Biodata of **Laszlo N. Csonka**, author of ***"What We Can Deduce about Metabolism in the Moderate Halophile Chromohalobacter salexigens from its Genomic Sequence"***

Dr. Laszlo Csonka is a professor in the Department of Biological Sciences at Purdue University. He received his Ph. D. in 1975 at Harvard Medical School in the laboratory of Dr. Dan Fraenkel, studying the pathways of NADPH formation in *Escherichia coli*. His main research interests are the analysis of the response to osmotic stress in bacteria and plants, the connection between osmotic adaptation and thermotolerance, and carbon-source metabolism in bacteria.

E-mail: **lcsonka@bilbo.bio.purdue.edu**

WHAT WE CAN DEDUCE ABOUT METABOLISM IN THE MODERATE HALOPHILE *CHROMOHALOBACTER SALEXIGENS* FROM ITS GENOMIC SEQUENCE

LASZLO N. CSONKA[1], KATHLEEN O'CONNOR[1], FRANK LARIMER[2], PAUL RICHARDSON[3], ALLA LAPIDUS[3], ADAM D. EWING[4], BRADLEY W. GOODNER[4] and AHARON OREN[5]

[1]*Department of Biological Sciences, Purdue University, West Lafayette IN 47907-1392, USA;* [2]*Genome Analysis and Systems Modeling, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge TN 37831, USA;* [3]*DOE Joint Genome Institute, Walnut Creek CA 94598, USA*; [4]*Department of Biology, Hiram College, Hiram OH 44234, USA;* [5]*The Institute of Life Science, and the Moshe Shilo Minerva Center for Marine Biogeochemistry, The Hebrew University of Jerusalem, 91904, Israel*

1. Introduction

A draft sequence of the genome of the moderate halophile *Chromohalobacter salexigens* (formerly *Halomonas elongata*) DSM3043 (Fig. 1) has been determined to an 8X coverage by the Joint Genome Institute (JGI) of the US Department of Energy. From this sequence, 3370 predicted protein-coding genes were identified and provisionally annotated by computer analysis at the JGI. The draft sequence and annotation are available at http://genome.jgi-psf.org/microbial/index.html. In this chapter, we present a brief overview of insights we gained into metabolism in *C. salexigens* from the sequence. In view of the close phylogenetic relationship of *C. salexigens* to the Pseudomonaceae and other γ-Proteobacteria (see below), we used known enzyme sequences from Pseudomonaceae, *Acinetobacter*, *Escherichia coli*, other Enterobacteriaceae, and other bacteria (in this order of priority) as query sequences for orthologs in *C. salexigens*. Any match with a probability ("expect value" in tblastn; http://www.ncbi.nlm.nih.gov/BLAST) ≤ e-20 was considered to be highly significant, values between e-5 and e-20 to be marginal, and ≥ e-5 to be not significant. The usual disclaimers about inferences from genomic sequence are in order: that all of the annotations that we are assigning are provisional and await confirmation by biochemical or functional genetic analyses, and that the predictions of the functions of the *C. salexigens* gene products are only as valid as the accuracy of the assignment of the function of matching proteins in other organisms.

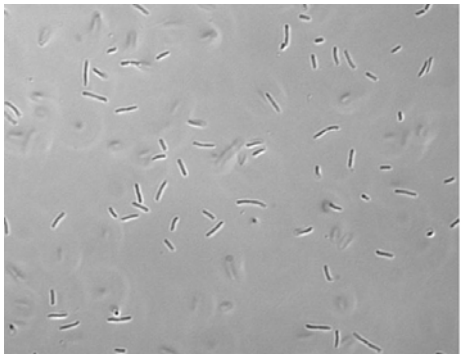

Figure 1. *Chromohalobacter salexigens* DSM3043, grown in LB medium supplemented with 1 M NaCl. Photograph: Kathleen O'Connor and Christopher Staiger.

2. Overview of the Sequence of the *C. salexigens* Genome

Pulsed-field gel electrophoretic analysis of total DNA cut with the rare-cutting enzymes *Pac*I and *Pme*I suggested that *C. salexigens* possesses a main chromosome of ~3.9 Mbp and a low copy plasmid of < 100 kbp (data not shown). The JGI generated a total of 3.7 Mbp unique sequences, and therefore, depending on the accuracy of our estimate of the genome size, we may be very near the complete coverage of the genome. The *C. salexigens* genome is 63.9% G + C. This value is close to the 66.6% G + C content found in *Pseudomonas aeruginosa* (Stover et al., 2000), a related non-halophilic γ-proteobacterium, but differs substantially from the 40.3% G + C composition in a second closely related γ-proteobacterium, *Acinetobacter* ADP1 (Barbe et al., 2004). However, despite the difference in nucleotide composition, the genome of *C. salexigens* is more similar in its size to the 3.6 Mbp (3325 ORF) genome of *Acinetobacter* than to the 6.3 Mbp genome of *P. aeruginosa*. The classification of *C. salexigens* among the γ-Proteobacteria is confirmed by the high similarity that the ORFs of this organism show to orthologs from other organisms in this subdivision (Table 1).

