonate; DOP, 2,5-dioxopentanoate; ARA-DH, arabinose dehydrogenase; ARA-D, arabinose dehydrogenase; KDA, 2-keto-3-deoxy-D-arabinonate dehydratase; DOPDH, 2,5-dioxopentanoate dehydrogenase.
Missing Links in Prokaryotic Pentose Oxidation Pathways

Synthesis of Organic Compounds

D-Arabinonate was synthesized from D-Ara as previously described (26). The aldehyde 2,5-dioxopentanoate was synthesized from 1,4-dibromobutane according to reported procedures (27–29).

Gene Cloning and Protein Overexpression

The genes araDH (Sso1300), araD (Sso3124), kdaD (Sso3118), and dopDH (Sso3117) were amplified by PCR from genomic DNA using Pfu TURBO polymerase (Stratagene) and cloned in expression vector pET24d (Novagen) (supplemental materials Table 1). The resulting plasmids were harvested from Escherichia coli HB101 transformants by Miniprep (Qiagen), sequenced by Westburg genomics (Leusden, Netherlands), and transformed to E. coli expression strain BL21(DE3) containing the tRNA accessory plasmid pRIL (Stratagene).

All proteins were produced according to standard procedures in four 1-liter shaker flasks containing LB medium, but with some exceptions. When the culture A660 reached 0.5, the cultures were cold-shocked by placing them on ice for 30 min to induce host chaperones (20). After that, the expression was started by adding 0.5 mm isopropyl β-d-thiogalactopyranoside, and the cultures were incubated for 12–16 h at 37 °C after which they were spun down (10 min, 5000 × g, 4 °C). At the time of induction, the arabinose dehydrogenase (AraDH) and AraD overexpression cultures were supplemented with 0.25 mm ZnSO4 (30) and 20 mm MgCl2, respectively.

Protein Purification

Pelleted E. coli and S. solfataricus cells were resuspended in buffer and disrupted by sonication at 0 °C. Afterward, insoluble cell material was spun down (30 min, 26,500 × g, 4 °C) and the E. coli supernatants were subjected to heat treatment for 30 min at 75 °C. Denatured proteins were removed by centrifugation (30 min, 26,500 × g, 4 °C) yielding the heat-stable cell-free extract (HSCFE).

AraDH—HSCFE in 20 mM Tris-HCl (pH 7.5) supplemented with 50 mM NaCl was applied to a 20-ml Matrix Red A affinity column (Amicon). After washing the bound protein with 2 column volumes of buffer, the recombinant protein was eluted by a linear gradient of 2 mM NaCl.

AraD—HSCFE in 50 mM HEPES-KOH (pH 8.0) supplemented with 50 mM NaCl was applied to a 70-ml Q-Sepharose Fast Flow (Amersham Biosciences) anion exchange column, and eluted in a 2 mM NaCl gradient. Fractions containing the recombinant protein were pooled, concentrated with a 30-kDa cut-off filter (Vivaspin), and purified by size exclusion chromatography using a Superdex 200 HR 10/30 column (Amersham Biosciences) and 50 mM HEPES-KOH buffer (pH 8.0) supplemented with 100 mM NaCl as an eluent.

2-Keto-3-deoxy-D-arabinonate Dehydratase (KdaD)—HSCFE in 25 mM NaP, buffer (pH 6.8) was applied to a 70-ml Q-Sepharose Fast Flow (Amersham Biosciences) anion exchange column. Flow-through fractions containing KdaD were collected, loaded onto a 46-ml Bio-Gel hydroxyapatite column (Bio-Rad), and eluted by a linear gradient of 0.5 mM NaP, buffer (pH 6.8). Fractions containing the recombinant proteins were pooled,
and dialyzed overnight in 50 mM HEPES-KOH (pH 8.0) supplemented with 0.5 mM dithiothreitol (DTT). 2,5-Dioxopentanoate Dehydrogenase (DopDH)—HSCFE in 20 mM HEPES-KOH (pH 8.0) supplemented with 200 mM NaCl and 7.5 mM DTT was purified by affinity chromatography, as described for AraDH. Fractions containing the protein were pooled, concentrated using a 30-kDa cut-off membrane (Vivaspin), and purified by size exclusion chromatography as described for AraD.

**Enzyme Assays**

Unless stated otherwise, all enzymatic assays were performed in degassed 100 mM HEPES-KOH buffer (pH 7.5) at 70 °C. The optimal pH of catalysis was determined using a 30 mM citrate-phosphate-glycine buffer system that was adjusted in the range of pH 3–11 at 70 °C. Thermal inactivation assays were performed by incubating 50 μg/ml of enzyme at 70, 80, 85, and 90 °C and drawing aliquots at regular intervals during 2 h followed by a standard activity assay.

**Dehydrogenase Assays**

Sugar dehydrogenase activity was determined on a Hitachi U-1500 spectrophotometer in a continuous assay using 10 mM D- and L-arabinose, D- and L-xylose, D-ribose, D-lyxose, D- and L-fucose, D- and L-galactose, D-mannose, and D-glucose as a substrate, and 0.4 mM NAD+ or NADP+ as a cofactor. Aldehyde dehydrogenase reactions were performed using 5 mM 2,5-dioxopentanoate, glycolaldehyde, DL-glyceraldehyde, DL-glyceraldehyde, acetalddehyde, and propionaldehyde in the presence of 10 mM DTT. Initial enzymatic activity rates were obtained from the increase in absorption at 340 nm (A340), and calculated using a molar extinction coefficient of 6.22 mM−1 cm−1.

