Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myoinositol-phosphate

Dmitry A. Rodionov, Oleg V. Kurnasov, Boguslaw Stec, Yan Wang, Mary F. Roberts, and Andrei L. Osterman

PNAS published online Mar 2, 2007; doi:10.1073/pnas.0609279104

This information is current as of March 2007.

Supplementary Material
Supplementary material can be found at:
www.pnas.org/cgi/content/full/0609279104/DC1

This article has been cited by other articles:
www.pnas.org#otherarticles

E-mail Alerts
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

Rights & Permissions
To reproduce this article in part (figures, tables) or in entirety, see:
www.pnas.org/misc/rightperm.shtml

Reprints
To order reprints, see:
www.pnas.org/misc/reprints.shtml

Notes:
Di-myo-inositol-phosphate (DIP) is a major osmoprotecting metabolite in a number of hyperthermophilic species of archaea and bacteria. Although the DIP biosynthesis pathway was previously proposed, genes encoding only two of the four required enzymes, inositol-1-phosphate synthase and inositol monophosphatase, were identified. In this study we used a comparative genomic analysis to predict two additional genes of this pathway (termed dipA and dipB) that remained missing. In *Thermotoga maritima* both candidate genes (in an originally misannotated locus TM1418) form an operon with the inositol-1-phosphate synthase encoding gene (TM1419). A predicted inositol-monophosphate cytidylyltransferase activity was directly confirmed for the purified product of *T. maritima* gene dipA cloned and expressed in *Escherichia coli*. The entire DIP pathway was reconstituted in *E. coli* by cloning of the TM1418–TM1419 operon in pBAD expression vector and confirmed to function in the crude lysate. 31P NMR and MS analysis revealed that DIP synthesis proceeds via a phosphorylated DIP intermediate, P-DIP, which is generated by the dipB-encoded enzyme, now termed P-DIP synthase. This previously unknown intermediate is apparently converted to the final product, DIP, by an inositol monophosphatase-like phosphatase. These findings allowed us to revise the previously proposed DIP pathway. The genomic survey confirmed its presence in the species known to use DIP for osmoregulation. Among several newly identified species with a postulated DIP pathway, *Aeropyrum pernix* was directly proven to produce this osmolyte.

**Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate**

**Dmitry A. Rodionov***,†, Oleg V. Kurnasov*, Boguslaw Stec*, Yan Wang‡, Mary F. Roberts‡, and Andrei L. Osterman**§

*Burnham Institute for Medical Research, La Jolla, CA 92037; †Mercket Chemistry Center, Boston College, Chestnut Hill, MA 02467; and §Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia; and ‡Fellowship for Interpretation of Genomes, Burr Ridge, IL 60527

Edited by Emil C. Gotschlich, The Rockefeller University, New York, NY, and approved January 5, 2007 (received for review October 19, 2006)

**Di-myo-inositol 1,1′-phosphate (DIP) is a major osmoprotecting metabolite in a number of hyperthermophilic species of archaea and bacteria. Although the DIP biosynthesis pathway was previously proposed, genes encoding only two of the four required enzymes, inositol-1-phosphate synthase and inositol monophosphatase, were identified. In this study we used a comparative genomic analysis to predict two additional genes of this pathway (termed dipA and dipB) that remained missing. In *Thermotoga maritima* both candidate genes (in an originally misannotated locus TM1418) form an operon with the inositol-1-phosphate synthase encoding gene (TM1419). A predicted inositol-monophosphate cytidylyltransferase activity was directly confirmed for the purified product of *T. maritima* gene dipA cloned and expressed in *Escherichia coli*. The entire DIP pathway was reconstituted in *E. coli* by cloning of the TM1418–TM1419 operon in pBAD expression vector and confirmed to function in the crude lysate. 31P NMR and MS analysis revealed that DIP synthesis proceeds via a phosphorylated DIP intermediate, P-DIP, which is generated by the dipB-encoded enzyme, now termed P-DIP synthase. This previously unknown intermediate is apparently converted to the final product, DIP, by an inositol monophosphatase-like phosphatase. These findings allowed us to revise the previously proposed DIP pathway. The genomic survey confirmed its presence in the species known to use DIP for osmoregulation. Among several newly identified species with a postulated DIP pathway, *Aeropyrum pernix* was directly proven to produce this osmolyte.