As can be seen in this Table, > 60% of the ORFS of *C. salexigens* show the highest similarity to orthologs from γ-Proteobacteria. In spite of the halophilic characteristics of *C. salexigens*, 37% of its ORFs are most similar to orthologs from Pseudomonaceae (which include *Acinetobacter*), in accord with the close phylogenic relationship between these organisms.

3. Carbon Source Metabolic Enzymes

We used the sequence data to make predictions about the carbon source metabolic repertoire of *C. salexigens*. The constituent enzymes of various pathways that we searched in *C. salexigens* were derived from the following references: central metabolism

was based on the chapter by Fraenkel (1996), the tricarboxylic acid (TCA) cycle by Cronan and LaPorte (1996), peripheral carbon source metabolism by Lin (1996) and Gottschalk (1985), and aromatic compound metabolism by Horn et al. (1991) and Nakazawa and Yokota (1973). For the tracing of the aromatic degradative pathways we also made extensive use of the University of Minnesota Biocatalysis/Biodegradation Database (http://umbbd.ahc.umn.edu). Most of the amino acid sequences of enzymes that were used as to search the *C. salexigens* genome were obtained from the Expert Protein Analysis System (ExPASy; http://www.expasy.org).

TABLE 1. The similarity of the predicted ORFs of *C. salexigens* to ORFs from other organisms.

Organism	% of ORFs of *C. salexigens* that show highest tblastn similarity to an ORF in the indicated organism
γ-Proteobacteria	61
Pseudomonaceae	37
Enterobacteriaceae	9
Vibrionaceae	6
Alteromonaceae	4
Xanthomonaceae	2
Other γ-Proteobacteria	3
α-Proteobacteria	13
β-Proteobacteria	11
Bacilli	3
Eukaryotes	3
δ-Proteobacteria	2
Actinobacteria	2
Cyanobacteria	1
Other Eubacteria	4
Archaea	0.5

3.1. ENZYMES OF CENTRAL METABOLISM

We found good matches in *C. salexigens* for all of the enzymes of the Embden-Meyerhof-Parnas pathway from phosphoglucose isomerase to pyruvate kinase. *C. salexigens* has the "standard" suite of glycolytic enzymes as found in *E. coli*, and it differs from *P. aeruginosa*, which lacks phosphofructokinase, and from *Acinetobacter*, which is devoid of both phosphofructokinase and pyruvate kinase (Barbe et al., 2004). We detected all of the enzymes of the oxidative and non-oxidative branches of the hexose monophosphate shunt as well as the Entner-Doudoroff pathways.

C. salexigens has good matches to the three subunits (E1, E2, and Lpd) of the NAD-dependent pyruvate dehydrogenase. There is no indication that it has an ortholog for pyruvate formate lyase, which is used in other organisms under anaerobic conditions. We found good matches to all of the usual TCA cycle enzymes. The one curiosity in this regard is the presence of both an NADP-linked isocitrate dehydrogenase (EC 1.1.1.42) and an NAD-linked isoenzyme (EC 1.1.141). The latter enzyme is not present in *E. coli*

and *Salmonella typhimurium*, and the difference in the function of the two is not known. *C. salexigens* has the capacity to form oxalacetate to fixing CO_2 via phosphoenolpyruvate (PEP) carboxylase.

3.2. GLUCONEOGENESIS

Based on our understanding of metabolism in other organisms, growth on gluconeogenic carbon sources requires enzymes that enable the cell to circumvent "irreversible" reactions to generate pyruvate and PEP from TCA cycle intermediates and other 3-carbon compounds and to form fructose-6-P from fructose-1,6-P_2.

3.2.1. TCA Cycle to PEP

We found good matches for phosphoenolpyruvate synthase, which gives PEP from pyruvate, and PEP caboxykinase, which decarboxylates oxalacetate to yield PEP. The organism has three potential malic enzymes, which carry out the oxidative decarboxylation of malate to pyruvate, two of which appear to NADP-linked and one NAD-linked. Finally there is a membrane-bound oxalacetate decarboxylase, which produces pyruvate from oxalacetate. The latter reaction is catalyzed by a membrane bound-enzyme whose activity is linked to Na^+ extrusion, and therefore it may play a role in the halophilism or halotolerance of the organism.