**Dehydratase Assay**

Standard reactions were performed using 10 mM potassium D-arabinonate in the presence of 1 mM MgCl2. The formation of 2-keto-3-deoxy-acid reaction products was determined with the thiobarbiturate assay at 549 nm using a molar extinction coefficient of 67.8 mM−1 cm−1 (31, 32). The effect of different divalent cations on enzymatic activity was investigated by a pre-treatment of the enzyme with 1 mM EDTA for 20 min at 70 °C, followed by a standard assay in the presence of 2 mM divalent metal ions.

**Formation of 2-Oxoglutarate and Pyruvate**

Enzyme reactions were performed with cell-free extract (CFE) from *S. solfataricus* cultures grown on either D-Ara or D-Glu, which were harvested at mid-exponential phase. The reaction was started by adding 25 μl of 3.5 mg/ml CFE to a mixture containing 10 mM potassium D-arabinonate, 1 mM MgCl2, and either 0.4 mM NAD+ or NADP+. After an incubation of 2 h at 75 °C, the reactions were stopped by placing the tubes on ice. Identical reactions were set up in which the CFE was replaced by the purified enzymes AraD (4.2 μg), KdaD (13.4 μg), and DopDH (3.8 μg). The amount of 2-oxoglutarate in these mixtures was then determined by the reductive amination of 2-oxoglutarate to L-glutamate using purified recombinant *Pyrococcus furiosus* glutamate dehydrogenase at 60 °C (33). The detection reaction was started by the addition of 5 units of glutamate dehydrogenase to a sample that was supplemented with 10 mM NH4Cl and 0.12 mM NADPH. The formation of pyruvate was determined at 30 °C using 4 units of chicken heart lactate dehydrogenase and 0.1 mM NADH. The conversion of 2-oxoglutarate or pyruvate was continuously monitored on a Hitachi U-1500 spectrophotometer by following the decrease in A340 until substrate depletion occurred. Changes in concentrations of NAD(P)H were calculated as described above.

**Determination of the Protein Oligomeric State**

The oligomerization state of AraDH, AraD, KdaD, and DopDH was determined by nanoflow electrospray ionization mass spectrometry. For this, the protein was concentrated in the range of 5–15 μM and the buffer was exchanged to 50 or 200 mM ammonium acetate (pH 6.7 or 7.5) by using an Ultrafree 0.5-ml centrifugal filter device with a 5-kDa cut-off (Millipore). Protein samples were introduced into the nanoflow electrospray ionization source of a Micromass LCT mass spectrometer (Waters), which was modified for high mass operation and set in positive ion mode. Mass spectrometry experiments were performed under conditions of increased pressure in the interface region between the sample and extraction cone of 8 mbar by reducing the pumping capacity of the rotary pump (34, 35). Capillary and sample cone voltages were optimized for the different proteins and were in the range of 1.4–1.6 kV and 75–150 V, respectively.

**Bioinformatic Analyses**

Upstream sequences of the differentially expressed genes were extracted between –200 and +50 nucleotides relative to the open reading frame translation start site. These sequences were analyzed using the Gibbs Recursive Sampler algorithm (36). Possible sequence motifs were checked against all upstream sequences and the complete genome of *S. solfataricus*. A diagram of the sequence motif was created using the WebLogo server.

Protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) and analyzed using PSI-BLAST on the non-redundant data base, and RPS-BLAST on the conserved domain data base. Multiple sequence alignments were built using either ClustalX or TCOffee software. Gene neighborhood analyses were performed using various webserver tools: STRING at the EMBL, Gene Ortholog Neighborhoods at the Integrated Microbial Genomes server of the Joint Genome Institute, and pinned regions at the ERGO bioinformatics suite.

**RESULTS AND DISCUSSION**

*S. solfataricus* is a model archaeon for studying metabolism and information processing systems, such as transcription, translation, and DNA replication (37, 38). Several halophilic and thermophilic Archaea have been reported to assimilate pentose sugars, but neither the catabolic pathways for these 5-carbon sugars nor the majority of its enzymes are known (17, 19). To close this knowledge gap, we have studied the growth of *S. solfataricus* on the pentose D-Ara using a multidisciplinary approach.
Missing Links in Prokaryotic Pentose Oxidation Pathways

Genomics approach, and compared the results to growth on the hexose D-Glu. Both culture media supported growth to cell densities of ~2 × 10^6 cells/ml (A_{600} 2.5) with similar doubling times of around 6 h.

Several enzyme activity assays were performed with CFEs from both cultures to establish a mode of D-Ara degradation (Fig. 1). A 12.3-fold higher NADP^+ dependent D-Ara dehydrogenase activity (45.5 milliunits/mg) was detected in D-Ara CFE (Table 1), which indicated the presence of an inducible D-Ara dehydrogenase. D-Ara reductase, D-arabinitol dehydrogenase, and D-Ara isomerase activity were not detected. Activity assays using D-arabinonate indicated that D-Ara CFE contained a 13.9-fold higher D-arabinonate dehydratase activity (7.4 milliunits/mg) than D-Glu CFE (Table 1). Moreover, the multistep conversion of D-arabinonate to 2-OG could readily be demonstrated with D-Ara CFE in the presence of NADP^+ (Fig. 2). The formation of pyruvate as one of the products from D-arabinonate was not observed, whereas control reactions with both CFEs and D-glucanate as a substrate did yield pyruvate (data not shown), indicating that the enzymes of the Entner-Doudoroff pathway were operative. In the final step of the pathway, D-Ara CFE contained a 3.6-fold higher activity (255 milliunits/mg) toward the aldehyde 2,5-dioxopentanoate (DOP) using NADP^+ as a cofactor. The data suggest that S. solfataricus employs an inducible enzyme set that converts D-Ara into the tricarboxylic acid cycle intermediate 2-OG via the pentose oxidation pathway (Fig. 1C).