The most common mechanism of osmoadaptation in microorganisms involves the accumulation of specific organic osmolytes, so-called compatible solutes, amino acids, sugars, and polyols, that can be taken up from the environment or synthesized de novo (1–3). In thermophiles and hyperthermophiles, compatible solutes are generally different from those found in mesophiles (4), and many of them additionally contribute to thermoprotection.

**Di-myo-inositol 1,1′-phosphate (DIP) is one of the major compatible solutes in a number of thermophilic archaea of the genera *Pyrococcus*, *Methanococcus*, *Thermococcus*, and *Archaeoglobus* and bacteria belonging to the genera *Aquifex*, *Rubrobacter*, and *Thermotoga*, including *Thermotoga maritima* and *Thermotoga neurophilosa* (4–7). DIP levels are highly increased at supraoptimal growth temperatures in both *Methanococcus igneus* and *Thermococcus*, suggesting that this solute also plays a thermoprotective role (6, 8).

Based on the study in *M. igneus* (9) and the reported stereochemistry of DIP (10), the pathway of DIP biosynthesis was originally proposed to occur in four steps: (i) synthesis of L-myo-inositol-1-phosphate (L-I-1-P) from glucose-6-phosphate by NAD+–dependent L-I-1-P synthase [inositol-1-phosphate synthase (IPS)]; (ii) hydrolysis of some of the L-I-1-P by inositol monophosphatase (IMP); (iii) coupling of the L-I-1-P with CTP to form CDP-inositol by CTP:inositol monophosphate cytidylyltransferase (IMPT); and (iv) generation of DIP by condensation of CDP-inositol with inositol via a DIP synthase (DIPS).

Two of these enzymes, IPS and IMP, are not solely committed to DIP biosynthesis as inositol derivatives play an important role in cell wall biogenesis, signaling, and other pathways in a broad spectrum of species including animals, plants, fungi, and mesophilic bacteria (11–13). Both enzymes are broadly conserved, and examples of each have been well characterized biochemically and structurally, including IMP and IPS from the hyperthermophiles *Methanocalo*dococcus *jannaschii* (14), *Archaeoglobus fulgidus* (15–19), *T. maritima* (20), and *Thermococcus kodakaraensis* (21).

However, two additional enzymes required for DIP biosynthesis have not yet been identified in any organism (3). We used a subsystems-based comparative genomic analysis [as implemented in the SEED platform (22)] to explore the DIP biosynthesis pathway (further termed DIP pathway) in thermophilic species with completely sequenced genomes. Analysis of chromosomal clustering and co-occurrence profiles implicated two previously uncharacterized genes (termed *dipA* and *dipB*; see Fig. IA) as candidates for the two missing genes of the DIP pathway. Both representative genes from *T. maritima* were cloned and expressed in *Escherichia coli*. The in vitro reconstitution of the entire pathway in the crude lysate of *E. coli* and the analysis of the individual enzymatic steps revealed the role of each gene in the DIP pathway. Whereas the *dipA* gene was confirmed to encode a predicted IMPCT enzyme, the product of the *dipB* gene involved in the next step of the pathway appeared to generate a previously unknown phosphorylated DIP intermediate (P-DIP) precursor via the condensation of CDP-inositol with L-I-1-P. The existence of this novel enzymatic activity (termed here PDIPS) instead of the previously postulated DIPS suggested a revision to the previously proposed version of the pathway, where the last step is a dephosphorylation of P-DIP by a phosphatase of the IMP family (Fig. 1B).

The comparative genomic analysis of the entire collection of species with completely sequenced genomes integrated in the SEED database (http://theseed.uchicago.edu/FIG/index.cgi) confirmed the presence of the DIP pathway in all archaeal and bacterial species where it was previously detected by biochemical methods.

**Author contributions:** O.V.K., B.S., M.F.R., and A.L.O. designed research; D.A.R., O.V.K., B.S., Y.W., and M.F.R. performed research; D.A.R., M.F.R., and A.L.O. analyzed data; and D.A.R., M.F.R., and A.L.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

**Abbreviations:** DIP, di-myo-inositol 1,1′-phosphate; L-I-1-P, l-myo-inositol-1-phosphate; DIPS, DIP synthase; IMP, inositol monophosphatase; IMPCT, CTP:inositol monophosphate cytidylyltransferase; IPS, inositol-1-phosphate synthase; PDIPS, P-DIP synthase.

†To whom correspondence should be addressed at: Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037. E-mail: osterman@burnham.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0609279104DC1.