3.2.2. Fructose-1,6-bisphosphatase

In contrast to the pyruvate or PEP-forming enzymes, we have not been able to find a good match to fructose-1,6-bisphosphatase (Fbpase) that can generate fructose-6-phosphate during gluconeogenic growth. There are at least six classes of proteins which have been recognized as having Fbpase activity. Class I Fbpase is the standard AMP-sensitive enzyme (Fraenkel, 1996) that is present in plants, animals, a variety of Eubacteria, and the extremely halophilic Archaea *Halobacterium salinarum* NRC-1 and *Haloarcula marismortui* ATCC 43049. This protein has been demonstrated by extensive genetic and biochemical analysis to be the physiologically important activity in *E. coli* used during gluconeogenic growth. Class II is represented by the GlpX protein, which was discovered as part of the *glp* (glycerol-3-phosphate) regulon in *E. coli* (Donahue et al., 2000) and shown to be the important Fbpase in *Mycobacterium tuberculosis* and a number of other Actinomycetes (Movahedzadeh et al., 2004). Although the GlpX protein of *E. coli* has Fbpase activity, its function is a mystery, because it is not adequate to support growth on gluconeogenic carbon sources in *fbp* mutants (which lack the Class I Fbpase), even when the *glp* regulon is induced by glycerol. Class III Fbpase contains the *yydE* gene product of *Bacillus subtilis* (Fujita et al., 1998). This enzyme has been shown to have Fbpase activity by in vitro assays and can complement a *Δfbp* mutation in *E. coli*, but its function is also unclear, because *B. subtilis* strains carrying a deletion of the *yydE* gene are not defective in growth on gluconeogenic carbon sources. Class IV of Fbpase has been identified in Archaea as a protein that has both Fbpase and inositol 1-phosphatase (I-1-Pase) activities (Stec et al., 2000). In addition to this enzyme, there is a Class V Fbpase in some but not all Archaea, which has nearly exclusive affinity for fructose-1,6-bisphosphate (Sato et al., 2004). *Thermococcus kodakaraensis* is one of the

Archaea that has orthologs of both the Class IV and Class V enzymes. Deletions were targeted to the genes encoding these proteins, with the result that the loss of the Class V enzyme abolished gluconeogenic growth, whereas removal of the Class IV enzyme did not result in an obvious phenotype (Sato et al., 2004). The implication is that the latter is the important Fbpase, at least in the Archaea that have both the Class IV and Class V enzymes. Finally, Class VI comprises sedoheptulose-1,7-bisphosphatase, a member of the Calvin-Benson-Bassham cycle, which has been found to be able hydrolyze fructose-1,6-bisphosphate also (http://us.expasy.org).

To make a long story short, we did not find significant matches (Expect values < e-5) in the *C. salexigens* genome to representatives of any of the classes of Fbpases, except for Class IV, the archaeal type Fbpase/I-1-Pase. However, even here, the matches were mostly borderline: when used as query in a tblastn search against the 23 sequenced Archaea in the data base, the *C. salexigens* ortholog of the Fbpase/I-1-Pase gave a match with an expect value of e-29 against a gene in the extreme halophile *Har. marismortui* ATCC 43049, but the other 22 had matches with expect values in the range of e-19 to e-0.8. The *C. salexigens* Fbpase/I-1-Pase gene has good matches to a gene found in most of not all Bacteria (e-20 to e-40). It was demonstrated in vitro that the product of the *E. coli* orthologous gene, called *suhB* (expect value of match to *C. salexigens* orthologs = 6e-34), has I-1-Pase activity, but the function of this enzyme in *E. coli* is yet to be discovered (Chen and Roberts, 2000).

In view of the facts that this putative *C. salexigens* Fbpase/I-1-Pase gene shows only low similarity to orthologs in Archaea and that the counterpart gene in *T. kodakaraensis* has been shown to be unnecessary for gluconeogenesis, we feel that that probably this gene product is not the authentic Fbpase in *C. salexigens*, and that the real Fbpase has not been revealed by the genomic analysis. In this regard, *C. salexigens* is similar to *Agrobacterium tumefaciens* (*Rhizobium radiobacter*) in which it has not been possible to infer an Fbpase gene from genomic analysis (B. Goodner, unpublished). Thus, there may be yet another form of Fbpase (Class VII?) in *C. salexigens* (and *A. tumefaciens*) that needs to be identified. However, another more interesting, albeit speculative possibility could be that there might be a novel pathway from 3- or 2-carbon compounds to fructose-6-phosphate or a pentose phosphate that bypasses fructose-1,6-bisphosphate.

3.3. SPECIFIC CARBON SOURCES

The enzymes in the likely metabolic pathways of various carbon and energy sources that we found in the *C. salexigens* are summarized in Table 2. This analysis is mainly concentrated on the catabolic pathways of simple sugars and other common carbon sources, and is not intended to be comprehensive. *C salexigens* can degrade a number of amino acids and nucleotides (Arahal et al., 2001), which are not considered. Because of difficulties in assigning the specificities of transport systems, we were not able to identify the transport genes for most carbon sources with certainty. In the subsequent analysis, we made the assumption that if all of the components of a pathway subsequent to transport can be identified, the organism would be able to metabolize the respective carbon sources. NAD-linked dehydrogenases were also problematic, because the queries for these enzymes often showed significant high similarity to more than one ORF in *C.*

salexigens, complicating the unambiguous assignment of enzymes catalyzing dehydrogenase reactions.