Transcriptomics—The global transcriptional response of S. solfataricus growing exponentially on D-Ara or D-Glu was determined by DNA microarray analysis. The transcriptome comparison between both growth conditions showed that a small set of genes was differentially expressed 3-fold or more (Table 2). The highly expressed genes under D-Ara conditions included all four subunits of the Ara ABC transporter (Sso3066–3069) (39), a putative sugar permease for D-Ara (Sso2718), five of six subunits of the SoxM quinol oxidase complex (Sso2968–2973) (40), and five metabolic genes with general function predictions only (Sso1300, Sso3124, Sso3117, Sso3118, and Sso1303). The differential expression of the gene for the remaining SoxM subunit, i.e. the

TABLE 1

Properties of the D-Ara degrading enzymes of S. solfataricus

| Description | AraDH | AraD | KdaD | DopDH |
|-------------|-------|------|------|-------|
| EC number  | 1.1.1.117 | 4.2.1.5 | 4.2.1.- | 1.2.1.26 |
| Locus ID   | Sso1300 | Sso3124 | Sso3118 | Sso3117 |
| Uniprot ID | Q97YM2 | Q97U96 | Q97UA0 | Q97UA1 |
| COG        | 1,064 | 4,948 | 3,970 | 1,012 |
| PFAM       | 00107 | 01188 | 01557 | 00171 |
| Specific activity in S.s extracts (milliunits/mg) | 45.5 | 7.4 | ND* | 255 |
| Fold A/G   | (12.3) | (13.9) | ND* | (3.6) |
| ARA-box present | Yes | Yes | Yes | Yes |
| Subunit size (kDa) | 37.3 | 42.4 | 33.1 | 52.3 |
| Oligomerization | Tetramer | Octamer | Tetramer | Tetramer |
| Substrate specificity, turnover number (s^-1) | D-Arabinose (23.8) | D-Arabinonate (1.8) | ND | 2,5-Dioxopentanoate (8.6) |
| Cofactor | NADP^+ , Zn²⁺ | Mg²⁺ | ND | NADP^+ |
| Apparent pH optimum (>50% activity) | 8.2 (7.3–9.3) | 6.7 (5.2–10.2) | ND | 7.8 (6.7–8.2) |
| Apparent T optimum (°C) (>50% activity) | 91 (74–100) | 85 (75–92) | ND | ND |
| Thermal inactivation | 42 min at 85 °C | 18 min at 85 °C | ND | ND |
| Half-life time | 26 min at 90 °C | <10 min at 90 °C | | |

* ND, not determined.

FIGURE 1. Schematic representation of three types of pentose degrading pathways (A, B, and C). Arrows with an open or closed arrowtail represent enzymatic steps that are performed by unknown proteins or known proteins, respectively. Abbreviations: Ara, arabinose; Xyl, xylose; Rib, ribose; Ru, ribulose; Xu, xylulose; Ai, arabinitol; Xi, xylitol; Al, arabinonolactone; Xi, xylopyranose; Ri, ribonolactone; At, arabinonate; Xt, xylulose; Rt, ribulose; KDA, 2-keto-3-deoxy-arabonate (also called 2-oxo-4,5-dihydroxypentanoate); DOP, 2,5-dioxopentanoate (also called 2-oxoglutarate semialdehyde); GA, glycolaldehyde.
Missing Links in Prokaryotic Pentose Oxidation Pathways

sulfocyanin subunit SoxE (Sso2972), was just below the threshold level (supplemental materials Table 2). Whereas the expression of the ABC-type transport system genes had been shown to be induced in Ara media previously (39, 41), the differential expression of the SoxM gene cluster was not anticipated.

The genes that were up-regulated under δ-Glu conditions encode seven uncharacterized proteins (Sso3073, Sso3089, Sso3104, Sso1312, Sso2884, Sso3085, and Sso3100), the SoxB subunit of the SoxABCD quinol oxidase complex (Sso2657) (42), and a glutamate dehydrogenase (Sso2044) (43) (Table 2).

The Glu ABC transporter was not differentially expressed, confirming previous observations (41). The difference in gene expression of subunits SoxA (Sso2658), SoxC (Sso2656), and SoxD (Sso10828) was just below the threshold level (supplemental materials Table 2). Next to the SoxABCD genes, a small gene cluster containing the Rieske iron-sulfur cluster protein SoxL-1 (Sso2660) and Sox661 to Sox663 appeared to be expressed with a 2–3-fold difference (supplemental materials Table 2). It thus appears that under δ-Glu conditions, the SoxABCD quinol oxidase complex is preferentially used, whereas under δ-Ara conditions the SoxM-mediated terminal quinol oxidation is favored. Differential use of both oxidase complexes was recently also found in Metallosphaera sedula. Here the SoxABCD genes were expressed at high levels during growth on sulfur, whereas heterotrophic growth on yeast extract induced the production of the SoxM complex (44). Because the aeration and cell density of the δ-Ara and δ-Glu cultures was similar, the trigger for the differential expression of the two oxidase complexes in S. solfataricus is currently unknown.

Proteomics—Protein expression in the soluble proteomes of δ-Ara and δ-Glu grown S. solfataricus cells was compared using a combination of two-dimensional gel electrophoresis, stable isotope labeling, and tandem mass spectrometry. By employing

![Image](https://example.com/image.png)

**FIGURE 2.** The formation of 2-oxoglutarate from δ-arabinonate by extracts of S. solfataricus and by recombinant enzymes AraD, KdaD, and DopDH, in the presence of 0.4 mM NAD⁺ (dark gray bars) or NADP⁺ (light gray bars).