© 2007 by The National Academy of Sciences of the USA
Results

Prediction of Novel DIP Pathway Genes by Comparative Genomics. Genome context analysis including chromosomal gene clustering, protein fusions, and occurrence profiles were applied to the available genomes of the DIP-producing microorganisms to identify candidate genes for the last two steps of the DIP biosynthesis pathway. The detailed results of this analysis are captured in the SEED subsystem available online (http://theseed.uchicago.edu/FIG/subsys.cgi; see “Di-Inositol-Phosphate biosynthesis”) and illustrated in Fig. 1 and Table 1.

An observed chromosomal clustering of two previously uncharacterized genes (e.g., APE1514 and APE1516 in A. pernix) with an ips gene (e.g., APE1517) in a number of thermophilic archaea and bacteria (Fig. 1A) provided us with the first strong evidence of their possible role in the DIP biosynthesis pathway. In some of these species, the aforementioned genes are colocalized with the imp gene, revealing an additional link to this pathway. Moreover, orthologs of these two hypothetical genes (termed here dipA and dipB) are always either located next to each other on the chromosome or, more often, fused together to form a single gene (dipAB), further supporting their strong functional coupling.

An additional scanning of the metagenomic libraries of environmental samples revealed clusters of contiguous ips, imp, and dipAB genes in two uncultured organisms: an archaeon from Eel River sediment from northwestern California and a bacterium from a deep-sea sediment of central Pacific northwest province, China (Fig. 1A).

Additional evidence supporting the possible involvement of these genes in DIP synthesis came from their occurrence only in the genomes containing ips and imp genes (Table 1). This co-occurrence profile is not in contradiction with the existence of many microbial genomes that contain ips and imp genes but do not contain dipA-dipB genes. For example, in Mycobacterium tuberculosis and other actinobacteria IPS plays an essential role in the production of the major inositol-containing thiol and cell wall lipoplys (23).

Notably, dipA and dipB genes identified in this study are absent from the current GenBank annotation of the T. maritima genome as the respective locus TM1418 was deemed to contain an authentic frame shift. An alternative interpretation was available in the SEED database (corresponding protein IDs: fig243274.1.pep.1893 and fig243274.1.pep.1892 at http://theseed.uchicago.edu/FIG/index.cgi) (Fig. 2) suggesting the presence of the two protein-encoding genes (here referred to as TM1418a and TM1418b) overlapping by 7 bp. Although most species appear to contain a fusion of dipA and dipB genes within a single reading frame dipAB, the existence of two separate genes in T. maritima was additionally supported by the presence of single-gene orthologs in A. pernix and Novosphingobium aromaticivorans.

Table 1. Phylogenetic distribution of the DIP biosynthesis pathway in bacteria and archaea

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Four-step DIP biosynthesis pathway</th>
<th>IPS</th>
<th>IMPCT</th>
<th>PDIPS</th>
<th>IMP</th>
<th>DIP production</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. maritima</td>
<td>TM1419</td>
<td>TM1418a</td>
<td>TM1418b</td>
<td>TM1415</td>
<td></td>
<td>Known</td>
</tr>
<tr>
<td>Aquifex aeolicus</td>
<td>aq.1763</td>
<td>aq.1367</td>
<td>aq.1367</td>
<td>aq.1983</td>
<td></td>
<td>Known*</td>
</tr>
<tr>
<td>Rubrobacter xylanophilus</td>
<td>Rxy021258</td>
<td>Rxy021259</td>
<td>Rxy021259</td>
<td>Rxy021693</td>
<td></td>
<td>Known</td>
</tr>
<tr>
<td>Novosphingobium aromaticivorans</td>
<td>Saro3074</td>
<td>Saro3073</td>
<td>Saro3075</td>
<td>Saro2521</td>
<td></td>
<td>Predicted</td>
</tr>
<tr>
<td>Uncultured bacterium 17H9</td>
<td>17H9.20</td>
<td>17H9.22</td>
<td>17H9.22</td>
<td>17H9.21</td>
<td></td>
<td>Predicted</td>
</tr>
<tr>
<td>A. pernix</td>
<td>APE1517</td>
<td>APE1514</td>
<td>APE1516</td>
<td>APE1798</td>
<td></td>
<td>Confirmed</td>
</tr>
<tr>
<td>Ar. fulgidus</td>
<td>AF1794</td>
<td>AF0263</td>
<td>AF0263</td>
<td>AF2372</td>
<td></td>
<td>Known</td>
</tr>
<tr>
<td>Pyrococcus horikoshii</td>
<td>PH1605</td>
<td>PH1219</td>
<td>PH1219</td>
<td>PH1897</td>
<td></td>
<td>Predicted</td>
</tr>
<tr>
<td>Pyrococcus abyssi</td>
<td>PAB1989</td>
<td>PAB2433</td>
<td>PAB2433</td>
<td>PAB0189</td>
<td></td>
<td>Predicted</td>
</tr>
<tr>
<td>Pyrococcus furiosus</td>
<td>PF1616</td>
<td>PF1058</td>
<td>PF1058</td>
<td>PF2014</td>
<td></td>
<td>Known</td>
</tr>
<tr>
<td>Thermococcus kodakarenzii</td>
<td>TK2278</td>
<td>TK2279</td>
<td>TK2279</td>
<td>TK2787</td>
<td></td>
<td>Known*</td>
</tr>
<tr>
<td>Uncultured archaeon GZfos13E1</td>
<td>GZ13E1.33</td>
<td>GZ13E1.32</td>
<td>GZ13E1.32</td>
<td>GZ13E1.31</td>
<td></td>
<td>Predicted</td>
</tr>
</tbody>
</table>