TABLE 2. Our predictions about metabolic pathways in *C. salexigens* based on the genomic sequence.

Compound	Enzymes in that could be predicted from the sequence	Predicted enzymes in catabolic pathway that could not be found in the sequence	Can we account for all the necessary metabolic enzymes?
1. Compounds observed to support growth			
D-Glucose[1]	several potential ABC and Δψ-coupled permeases; two glucokinases; PQQ-dependent glucose oxidase; gluconolactonase		Yes
D-Gluconate[2]	H^+-coupled transporter; glucokinase		Yes
D-Fructose[1,2]	PTS enzyme I; PTS Hpr; PTS Enzyme $IIBC^{Fru}$; fructose-1-phosphate kinase; fructose-6-phosphate kinase		Yes
D-Mannose[1]	hexokinase?; phosphomannose isomerase		Yes
D-Galactose[1,2]	hexokinase?; UDP-glucose pyrophosphorylase; UDP-glucose 4-epimerase	galactose-1-P uridyl transferase	No
myo- (*meso)*-Inositol[1]	inositol dehydrogenase, poor match; 2-keto-*myo*-inositol dehydrase, poor match; possible 2,3-diketo-4-d-inositol hydratase; possible 2-deoxy-5-ketogluconate-P aldolase; malonyl (methylmalonyl?) semialdehyde dehyrogenase; *myo*-inositol 1-P phosphatase (I-1-Pase), function unknown	2-d-5-ketogluconate kinase	No
Lactose[3]		β-galactosidase 6-P-β-galactosidase	No
Maltose[1,2]	α-glucosidase (maltase)		Yes
Trehalose[1]	α,α-phosphotrehalase (α-glucosidase?)	periplasmic trehalase	Maybe
Mannitol[1]	mannitol → fructose dehydrogenase		Yes
Sorbitol[1,2]	sorbitol → fructose dehydrogenase		Yes
Galactitol (dulcitol)[1]	tagatose-6-P kinase; possible tagatose (fructose?)-1,6-P_2 aldolase	transport or phosphorylation system; galactitol-P dehydrogenase	No
D-Glucarate (saccharate)[1]	glucarate dehydratase; 5-keto-4-d-D-glucarate aldolase		Yes
Sucrose[1,2]	α-glucosidase (maltase)		Yes

L-Arabinose[1,2]	ribulokinase?	arabinose ↔ ribulose isomerase; ribulose-5-P epimerase	No
D-Ribose[1,2]	ribokinase		Yes
D-Xylose[1]	xylulose kinase	xylose ↔ xylulose isomerase	No
D-Erythritol[1]	kinase?	erythritol-1-P dehydrogenase; erythrulose-1-P dehydrogenase	No
TCA cycle intermediates: citrate[1,2], α-ketoglutarate[1], succinate[1,2], fumarate[1], malate[1]	several potential Na^+ and H^+ linked tri- and dicarboxylic acid transporters, specificity can't be inferred; subsequent metabolism via TCA cycle		Yes
Acetate[1,2]	acetyl CoA synthetase; isocitrate lyase; malate synthase		Yes
Ethanol[1,2]	alcohol dehydrogenases (uncertain specificity); aldehyde dehydrogenases (uncertain specificity)		Maybe
Glycerol[1,2]	glycerol kinase; glycerol-3-phosphate dehydrogenase		Yes
D-Tartrate[1]	tartrate dehydratase α, β subunits		Yes
D,L-Glycerate[1]	D-glycerate kinase		Yes on D-glycerate
Propionate[1]	propionyl CoA synthetase; 2-methylcitrate synthase; 2-methylcitrate dehydratase; 2-methylisocitrate lyase		Yes
Malonate[1]	malonyl-CoA: ACP-SH transferase malonate decarboxylase β, γ subunits, poor match	malonate decarboxylase α, γ subunits 2-(5"-triphosphoribosyl)-3'-dephospho-CoA synthase; phosphoribosyl-dephospho-CoA transferase	No
Benzoate[2]	See Figure 1		Maybe
Protocatechuate (3,4-dihydroxy-benzoate)[2]	protocatechuate 3,4-dioxygenase α, β subunits; 3-carboxy-*cis*,*cis*-muconate cycloisomerase; 4-carboxy- muconolactone decarboxylase		Yes, if 3-oxoadipate enol-lactone can be metabolized; See Fig. 2
4-Hydroxy-benzoate[2]	4-hydroxybenzoate 3-monooxygenase		Yes; see protocatechuate
Toluene[2]	toluene 2,3-dioxygenase α subunit toluene *cis*-dihydrodiol dehydrogenase catechol 2,3-dioxygenase I, II 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase; poor match to 4-hydroxy-2-oxovalerate aldolase	toluene 2,3-dioxygenase β subunit; 2-oxopent-4-enoate hydratase; toluene 2-monooxygenase; toluene 3-monooxygenase; toluene 4-monooxygenase; toluene / o-xylene monooxygenase; 4-cresol dehydrogenase; p-hydroxybenzaldehyde dehydrogenase	No