**TABLE 2**

| Locus name | Microarray log² fold (q value)α | Proteomics fold | Description | Ref. |
|------------|--------------------------------|----------------|-------------|------|
| Sso3066    | 4.02 ± 0.58 (1.1)               | >20            | Arabinose ABC transporter, arabinose-binding protein (AraS) | 39   |
| Sso3089    | 3.71 ± 0.97 (1.1)               | ND            | Arabinose ABC transporter, permease | 39   |
| Sso1300    | 3.64 ± 0.95 (1.1)               | >20            | Alcohol dehydrogenase IV δ-arabinose 1-dehydrogenase (AraDH) | This study |
| Sso3067    | 3.37 ± 0.49 (1.1)               | ND            | Arabinose ABC transporter, permease | 39   |
| Sso3069    | 2.97 ± 0.18 (2.5)               | ND            | Arabinose ABC transporter, ATP binding protein | 39   |
| Sso2968    | 2.56 ± 1.30 (1.1)               | ND            | Quinol oxidase subunit (SoxM complex), SoxL | 40   |
| Sso3124    | 2.44 ± 1.13 (1.1)               | >20            | Mandelate racemase/mucunate lactonizing enzyme δ-arabinonate dehydrogenase (AraD) | This study |
| Sso3117    | 2.39 ± 0.62 (1.1)               | >20            | Aldehyde dehydrogenase 2,5-dioxopentanoate dehydrogenase (DopDH) | This study |
| Sso2971    | 2.24 ± 1.17 (1.1)               | ND            | Quinol oxidase subunit (SoxM complex), SoxF (Rieske Fe-S protein) | 40   |
| Sso2973    | 2.10 ± 1.48 (2.6)               | ND            | Quinol oxidase subunit (SoxM complex), SoxM (I + III) | 40   |
| Sso2970    | 2.09 ± 1.53 (2.6)               | ND            | Quinol oxidase subunit (SoxM complex), SoxG (cytochrome a₅) | 40   |
| Sso2718    | 1.99 ± 1.10 (1.1)               | ND            | Putative sugar permease | 40   |
| Sso3046    | 1.89 ± 0.87 (1.1)               | ND            | Putative ABC sugar transporter, ATP-binding protein | 40   |
| Sso2969    | 1.86 ± 1.09 (1.1)               | ND            | Quinol oxidase (SoxM complex), SoxH subunit (II) | 40   |
| Sso3118    | 1.78 ± 0.45 (1.1)               | >20            | Conserved hypothetical protein '2-keto-3-deoxy-δ-arabinonate dehydrogenase (KdaD) | This study |
| Sso1303    | 1.77 ± 0.57 (1.1)               | ND            | Putative pentonate acid dehydratase | 45   |
| Sso1333    | NDE                         | 3.48 ± 0.62 | Isocitrate lyase | 45   |
| Sso0527    | NDE                         | 3.45 ± 0.72 | Phosphoglycerate kinase | 46   |
| Sso2869    | NDE                         | 3.06 ± 1.06 | Malic enzyme | 47   |

**High expression on δ-Glu**

Sso3073  -2.59 ± 0.71 (1.1) ND Putative sugar permease
Sso3089  -2.12 ± 1.12 (1.1) ND Hypothetical membrane protein component
Sso3104  -2.04 ± 0.31 (1.1) ND Hypothetical protein
Sso1312  -2.02 ± 0.52 (1.1) ND Putative ring oxidation complex/phenylacetic acid degradation rel.
Sso2884  -1.87 ± 0.37 (1.1) ND Putative 4-carboxymuconolactone decarboxylase
Sso2657  -1.77 ± 0.87 (1.1) ND Quinol oxidase (SoxABCD complex), cytochrome aa₃ subunit (SoxB) 42
Sso3085  -1.63 ± 0.86 (1.1) ND Conserved hypothetical membrane protein
Sso3100  -1.60 ± 0.88 (1.1) ND Hypothetical membrane protein component
Sso2049  -1.60 ± 0.48 (1.1) ND L-Glutamate dehydrogenase

α q value <5 indicates statistically significant differential expression.

β ND, not determined.

γ Original annotation.

δ NDE, not differentially expressed.
this strategy, five proteins were found with more than a 20-fold difference in expression level (supplemental materials Fig. 1, B–F), including the Ara-binding protein from the Ara ABC transporter (AraS, Sso3066) (39), Sso1300, Sso3124, Sso3118, and Sso3117 (Table 2). Interestingly, the difference in expression level of these genes at the protein level appeared to be more pronounced than at the mRNA level, which ranged from 3.4- to 16-fold. Three other proteins were also produced in higher amounts during growth on D-Ara, albeit only up to a 3-fold difference (Table 2). These were the isocitrate lyase (Sso1333) (45), the phosphoglycerate kinase (Sso0527) (46), and the malic enzyme (Sso2869) (47).