Genes encoding four enzymes of DIP pathway are shown by standard GenBank identifiers. N and C superscripts denote N-terminal and C-terminal domains in the fusion IMPCT–PDIPS proteins. Organisms experimentally shown to produce DIP are indicated in the last column (4, 7). DIP production in A. pernix was predicted and confirmed in this study.

*Cases when DIP production was confirmed in subspecies related to but distinct from those with available genomic sequences.

Fig. 1. Reconstruction of the DIP biosynthesis pathway in prokaryotic genomes. (A) Genome context of DIP pathway genes in bacterial and archaeal genomes. (Hyper)thermophilic species are marked with “T.” Orthologs are shown by matching colors of individual genes or respective segments of fused genes. Predicted novel genes (dipA and dipB) are fused (dipAB) or occur in chromosomal clusters with each other and, often, with the two known genes of the DIP pathway (ips and imp). (B) A diagram of the revised DIP synthesis pathway. IPS, EC 5.5.1.4; IMPCT, 2.7.7.2; PDIPS, 2.7.8.2; IMP, 3.1.3.25.
Methods was cloned and overexpressed in family, we conjectured that DipB is the most likely candidate for the similarity of chemical reactions catalyzed by the previously characterized. Because the product of dipA gene was expected to be membrane-bound and, hence, insoluble, its activity was tested in a crude lysate of T. maritima dipB gene encodes a 205-aa protein that belongs to the CDP-alcohol phosphatidyltransferase class-I protein family (COG1213). Among characterized members of this protein family are glucose-1-phosphate cytidylyltransferase from Salmonella typhimurium, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase from E. coli, and mannose-1-phosphate guanylyltransferase from yeast [see supporting information (SI) Fig. 8]. Based on these observations, we tentatively assigned the role of missing IMPCT to the dipA gene product.

The T. maritima dipB gene encodes a 205-aa protein that belongs to the CDP-alcohol phosphatidyltransferase class-I protein family (COG1213). A BLAST search revealed similarity of DipB to phosphatidylinositol synthases from bacteria, and phosphatidylinositol synthases from eukaryotes, phosphatidylglycerophosphate synthases from bacteria, and phosphatidylserine synthases from eukaryotes, phosphatidylglycerophosphate synthases from bacteria, and phosphatidylinositol synthases from eukaryotes. This similarity is particularly striking in the region between amino acids 70 and 100, showing a perfect match to the consensus sequence D(X)2DG(X)2AR(X)2N(X)5G(X)3D(X)3D, characteristic of several alcohol phosphatidylinosferases (24). Based on the overall similarity of chemical reactions catalyzed by the previously characterized proteins from the CDP-alcohol phosphatidyltransferase family, we conjectured that DipB is the most likely candidate for the role of the missing DIPS.