2. Compounds observed as not being able to support growth

D-Glucono-lactone[1]	periplasmic gluconolactonase		Yes
Cellobiose[1]		cellobiose phosphorylase; cellobiose-6-P hydrolase (6-β-glucosidase)	No
Melibiose[1]		α-galactosidase	No
L-Fucose[1]	fuculose kinase; fuculose-1-P aldolase	fucose ↔ fuculose isomerase	No
L-Rhamnose[1]		rhamnulokinase; rhamnulose-1-P aldolase	No
Raffinose[1,2]		raffinose invertase (β-fructosidase); α-galactosidase	No
Inulin[1]		endo- or exoinulinase; inulin fructotransferase	No
Ribitol (adonitol)[1]	ribitol kinase ribitol dehydrogenase (specificity questionable)	ribitol-5-P dehydrogenase	Yes
Lactate (D-, L-form not specified)[1]	NAD-linked D-lactate dehydrogenase, probably fermentative; cytochrome-linked L-lactate dehydrogenase	FAD-dependent D-lactate dehydrogenase	No growth on D-lactate; possibly growth on L-lactate
Oxalate[1]	possible formyl-CoA, oxalate CoA-transferase; possible oxalyl-CoA decarboxylase	oxalyl CoA reductase	No

3. Not tested

N-Acetyl-D-glucosamine		N-acetylglucosamine-6-P deacetylase; glucosamine-6-P deaminase	No
D-Glucosamine	glucosamine-fructose-6-P aminotransferase, but this is probably biosynthetic	glucosamine-6-P deaminase	No
N-Acetyl-D-galactosamine		N-acetylgalactosamine-6-P deacetylase	No
D-Galactosamine		galactosamine-6-P isomerase / deaminase	No
Sorbose	possible sorbitol-6-P dehydrogenase	sorbose-1-reductase	No
Tagatose	possible tagatose-6-P kinase; possible tagatose (fructose?)-1,6-P_2 aldolase		Maybe
D-Altronate	altronate (mannonate?) dehydrates; 2-keto-3-d-gluconokinase		Yes
D-Tagaturonate	tagaturonate ↔ altronate oxidoreductase		Yes
D-Glucuronate		glucuronate ↔ fructuronate isomerase	No
D-Galacturonate		galacturonate ↔ tagaturonate isomerase	No
D-Mannonate	mannonate (altronate?) dehydrates; 2-keto-3-d-gluconokinase		Yes

D-Fructuronate	fructuronate → mannuronate reductase		Yes
D-Arabinose	D-arabinose-P-isomerase	specific kinase	Maybe
D-Xylulose	xylulose kinase		Maybe
Arabitol	arabitol dehydrogenase		Maybe
Hydroxypyruvate	hydroxypyruvate isomerase; tartronate semiladehyde reductase		Yes
Methanol	alcohol dehydrogenases (uncertain specificity)	PQQ-dependent methanol dehydrogenase	Maybe
Xylene		xylene monooxygenase, hydroxylase and electron transfer components; aryl alcohol dehydrogenase; aryl aldehyde dehydrogenase	No
Salicylate	salicylate 1-mono-oxygenase; poor match		Maybe
Benzene		benzene 1,2-dioxygenase α, β subunits; *cis*-1,2-dihydrobenzene-1,2-diol dehydrogenase	No
Vanillin	vanillin dehydrogenase; vanillate O-demethylase oxidoreductase		Yes, if we can explain the protocatechuate pathway
Naphthalene		1,2-dihydronaphthalene demethylase / oxidoreductase	No
Phenol	possible phenol 2-monooxygenase subunit P5	phenol 2-monooxygenase subunits P0 to P4	No

[1]Reported by Arahal et al. (2001).
[2]K. O'Connor and L.N. Csonka, unpublished observations.
[3]We found that lactose supported growth (K. O'Connor and L.N. Csonka, unpublished observation), whereas Arahal et al. (2001) reported that lactose supported acid production but not growth.

3.3.1. Glucose

Like pseudomonads and *Acinetobacter*, *C. salexigens* does not appear to have a PEP-dependent phosphotransferase system (PTS) for the uptake and phosphorylation of any sugars, except fructose (see below). It has several genes specifying ABC-type and ion co-transport driven transport systems for sugars, but the substrate specificities of these gene products is not evident from their sequences. There are two predicted genes for glucokinase that could generate glucose-6-phosphate, but these may have broader specificities for other sugars. As discussed above, the internal glucose-6-phosphate could be catabolized via glycolysis, the hexose monophosphate shunt, or the Entner-Doudoroff pathway.

C. salexigens has predicted open reading frames for the pyrroquinoline quinone (PQQ)-dependent glucose oxidase, gluconolactonase, and glucokinase, as well as the entire pathway for the synthesis of PQQ. Thus, like its pseudomonad and *Klebsiella* relatives, *C. salexigens* might be able to catabolize glucose by the non-phosphorylated route to gluconate and then by the Entner-Doudoroff pathway.