Promoter Motif Analysis—The promoters of the differentially expressed genes were analyzed for the occurrence of DNA sequence motifs that could play a role as cis-acting elements in the coordinated transcriptional control of these genes. The analysis indeed revealed the presence of a palindromic motif (consensus: AACATGTT) in the promoters of Sso3066 (araS), Sso1300, Sso3124, Sso3118, and Sso3117 genes (Fig. 3). This motif was designated the ARA box and it was always located upstream of the predicted TATA box with a separation of 10 bases. A conserved transcription factor B recognition element upstream of the predicted TATA box with a separation of 10 bases. A conserved transcription factor B recognition element appeared to be absent from the interspaced sequence between both boxes. Additional copies of the ARA box were identified further upstream of both Sso3066 and Sso1300. Although primer extension analysis was only performed for the araS gene (41), the promoter architecture suggests that the transcript leader of Sso1300, Sso3124, Sso3118, and Sso3117 will either be very short, or absent. This is in good agreement with the fact that a large proportion of the S. solfataricus genes is predicted to be transcribed without a leader (48). The inducibility of the araS promoter has recently been exploited in viral expression vectors that enable recombinant protein production in S. solfataricus (49).

Biochemical Characterization of the D-Ara-induced Proteins—The genes that were differentially expressed and contained an Ara box in their promoter were selected and cloned in an E. coli expression vector. The resulting proteins were over-produced, purified, and characterized to investigate their role in the metabolism of D-Ara.

**AraDH**—The putative zinc-containing, medium-chain alcohol dehydrogenase encoded by Sso1300 was efficiently produced and purified using a single step of affinity chromatography (Fig. 4). The enzyme was most active on l-fucose (6-deoxy-l-galactose) \(k_{\text{cat}} = 26.8 \text{ s}^{-1}\), followed by D-Ara \(k_{\text{cat}} = 23.8 \text{ s}^{-1}\), using preferentially NADP\(^+\) (\(K_m = 0.04 \pm 0.01 \text{ mM}\)) over NAD\(^+\) (\(K_m = 1.25 \pm 0.45 \text{ mM}\) as a cofactor. This enzyme was thus likely to account for the elevated D-Ara dehydrogenase activities in S. solfataricus CFE. AraDH could also oxidize l-galactose and the D-Ara C2-epimer D-ribose with similar rates \(k_{\text{cat}} = 17.7 \text{ s}^{-1}\) (Table 1). Enzyme activity toward other sugars remained below 7% of the highest activity. Similar substrate specificities and affinities have been found previously for mammalian and bacterial l-fucose or D-Ara dehydrogenases, although these enzymes prefer NAD\(^+\) as a cofactor (50, 51). AraDH was more than 50% active in a relatively narrow pH range from 7.3 to 9.3, with optimal catalysis proceeding at pH 8.2. The thermophilic nature of the enzyme is apparent from its optimal catalytic temperature of 91 °C. The enzyme maintained a half-life of 42 and

![FIGURE 3. Putative cis-regulatory element (ARA box) and TATA box in upstream sequences of the D-Ara-induced genes. The predicted transcription start site is indicated by +1. Transcripts are underlined (41). Coding sequences are in bold. Additional ARA boxes were found for Sso3066 at −90 to −83 and Sso1300 at −235 to −228 relative to the transcription start sites. Note: a single ARA box is present in the intergenic region between the divergently oriented genes Sso3118 and Sso3117.](image1)

![FIGURE 4. A digital photograph is shown of a Coomassie Blue-stained 8% SDS-PAGE gel that was loaded with purified recombinantly produced enzymes from the D-Ara oxidation pathway of S. solfataricus. Marker sizes are indicated in kDa.](image2)
Missing Links in Prokaryotic Pentose Oxidation Pathways

26 min at 85 and 90 °C, respectively, indicating that the enzyme is thermostable at physiological growth temperatures of *S. solfataricus*. Native mass spectrometry experiments showed that the intact recombinant AraDH has a molecular mass of 149,700 ± 24 Da. Comparing these data with the expected mass on the basis of the primary sequence (37,291 Da) clearly showed that the protein has a tetrameric structure and contains two zinc atoms per monomer. This is in good agreement with the tetrameric structure that has been reported for another alcohol dehydrogenase from *S. solfataricus* (Sso2536), which has a 33% identical protein sequence (30). This dehydrogenase, however, prefers aromatic or aliphatic alcohols as a substrate, and NAD + over NADP + as a cofactor. A structural study of AraDH is currently ongoing to explain the observed differences in substrate and cofactor selectivity.

**Arabinonate Dehydratase (AraD)**—The protein encoded by gene Sso3124 was originally annotated as a member of the mandelate racemase and muconate lactonizing enzyme family. This superfamily, which additionally comprises of aldonic acid dehydratases, is mechanistically related by their common ability to abstract α-protons from carboxylic acids (52). Production of the enzyme in *E. coli* yielded ~10% soluble recombinant protein, which was purified using anion exchange and size exclusion chromatography (Fig. 4). The enzyme was shown to catalyze the strictly Mg 2+ -dependent dehydration reaction of D-arabinonate to 2-keto-3-deoxy-D-arabinonate (KDA) (supplemental materials Fig. 2A). It is therefore conceivable that this enzyme is largely responsible for the increased levels of D-arabinonate dehydratase activity in *S. solfataricus* extracts. AraD displayed a maximum turnover rate of 1.8 s⁻¹ at a substrate concentration of 8 mM, whereas higher substrate concentrations imposed severe inhibitory effects on the enzyme (supplemental materials Fig. 2B). No activity was measured with D-glucuronate up to 20 mM. More than 50% enzyme activity was observed in a broad pH range of 5.2 to 10.2 with an optimum at pH 6.7 (Table 1). The enzyme was optimally active at 85 °C during which it maintained a half-life time of 18 min. Native mass spectrometry revealed that the protein had a molecular mass of 340,654 ± 63 Da, which corresponds well to an octameric protein assembly (expected monomeric mass is 33,143 Da). The catalytic activity of the protein was investigated by performing indirect enzyme assays using AraD with D-arabinonate as a substrate. A 50% decrease in the yield of KDA was observed when both enzymes were co-incubated in the presence of Mg 2+, but this did not result in the formation of either 2-OG or pyruvate. Given the fact that D-arabinonate is converted to 2-OG in D-Ara CFE, this enzyme was anticipated to be responsible for the dehydration of D-KDA to the aldehyde DOP. However, due to the unavailability of D-KDA, it was not possible to show this in a direct enzyme assay. We therefore employed an indirect assay using AraD, the putative D-KDA dehydratase (KdaD) and the predicted aldehyde dehydrogenase. The results of this assay are described under “DopDH.”