Experimental Characterization of Novel DIP Pathway Enzymes from T. maritima. To assess the activity of the product of the T. maritima gene dipA predicted to encode a soluble IMPCT enzyme, this gene was cloned and overexpressed in E. coli (see SI Materials and Methods). The recombinant protein with the N-terminal His6 tag was purified by NiNTA chromatography and enzymatically characterized. Because the product of T. maritima gene dipB was predicted to be membrane-bound and, hence, insoluble, its activity was tested in a crude lysate of E. coli where this gene was expressed as a part of the T. maritima operon 3ips-dipA-dipB cloned in the pBAD expression vector. The results of these enzymatic analyses are described in the following subsections. However, so far we failed to express DipB enzyme alone in the active form, which may reflect the importance of an apparent tight association of DipA and DipB proteins (that are more often fused in one polypeptide in other species) for the proper membrane insertion and activity of DipB enzyme.

Enzymatic Characterization of IMPCT. The predicted enzymatic activity of the purified recombinant T. maritima IMPCT enzyme was confirmed by the formation of CDP-inositol as characterized by 31P NMR spectroscopy and mass spectrometry (MS). The IMPCT reaction converts the CTP and L-1-P into CDP-inositol and inorganic pyrophosphate. We used the detection of the latter product for quantitative enzymatic characterization of IMPCT using the coupled colorimetric assay for pyrophosphate. Because of significant substrate inhibition observed at higher concentrations of L-1-P (>1 mM), only an estimate of the apparent Km ≈ 0.2 mM was obtained for one of its specific substrates, L-1-P, at fixed saturating concentration of the second substrate (1.5 mM CTP). With L-1-P fixed at 1.0 mM, the Vmax at 80°C was 160 ± 10 mmol min⁻¹ mg⁻¹, and the apparent Km for CTP was 0.16 ± 0.04 mM (Fig. 3). The IMPCT appears to be specific for L-inositol-1-phosphate because no CDP-inositol was produced when D-inositol-1-phosphate (Cayman Chemical) was incubated with the recombinant enzyme. This enzyme is also highly selective for CTP because no activity could be detected by the pyrophosphatase coupled assay in the presence of ATP, GTP, or UTP. Moreover, no detectable adduct formation could be detected by 31P-NMR spectra even after extensive incubation with any of these alternative NTPs. Deoxy-CTP could serve as a substrate, although the rate was 20-fold lower (with 1.5 mM L-1-P and 3 mM deoxi-CTP) than with CTP. IMPCT activity was reduced >25-fold in the absence of added MgCl2, and it was fully suppressed by 5 mM EDTA. These observations confirm the participation of Mg2⁺ in its catalytic mechanism, characteristic of this nucleotidyltransferase family.

Identification of PDIPS Enzyme and Reconstitution of the Entire DIP Pathway. Because the product of the dipB gene was expected to be a membrane-protos, its activity was tested in the crude lysate of E. coli carrying the operon 3ips-dipA-dipB (TM1419, TM1418a, and TM1418b) from T. maritima. This experimental setup allowed us to combine testing of individual enzymatic steps with the pathway reconstitution as none of the respective enzymatic activities or metabolites are present in the E. coli host.

The 2,546-bp PCR-amplified segment of T. maritima chromosome was cloned into the pBAD-TOPO expression vector and transformed to E. coli BL21. The crude cell lysate was obtained after induction with arabinose. Product formation was monitored by 31P-NMR after incubation with respective substrates at 80°C and compared with available standards and with the samples prepared using the same host strain of E. coli carrying the same vector with the β-galactosidase gene as a negative control. The 1H-coupled 31P-NMR analysis of the reaction mixture obtained after incubation of the cell lysate with L-1-P and CTP confirmed the formation of CDP-inositol and of a novel phosphodiester compound with a chemical shift similar to that of DIP (Fig. 4). The identity of this compound as P-DIP was later confirmed by MS analysis and by its conversion to DIP upon incubation with an exogenous IMP phosphatase (see below).

The results obtained in this study provided an unambiguous...
assignment of the observed P-DIP synthase (PDIPS) activity to the product of the dipB gene as the properties of IPS and IMPCT encoded by the two other genes of the operon were characterized separately (in ref. 17 and in this study, respectively). Addition of the free inositol to the mixture did not lead to any additional formation of DIP, confirming that formation of the P-DIP intermediate via condensation of CDP-inositol with L-I-1-P (and not with free inositol as for a previously postulated DIPS enzyme) is the main and likely the only possible route of DIP synthesis. These findings allowed us to propose a revised version of the DIP pathway shown in Fig. 1B.