3.3.2. Fructose

We found genes for only one PTS system, which in *C. salexigens* appears to be specified by two genes. The first of these genes, which we designate as *fruA*, specifies a multi-functional protein that contains an N-terminal domain that is orthologous to the Enzyme IIA^{Fru} of *E. coli*, a central domain that is similar to the Hpr and Fpr components of PTS systems in other organisms, followed by a domain that is orthologous to Enzyme I. The second gene, *fruB*, specifies a protein that is similar to Enzyme $IIBC^{Fru}$ of *E. coli* and *P. aeruginosa*. In between *fruA* and *fruB*, there is another gene, *fruK*, which encodes a fructose-1-phosphate kinase. The similarity of these genes to loci in other organisms that have known function in fructose metabolism suggests that the PTS system specified by these genes recognizes fructose as its substrate, but in view of the generous substrate specificities of PTS systems in other organisms, this system might accept other sugars. On the basis of the sequence similarities to *P. aeruginosa* and *E. coli*, we predict that the fructose metabolic pathway in *C. salexigens* consists of uptake and phosphorylation to fructose-1-phosphate by the PTS and a second phosphorylation to fructose-1,6-bisphophosphate by the *fruK* gene product.

3.3.3. Problem Areas

Table 2 was based on the list of compounds that were tested by Arahal et al. (2001) and our laboratory (K. O'Connor and L.N. Csonka, unpublished data) for their ability to serve as carbon/energy sources for *C. salexigens*. From the sequence information, we were able to confirm the observed phenotypes of growth, or lack thereof, on various compounds, and in addition, we were able to make predictions on the ability of the organisms to metabolize several compounds that have not yet been tested. However, in a number of cases, we cannot account satisfactorily for the observed growth characteristics of the organism. In these problematic cases, *C. salexigens* could have the enzymes with the expected biochemical function, but these may not have sufficient similarity to the counterparts in organisms that were used as queries in the searches, or that *C. salexigens* may use novel biochemical reactions for some of the missing steps that have not been recognized in other organisms. In the section below, we discuss the compounds for which there is a discrepancy between the observed growth phenotype and our inferences about the metabolic pathways from the genome sequence.

D-Gluconolactone. It is not obvious why *C. salexigens* cannot grow on gluconolactone: there is a predicted periplasmic gluconolactonase, and the organism is able to metabolize gluconate, the expected product of this enzyme.

D-Galactose. This compound can clearly support growth, and we were able to find three of the enzymes of the Leloir pathway of galactose metabolism (two of which also have anabolic functions in the synthesis of galactose residues in the lipopolysaccharide). However, we were not able to see a good match for galactose-1-P uridyl transferase (*galT* gene product in *E. coli*), which converts galactose-1-P + UDP-glucose to glucose-1-P + UDP-galactose.

myo-Inositol. The *myo*-inositol catabolism has been best characterized in *B. subtilis*, and we looked for orthologs of the enzymes from that organism (Yoshida et al., 2004). We found only marginal matches for five of the enzymes of the *B. subtilis* pathway and no ortholog for a sixth enzyme (2-deoxy-5-ketogluconate kinase). It is not clear whether the poor match between to the *B. subtilis* enzymes is due to sequence divergence between the Gram-negative and Gram-positive organisms, or whether *C. salexigens* uses a different pathway. As discussed in connection with Fbpase, there is a good match to I-1-Pase, and it is not clear that this enzyme is required for the metabolism of inositol.

Lactose. There is a discrepancy in the data on the metabolism of lactose. Arahal et al. (2001) reported that *C. salexigens* cannot use lactose as sole carbon and energy source but can produce acid from it, and we found that it can grow on it as sole carbon source (K. O'Connor and L.N. Csonka, unpublished data). However, we could not detect a candidate gene for either β-galactosidase or 6-P-β-galactosidase and cannot account for the metabolism of this disaccharide.

Trehalose. There is no predicted periplasmic or cytoplasmic trehalase, but we found an α,α-phosphotrehalase, which could split trehalose-6-P to glucose and glucose-6-P. However, there is no obvious trehalose kinase, and so our ability to rationalize a trehalose metabolic pathway depends on identifying the transport and phosphorylation enzymes.

Galactitol. We did not observe a good match to a galactitol kinase or PTS, but it is possible that these reactions could be handled by enzymes that have a broader substrate specificity (e.g. the fructose PTS). In *E. coli*, galactitol-1-P is oxidized to tagatose-6-P, phosphorylated to tagatose-1,6-P_2, and cleaved to dihydroxyacetone-P + glyceraldehyde-3-P. There is no obvious galactitol-1-P dehydrogenase in *C. salexigens*, so we cannot rationalize how the organism metabolizes this hexitol.