According to the Clusters of Orthologous Groups of proteins classification system, the putative KDA dehydratase belongs to COG3970. The catalytic domain of these proteins resembles that of the eukaryal fumaracetoacetate hydrolase; an enzyme that catalyzes the Mg 2+ - or Ca 2+ -dependent hydrolytic cleavage of fumaracetoacetate to yield fumarate and acetoacetate as the final step of phenylalanine and tyrosine degradation (53). In humans, several mutations in the fumaracetoacetate hydrolase gene will lead to hereditary tyrosinemia type I, which is mainly characterized by liver defects (54). Members of COG3970 are also homologous to the C-terminal decarboxylase domain of the bifunctional enzyme HpcE from *E. coli*, which in addition consists of an N-terminal isomerase domain (55). This enzyme is active in the homoprotocatechuate pathway of aromatic compounds and is responsible for the Mg 2+ - dependent decarboxylation of 2-oxo-5-carboxy-hept-3-ene-1,7-dioic acid to 2-hydroxy-hepta-2,4-diene-1,7-dioic acid and its subsequent isomerization to 2-oxo-hept-3-ene-1,7-dioic acid (55). Although the function of these enzyme classes is rather diverse, their structures have revealed similarities in terms of a fully conserved metal-ion binding site and a relatively conserved active site architecture. Multiple sequence alignment analysis of KdaD indicated the presence of a metal binding site consisting of Glu-143, Glu-145, and Asp-164, which may target a metal dependent activity as well. Further structural and kinetic studies of KdaD are currently ongoing.

**DopDH**—The putative aldehyde dehydrogenase encoded by Sso3117 was overproduced in *E. coli*, which resulted in the formation of ~5% soluble protein. This protein fraction was purified using affinity and size exclusion chromatography (Fig. 4). From native mass spectrometry experiments we could determine a molecular mass of 210,110 Da, which is in reasonable agreement with the expected mass of the tetramer on the basis of the primary sequence (52,290 Da). The measured mass may be somewhat higher due to the binding of small molecules to the protein oligomer. The determined oligomerization state corresponds to that of the closely related aldehyde dehydrogenase ALDH-T from *Geobacillus stearotherophilus* (56). The aldehyde dehydrogenase was tested for the activity toward different aldehydes and cofac-
The effect of the introduction of excess 2-OG resulted from sequence of ability of their genome sequences (37, 57) as well as the genome components were essential for the enzymatic conversions, and that prevented the formation of 2-OG, indicating that all components were essential for the enzymatic conversions, and that KdAD was most likely responsible for the dehydration of D-KDA to DOP.

Extensive kinetic characterization of DOPDH proved to be rather complicated, because the enzyme lost nearly all its activity within 1 day after its purification, even in the presence of high concentrations of reducing agents, such as DTT or β-mercaptoethanol. This could be due to the fact that this class of enzymes contains a catalytic cysteine residue (in DOPDH Cys-293), which can become irreversibly oxidized, leading to a total loss of enzymatic activity. A rapid inactivation was also observed with ALDH-T from *G. stearothermophilus* (56).

Central Carbohydrate Metabolism—Some central metabolic routes, such as the glycolysis, gluconeogenesis, and the tricarboxylic acid cycle have been studied extensively in *S. solfataricus*, *Sulfolobus acidocaldarius*, and other Archaea. The availability of their genome sequences (37, 57) as well as the genome sequence of *Sulfolobus tokodaii* (58), has recently allowed a reconstruction of the genes involved in these pathways (23). The effect of the introduction of excess 2-OG resulting from the D-Ara oxidative pathway led to the differential expression of only a few additional genes in these central carbon metabolic routes (Table 2; supplemental materials Fig. 3). The isocitrate lyase, the phosphoglycerate kinase, and the malic enzyme were up-regulated at the protein level under D-Ara conditions. The induction of the malic enzyme might indicate that the main proportion of 2-OG is converted to malate, which is then decarboxylated to pyruvate and acetyl-CoA, respectively, and is then fully oxidized to two molecules of CO₂ in one round of the tricarboxylic acid cycle. Although this may seem energetically unfavorable, the net difference in yield between the full degradation of one molecule of D-Glu or D-Ara to CO₂ is only one NADPH reduction equivalent in favor of D-Glu, because both degradation schemes lead to 6 reduced ferredoxins, 2 FADH₂, 2 ATP, and 6 or 5 NADPH molecules, respectively. It is therefore not surprising that the growth rates under both conditions are similar. The phosphoglycerate kinase may be indicative of increased gluconeogenic activities that are required under D-Ara conditions. The isocitrate lyase is normally operative in the glyoxylate shunt, but high production levels of the enzyme have also been observed during growth on L-glutamate compared with D-Glu (25). Oxidative deamination of L-glutamate leads to the formation of 2-OG as well, which may inhibit the isocitrate dehydrogenase activity leading to an accumulation of isocitrate. This could trigger the production of the isocitrate lyase, which can bypass this step without the loss of CO₂.