Only lysates of cells that contained all three genes of a T. maritima DIP operon displayed all of the resonances (including a number of novel peaks) consistent with the proposed pathway and not present in control samples (Fig. 4). Under the conditions used, most of the L-I-1-P was converted to CDP-inositol as monitored by two resonances at −10.2 and −11.0 ppm (each appearing as an AB quartet). Likewise, most of the added CTP was consumed (little intensity for the β-phosphorus remained at ∼22 ppm, not shown in these graphs). A new triplet resonance appeared at −0.2 ppm, consistent with synthesis of a phosphodiester such as DIP; another new resonance ∼4 ppm (a doublet in proton-coupled spectra) consistent with a phosphomonoester had the same integrated intensity as the new phosphodiester peak. The same products in the 31P-NMR spectrum were produced when 3 mM glucose-6-phosphate was incubated with 1.5 mM NAD+°, 6 mM CTP, and 3 mM MgCl2 (Fig. 4C), indicating that all three enzymes function under these conditions.

Neither of the two peaks in the 31P-NMR spectra corresponding to CDP-inositol and P-DIP appeared in the absence of Mg2+° or in the presence of 6 mM EDTA. In the presence of Mg2+°, both of them increased with the time of incubation and exhibited strong temperature dependence. Under our experimental conditions and at the temperatures examined (2 h of incubation at 65, 75, and 85°C), nearly all of the L-I-1-P was converted to CDP-inositol. Activation energy of its conversion to P-DIP is 58 kcal/mol.

Verification of the Predicted DIP Production in A. pernix. The comparative genomic analysis allowed us to infer the presence of the DIP pathway in a number of species with completely sequenced genomes where DIP production has not been previously reported (Table 1). To test this conjecture, we checked for DIP production in the hyperthermophilic archaeon A. pernix. 31P-NMR spectrum of the ethanol extract of these cells showed one major resonance at −0.3 ppm consistent with DIP, without any indication of P-DIP presence (Fig. 7A). The 1H NMR spectrum (Fig. 7B) showed resonances consistent with DIP and mannosylglycerate. The identity of the resonances for these two compounds was confirmed by TOCSY and HMQC experiments (6, 26). 13C chemical shifts were obtained from an HMQC experiment and were consistent with published 13C and 1H shifts for these molecules (27). The MS analysis revealed the presence of 421.2 m/z peak consistent with the DIP anion and the 267.2 m/z peak likely corresponding to mannosylglycerate.

Discussion

Although our current knowledge of major metabolic pathways and respective genes in well-studied model bacteria such as E. coli is nearly comprehensive, it can be only partially projected on the

---

Fig. 4. 31P NMR (202.2 MHz) 1H-coupled spectra of soluble fractions from incubations (at 80°C) of E. coli BL21-DIP lysed cells with DIP precursors. (A) Control mixture with 3 mM L-I-1-P, 6 mM CTP, 3 mM MgCl2, and 6 mM EDTA incubated for 2 h. (B) Lysed cells incubated for 2 h with 3 mM L-I-1-P, 6 mM CTP, and 3 mM MgCl2. (C) Lysed cells incubated for 2.5 h with 3 mM glucose-6-phosphate, 1.5 mM NAD+°, 3 mM CTP, and 3 mM MgCl2. After centrifugation of cell debris, EDTA (6 mM) was added to all samples before analysis by 31P NMR spectroscopy. The asterisks indicate the appearance of the DIP-related peaks; the one at −0.2 ppm is the phosphodiester, and the one at −3 ppm is the phosphomonoester.

Fig. 5. Mass spectrum of the soluble fraction from incubation (at 80°C) of E. coli BL21-DIP lysed cells that had been incubated for 2 h with 3 mM L-I-1-P, 6 mM CTP, and 3 mM MgCl2. Identities of the major peaks are indicated. The asterisks highlight ions detected only in spectra from incubations with cell lysates containing the three DIP biosynthetic enzymes IPS, IMPCT, and DIPS.
rapidly growing number of sequenced genomes of more distant species. Deep-branched bacteria and archaea, including many extremophiles, represent a particular challenge because of a substantial number of unique pathways supporting their divergent lifestyles. In this study we used comparative genomics to address one such challenge and to predict two previously uncharacterized (missing) genes completing the biosynthetic pathway of DIP, a major osmoprotecting metabolite in a number of thermophiles archaea and bacteria (Fig. 1B).