L-Arabinose. C. salexigens is able to grow on L-arabinose, but we were not able to confirm the presence of arabinose isomerase and ribulose-5-P epimerase, which are part of the degradative pathway in *E. coli*. Two different pathways of L-arabinose catabolism that proceed through the non-phosphorylated intermediates L-arabinolactone, L-arabonate, and 2-keto-3-deoxy-L-arabonate, have been described in *Sinorhizobium meliloti* (Duncan and Fraenkel, 1979) and *Bradyrhizobium japonicum* (Pedrosa and Zancan, 1974). However, sequences of the enzymes of these pathways are not available in the ExPASy database, and we cannot tell whether L-arabinose is metabolized via the pathways found in the latter organisms.

D-xylose. In *E. coli*, D-xylose is metabolized via xylulose. There does not appear to be an appropriate isomerase in *C. salexigens*, so we cannot account for the metabolic pathway of this pentose.

D-Erythritol. A pathway for D-erythritol metabolism, consisting of a kinase, erythrose-1-P dehydrogenase, and erythrulose-1-P dehydrogenase, has been worked out in *Brucella*

(Sangari et al., 2000). Genes for the latter two enzymes were not detected in *C. salexigens*, and therefore the erythritol metabolic pathway needs to be established.

Malonate. An aerobic pathway of malonate degradation has been determined in *Klebsiella pneumoniae*, which entails a complicated set of reactions for the decarboxylation of malonate to acetate (Dimroth and Hilbi, 1997). Of the seven enzymes of this pathway, four could not be identified in *C. salexigens* with query proteins from *K. pneumoniae* or *P. putida*. *Malonomonas rubra*, which can grow on malonate anaerobically, has a Na^+ pumping membrane-bound malonate decarboxylase that is needed to generate a membrane potential (Dimroth and Hilbi, 1997). The malonate metabolic enzymes of *M. rubra*, including the Na^+ pumping decarboxylase, also did not show adequate similarities to predicted *C. salexigens* proteins.

Benzoate. This compound is oxidized to catechol (1,2-dihydroxybenzene), which can be metabolized by the *meta*- or *ortho*-cleavage pathways (Fig. 2; Nakazawa and Yokota, 1973). For the *meta*-pathway, the genome was queried with ORFs present on the *Pseudomonas putida* TOL plasmid pWWO (Horn et al., 1991), and for the *ortho*-pathway, the genome was searched against the enzymes from *Acinetobacter* sp. (Barbe et al., 2004).

Although *C. salexigens* can grow on benzoate as a single carbon source (K. O'Connor and L.N. Csonka, unpublished data), we cannot account satisfactorily for the metabolism of this important compound. In common pathway to from benzoate to catechol, we could not find a good ortholog for the β-subunit of benzoate 1,2-dioxygenase. Furthermore, 2-oxopent-4-enoate hydratase seems to be missing from the catechol *meta*-cleavage pathway, and muconolactone isomerase and the A subunit of 3-oxodaipate CoA transferase are missing from the *ortho*-pathway. In addition, some of the matches are very borderline, at least to the *meta* pathway enzymes specified by the *P. putida* TOL plasmid and the *ortho* pathway enzymes found in *Acinetobacter*. However, overall the matches are better for the *ortho*-pathway enzymes from *Acinetobacter* than for the *meta*-pathway enzymes from the TOL plasmid. Therefore, we feel that it is more likely that *C. salexigens* uses the former route than the latter one. However, it is possible that we could find better quality matches if we searched the *C. salexigens* genome with benzoate degradative enzymes from a broader spectrum of organisms. In *Acinetobacter*, the benzoate degradative enzymes are arranged as a tight cluster (Barbe et al., 2004), but this does not seem to be the case in *C. salexigens*.

Protocathecuate (3,4-dihydroxybenzoic acid). This compound is metabolized to 3-oxoadipate enol-lactone, an intermediate in the catechol *ortho*-cleavage pathway (Fig. 2). We can find excellent orthologs for the three enzymes that bring protocatechuate into the *ortho*-pathway. So, provided that the latter pathway exists in *C. salexigens*, we could account for the necessary enzymes of protocatechuate metabolism. Again, unlike in *Acinetobacter* (Barbe et al., 2004), the three unique enzymes of protocatechuate metabolism are not adjacent to each other in our organism.