**pentose Oxidation Gene Clusters**—The comprehensive analysis of conserved gene clustering in multiple genome sequences is becoming an increasingly important tool to predict functionally or physically associated proteins in prokaryotic cells (reviewed in Ref. 59). Genomic context analysis of the genes involved in the D-Ara oxidative pathway of *S. solfataricus* showed that kdaD and dopDH gene orthologs are often located adjacent in prokaryotic genomes. This finding supports the proposed enzymatic functions of an aldehyde producing and an aldehyde oxidative activity. In addition, the analysis uncovered the presence of putative pentose oxidative gene clusters in the genomes of several aerobic proteobacteria, such as members of the genera *Burkholderia*, *Rhizobium*, *Bradyrhizobium*, *Agrobacterium*, and *Pseudomonas*. In some cases, the presence of such a gene cluster correlates well with the ability of the organism to assimilate pentoses and with enzymatic activities present in cell extracts (7–11), whereas in other cases biochemical data is not available. Nonetheless, a few of these characteristic gene clusters have been demonstrated genetically to be linked to pentose degradation. Combined with the findings in *S. solfataricus*, this allows the identification of additional enzymatic components in the pentose oxidation pathway and prediction of their enzymatic functions (Fig. 5A).

A putative operon of five genes was found in the genome of the oligotrophic α-proteobacterium *Caulobacter crescentus*, which was 2.8–11.6-fold up-regulated during growth on D-Xyl as compared with D-Glu (60). Reporter fusion constructs of the CC0823 promoter to the β-galactosidase gene (*lacZ*) from *E. coli* confirmed that this promoter is highly induced during growth on D-Xyl, and repressed on D-Glu or proteinaceous media (60, 61). Moreover, the disruption of the CC0823 gene prevented the *C. crescentus* from growth on D-Xyl as a single carbon source (61).

A second pentose degradation gene cluster involved in L-Ara uptake and utilization was found on chromosome II of the pathogenic β-proteobacterium *Burkholderia thailandensis*. This cluster consisting of nine genes was proposed to be responsible for the L-Ara degradation to 2-OG (Fig. 5A) (62). Disruption of the araA, araC, araE, and aral genes led to an L-Ara negative phenotype. Reporter gene insertions showed that araC and araE gene expression was repressed during growth in D-Glu media, and induced in L-Ara media. The transfer of the gene cluster to the related bacterium *B. pseudomallei* enabled this organism to utilize L-Ara as a sole carbon source also (62). Interestingly, an L-Ara dehydrogenase with 80% sequence identity to AraE has recently been characterized from *Azospirillum brasilense* (63); an organism that is known to degrade L-Ara to 2-OG (12). The flanking sequences of this gene revealed close homologs of the *B. thailandensis* araD and araE, which would indicate a similar gene cluster in *A. brasilense* (63).

Apart from several bacteria, putative pentose oxidation clusters are also present in some Archaea. In the halophile *Halocarcula marismortui*, a gene cluster was found on chromosome I that seems to contain all of the necessary components for D-Xyl oxidation, including a gene that has been identified as a D-Xyl dehydrogenase (19) (Fig. 5A).

**Components of the Pentose Oxidation Pathway**—Careful inspection of the different pentose oxidation gene clusters...
shows that the gene encoding the final enzymatic step, from DOP to 2-OG, is fully conserved between the different pentose oxidation gene clusters. The remaining analogous enzymatic steps that convert D-Ara, D-Xyl, or L-Ara into DOP are performed by enzymes from distinct COGs (Clusters of Orthologous Groups of proteins) (64) (Fig. 5, A and B, pentose panels). Whereas some of this variation in enzyme use may simply be explained by substrate differences, other variations may be due...
to the individual adaptation of existing enzymes with similar reaction chemistry, such as the pentose dehydrogenases.

A striking difference between the set of enzymes responsible for d-Ara degradation in *S. solfataricus* on the one hand, and the predicted sets for d-Xyl degradation in *C. crescentus* and *H. marismortui* and l-Ara degradation in *B. thailandensis* on the other hand, is the apparent absence of an up-regulated lactonase in the hyperthermophile. This enzyme is responsible for the hydrolysis of the lactone, yielding the corresponding linear pentonic acid. Such ring opening reactions are reported to proceed spontaneously at ambient temperatures, albeit at slow rates (65). Overexpressing a lactonase may therefore be advantageous at mesophilic growth temperatures, whereas at 80 °C the spontaneous reaction may well proceed rapidly enough not to be rate-limiting. The pentose oxidation gene clusters seem to be predominated by lactonases of COG3386, which are often annotated as “senescence marker protein-30 family proteins.” The genome of *S. solfataricus* contains two of these genes (*Sso2705* and *Sso3041*), but they were not differentially expressed, indicating that they are either not involved or that their basal expression level is sufficient for arabinonolactone hydrolysis. The putative xylonolactonase from *H. marismortui*, however, is homologous to metal-dependent β-lactamases belonging to COG2220, which catalyze similar lactame-ring opening reactions (66).

Other non-orthologous enzyme components of the pentose oxidation pathway include the pentonic acid dehydratases. Whereas the d-arabino-3-deoxy-l-arabinonate dehydratase from *S. solfataricus* belongs to COG4948, the same function seems to be performed by members of COG0129 that are commonly annotated as dihydroxyacid dehydratases (IlvD) or 6-phosphogluconate dehydratases (Edd) (67). A member of this family has recently been characterized from *S. solfataricus* (DHAD, *Sso3107*), which revealed a broad substrate specificity for aldonic acids (68). However, this gene was not differentially expressed according to the transcriptome or proteome analysis.