The key evidence implicating two candidate genes termed dipA and dipB with the DIP pathway was provided by their clustering on the chromosome with the genes encoding IPS and IMP (Fig. 1A), the two previously characterized components of this pathway. Additional support for these conjectures came from the observed co-occurrence profile and a frequent fusion of dipA and dipB into one contiguous gene dipAB. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

Enzymatic activities of proteins encoded by the genes dipA and dipB were experimentally assessed by cloning and expression of the corresponding genes (termed here dipA and dipB) into one recombinant gene. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

Enzymatic activities of proteins encoded by the genes dipA and dipB were experimentally assessed by cloning and expression of the recombinant genes. The key evidence implicating two candidate genes termed dipA and dipB with the DIP pathway was provided by their clustering on the chromosome with the genes encoding IPS and IMP (Fig. 1A), the two previously characterized components of this pathway. Additional support for these conjectures came from the observed co-occurrence profile and a frequent fusion of dipA and dipB into one contiguous gene dipAB. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

Enzymatic activities of proteins encoded by the genes dipA and dipB were experimentally assessed by cloning and expression of the recombinant genes. The key evidence implicating two candidate genes termed dipA and dipB with the DIP pathway was provided by their clustering on the chromosome with the genes encoding IPS and IMP (Fig. 1A), the two previously characterized components of this pathway. Additional support for these conjectures came from the observed co-occurrence profile and a frequent fusion of dipA and dipB into one contiguous gene dipAB. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

Enzymatic activities of proteins encoded by the genes dipA and dipB were experimentally assessed by cloning and expression of the recombinant genes. The key evidence implicating two candidate genes termed dipA and dipB with the DIP pathway was provided by their clustering on the chromosome with the genes encoding IPS and IMP (Fig. 1A), the two previously characterized components of this pathway. Additional support for these conjectures came from the observed co-occurrence profile and a frequent fusion of dipA and dipB into one contiguous gene dipAB. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

The key evidence implicating two candidate genes termed dipA and dipB with the DIP pathway was provided by their clustering on the chromosome with the genes encoding IPS and IMP (Fig. 1A), the two previously characterized components of this pathway. Additional support for these conjectures came from the observed co-occurrence profile and a frequent fusion of dipA and dipB into one contiguous gene dipAB. This type of genome context analysis has been successfully applied for identification of missing genes in many metabolic pathways (for review, see ref. 28). General class functions tentatively assigned by homology to the protein products of genes dipA (sugar phosphate nucleotidytransferase, COG1213) and dipB (CDP-alcohol phosphatidytransferase, COG0558) were consistent with their expected functional roles in the DIP pathway.

An enzymatic analysis of the overexpressed and purified product of gene dipA (TM1418a) confirmed its predicted IMPCT activity, including previously inferred features, such as preference for CTP over other NTPs and a strict requirement of Mg2+ for activity. On the other hand, the analysis of the enzymatic activity of the product of gene dipB (TM1418b) provided some unexpected results. In keeping with the initial version of the DIP pathway (9), this enzyme was expected to generate DIP by a direct condensation of CDP-inositol (produced from L-I-1-P by IMPCT) with free myo-inositol (produced from L-I-1-P by IMP). However, the reconstitution of the pathway in the crude lysate of E. coli carrying the entire ips-dipA-dipB operon cloned in the expression vector revealed the formation of the previously unknown intermediary metabolite, P-DIP. Therefore, a condensation reaction catalyzed by the product of the dipB gene actually involves CDP-inositol and L-I-1-P rather than free inositol. Based on these results, the respective enzyme was termed PDIPS. The last step in the revised version of the DIP pathway is the dephosphorylation of the newly identified intermediate (Fig. 1B). This reaction is likely catalyzed by an IMP-like phosphatase as confirmed by the formation of DIP upon addition of the pure recombinant IMP enzymes from T. maritima and T. maritima to the reaction mixture containing P-DIP.

These findings illustrate a synergy between bioinformatics and experimental aspects of gene discovery. Although the comparative genomics techniques allowed us to accurately predict two novel genes of the DIP pathway, the consequent experimental analysis of the recombinant enzymes was required to confirm one of the inferred activities (IMPCT) and, more importantly, to revise an original prediction of the second activity (PDIPS) and of the entire pathway.

The phylogenetic pattern of occurrence of the DIP pathway genes is in perfect correlation with the previously established distribution of compatible solutes in thermophiles and hyperthermophiles (Table 1). Moreover, projection of the DIP biosynthesis subnetwork across the whole collection of integrated complete and almost-complete genomes (http://theseed.uchicago.edu/FIG/subsys.cgi?user=master&ssa_name=Di-Inositol-Phosphatase_biosynthesis&request=show_ssa) allowed us to infer this pathway in seven archaea and six bacteria (including two uncultured species; see Fig. 1A). This inference was further validated in this study by the experimental detection of DIP in the archaea A. pernix. Also in agreement with the available data, DIP pathway signature genes dipA and dipB are absent from the available genomes of thermophiles where the presence of DIP was not detected in Thermus thermophilus, Pyrobaculum aerophilum, Methanopyrus kandleri, Sulfolobus solfataricus, and Methanobacterium thermautotrophicum.

Materials and Methods
Genomes and Bioinformatics Tools. The bulk of comparative genomic analysis, including subsystem encoding and genome context analysis (chromosomal clustering and phylogenetic profiling), was performed by using the SEED genomic database and tools implemented therein (http://theseed.uchicago.edu/FIG/index.cgi) (22). Complete and nearly complete genomes of bacteria and archaea were from the GenBank database (www.ncbi.nlm.nih.gov/GenBank). Preliminary genome sequence data for T. maritima was obtained from The Institute for Genomic Research. We used ClustalX to construct multiple protein alignments (30), Psi-BLAST...
Cloning and Expression of T. maritima ips-dipA-dipB Operon. The 2,546-bp segment containing ips and two novel predicted genes, dipA and dipB, was PCR-amplified by using T. maritima MSB8 genomic DNA and oligonucleotide primers ATGatgaacctgatc-ctgg and TC/Acatgtgaccagaagtt. This fragment was cloned in the pBAD-TOPO expression vector (Invitrogen), and the DNA sequence of the entire fragment was verified. The expression of the operon-encoded genes was performed in E. coli strain BL-21 with arabinose induction. Crude cell lysates [0.2 g cells in 1 ml of 50 mM Tris-HCl (pH 7.5)] were heated at 80°C for 5 min to inactivate most NAD+ that was generated by incubating glucose-6-P in the presence of 2,546-bp segment containing and dipA operon-encoded genes was performed in a host cell. The expression of these genes was assessed by SDS-PAGE and Western blotting.

Electrospray Mass Spectrometry. Electrospray mass spectrometry of the cell lysates soon after preparation was performed by using negative ion mode in acetonitrile/H2O, 1:1 with 0.1% ammonia using an LCQ Classic spectrometer.

Growing of A. pernix for DIP Production. A. pernix strain JCM 9820, obtained from the American Type Culture Collection (Manassas, VA), was grown in Diffco Marine Broth 2216 to which Na2S2O3·5H2O (0.1 g per 100 ml of medium) had been added. The medium was inoculated with ice chips of frozen A. pernix, and the flask containing the cells was placed in a shaking water bath at 92°C and 120 rpm (37). After a 30-h lag time, the cells began to grow exponentially (doubling time of <1 h). Cells were harvested after ~40 h after the lag period when the cells reached OD600 = 0.53; at this point they appeared to be in stationary phase. The cell suspension was centrifuged and the pellet resuspended in 2% NaCl. After centrifugation, the pellet was mixed with 70% ethanol and the mixture was placed in a bath sonicator for 20 min to ensure lysis of the cells. After centrifugation, the pellet was extracted five times with 70% ethanol and all of the supernatants combined, frozen, and lyophilized. The dried sample was solubilized in D2O, and both 1H and 31P-NMR spectra were obtained.

Note. When this manuscript was in preparation, Borges et al. (38) proposed a similar revision of the DIP pathway based on the detection of the respective enzymatic activities and intermediary metabolites in crude extracts of A. fulgidus by NMR. The results of their study are in agreement with the results presented here and confirm that the DIP synthesis occurs via a P-DIP intermediate.

We thank Scott Lesley at the Joint Center for Structural Genomics for providing T. maritima chromosomal DNA and Ross Overbeck and the whole team of developers of the SEED genomic platform at the Fellowship for Internship of Genomics and Argonne National Laboratory for providing guidance and access to integrated genomes and tools for their comparative analysis. This study was partially supported by the National Institutes of Health Roadmap National Technology Networks on Networks and Pathways Grant 5 U54 RR020843-03, by Department of Energy Biociences DE-FG02-01ER20025 (to J.P.R.), and by National Institutes of Health Grant 1RO1 GM64841 (to B.S.).