The 2-keto-3-deoxy-d-arabino-3-deoxy-d-arabinonate dehydratase (COG3970), or a member of the homologous COG0179, appears to be present in d-Ara and d-Xyl degradation gene clusters. Interestingly, in several *Burkholderia* species, and in *A. brasiliense*, this gene is replaced by a member of the dihydrodipicolinate synthase family (COG0329, *B. th araD*). Members of this family catalyze either aldolase or dehydratase reactions via a Schiff base-dependent reaction mechanism by a strictly conserved lysine residue. Interestingly, a detailed study of an l-KDA dehydratase involved in the l-Ara metabolism of *P. saccharophila* was reported a few decades ago, but unfortunately, neither the N-terminal sequence of the protein nor the gene sequence was determined (69, 70). The authors found that this enzyme was enantioselective for l-KDA (2-oxo-4(R),5-dihydroxypentanoate), and that the reaction proceeds via a Schiff-base intermediate. The enzyme activity was not affected by the presence of 1 mM EDTA, which suggests a divergent metal-ion independent reaction. It seems likely that this enzyme is encoded by homologs of the *B. thailandensis araD* gene, and that the apparent enantioselectivity of this enzyme does not allow a function in the degradation of d-Ara or d-Xyl, which results in a 2-keto-3-deoxypentonic acid with the S-configuration (Fig. 5B).

The aldehyde dehydrogenase from COG1012 is fully conserved in the pentose oxidation gene clusters (Fig. 5A). Strikingly, close homologs of this gene can also be found in hexaric acid degradation gene clusters of *Bacillus* species (*ycbC–ycbI*) (71, 72) (Fig. 5A). The same holds for a gene cluster in *Pseudomonas putida* (PP1245–PP1249) that is likely to be involved in the breakdown of l-hydroxyproline, which is a major constituent of collagen and plant cell wall proteins (73, 74) (Fig. 5B). Apparently, because the degradation of both hexaric acids and l-hydroxyproline is also known to proceed through DOP (9), the genetic information for the conversion of DOP to 2-OG has been shared between multiple metabolic pathways during evolution (Fig. 5, A and B). Apart from the dopDH gene, orthologs of the d-glucarate dehydratase gene (*ycbF*, COG4948) are observed in the pentose degradation gene clusters of both *S. solfataricus* and *H. marismortui*, although remarkably, the keto-deoxy-acid dehydratase of COG0329 is found in all three pathways. In the hydroxyproline degradation pathway, this enzyme might function as a deaminase instead (75).

The apparent mosaic of orthologous and non-orthologous proteins involved in the pentose oxidation pathway suggests that some of these enzymatic steps may have evolved by recruitment events between enzymes from the hexaric acid or hydroxyproline degradation pathways, which also make use of DOP as an intermediate and produce 2-OG as the final product (76, 77). The low number of enzymes required, their common cofactor usage, and the large gain of obtaining the hub metabolite 2-OG as the end product of pentose oxidation, may have been the driving force in the creation of this pathway in aerobically respiring Bacteria and Archaea.

**REFERENCES**

1. Lee, N., Gielow, W., Martin, R., Hamilton, E., and Fowler, A. (1986) *Gene (Amst)*. 47, 231–244
2. Rosenfeld, S. A., Stevis, P. E., and Ho, N. W. (1984) *Mol. Gen. Genet.* 194, 410–415
3. Elsinghorst, E. A., and Mortlock, R. P. (1994) *J. Bacteriol.* 176, 7223–7232
4. Chiang, C., and Knight, S. G. (1969) *Nature* 188, 79–81
5. Fissett, D., Mortlock, R. P., Anderson, R. L., and Wood, W. A. (1964) *J. Biol. Chem.* 239, 2110–2115
6. Wojtkiewicz, B., Szymidzinski, R., Jeziarska, A., and Cocito, C. (1988) *Eur. J. Biochem.* 172, 197–203
7. Weinberg, R., and Doudoroff, M. (1955) *J. Biol. Chem.* 217, 607–624
8. Weinberg, R. (1961) *J. Biol. Chem.* 236, 629–635
9. Dagley, S., and Trudgill, P. W. (1965) *Biochem. J.* 95, 48–58
10. Duncan, M. J. (1979) *J. Gen. Microbiol.* 113, 177–179
11. Duncan, M. J., and Fraenkel, D. G. (1979) *J. Bacteriol.* 137, 415–419
12. Novick, N. J., and Tyler, M. E. (1982) *J. Bacteriol.* 149, 364–367
13. Mathias, A. L., Rigo, L. U., Funayama, S., and Pedrosa, F. O. (1989) *J. Bacteriol.* 171, 5206–5209
14. Palleroni, N. J., and Doudoroff, M. (1957) *J. Bacteriol.* 74, 180–185
15. Dahms, A. S., and Anderson, R. L. (1969) *Biochem. Biophys. Res. Commun.* 36, 809–814
16. Pedrosa, F. O., and Zancan, G. T. (1974) *J. Bacteriol.* 119, 336–338
17. Grogan, D. W. (1989) *J. Bacteriol.* 171, 6710–6719
18. Siebers, B., and Schonheit, P. (2005) *Curr. Opin. Microbiol.* 8, 695–705
19. Johnsen, U., and Schonheit, P. (2004) *J. Bacteriol.* 186, 6198–6207
20. Brouns, S. J., Smits, N., Wu, H., Snijders, A. P., Wright, P. C., de Vos, W. M., and van der Oost, J. (2006) *J. Bacteriol.* 188, 2392–2399