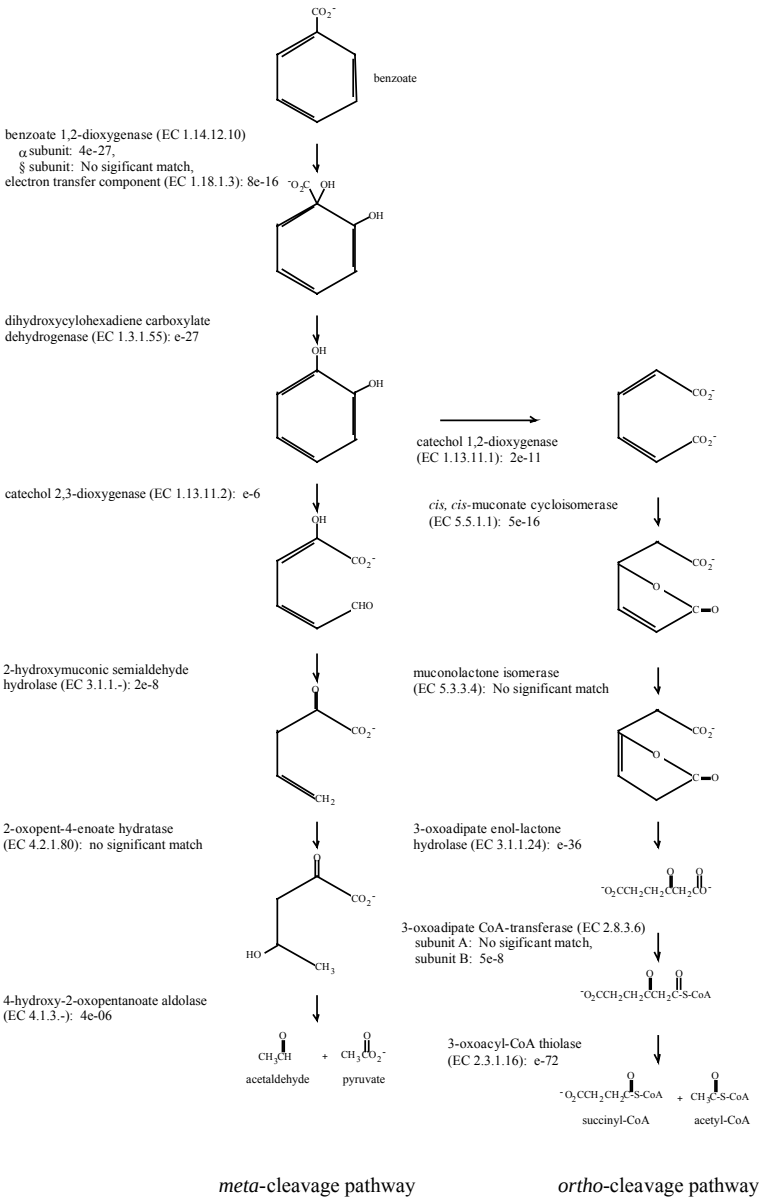

Figure 2. Benzoate metabolic pathways. The expect values indicate the tblastn similarity scores of *C. salexigens* ORFS against queries from genes from the *P. putida* TOL plasmid (for the reactions from benzoate to catechol and for the *meta* pathway) and from *Acenitobacter* (for the *ortho* pathway).

Toluene. There are five pathways of toluene degradation, beginning with the five enzymes xylene monooxygenase, toluene 2-monooxygenase, toluene 3-monooxygenase, toluene 4-monooxygenase, and toluene 1,2-dioxygenase (the University of Minnesota Biocatalysis / Biodegradation Database; http://umbbd.ahc.umn.edu). We did not detect acceptable similarities to enzymes of the first four pathways, but were able to recognize some of the enzymes of the fifth pathway, which goes through toluene-*cis*-1,2-dihydrodiol, 3-methylcatechol, *cis*,*cis*-2-hydroxy-6-oxohept-2,4-dienoate, *cis*-2-hydroxypenta-2,4-dienoate, and 4-hydroxy-2-oxovalerate to acetate, acetaldehyde plus pyruvate. We found good matches to all of the enzymes of this pathway, except for the β subunit of the first enzyme (toluene 2,3-dioxygenase) and the last enzyme (2-oxopent-4-enoate hydratase), and therefore we cannot properly account for the toluene metabolic pathway.

Xenobiotic compounds. We found suggestions for the presence of ORFs specifying various enzymes that are involved in the metabolism of the following xenobiotic compounds: acrylamide, benzamide, cyclohexanol, 1,2-dichloroethane, ethylbenzene, nitrophenyl phosphate, and phthalate. However, the complete metabolic pathways of these compounds could not be identified. Therefore, *C. salexigens* may be able to partially degrade these compounds, but it probably would not be able to grow on them as single carbon sources.

4. 1C metabolism

4.1. RIBULOSE BISPHOSPHATE CARBOXYLASE, LARGE SUBUNIT

The *C. salexigens* chromosome contains an ORF that shows high similarity to the large subunit of the Calvin-Benson-Bassham cycle ribulose bisphosphate carboxylase / oxygenase (RubisCO). Orthologous proteins exist in a variety of organisms, and among these, the *C. salexigens* sequence has highest similarity (E values in from e-119 to 5e-96) to ORFs in *Bordetella bronchiseptica* RB50, ND *Mesorhizobium loti* MAFF30309, and *Sinorhizobium meliloti* 1021. *B. subtilis* has a related enzyme, encoded by the *mtnW* (*ykrW*) gene that has been shown to carry out an enolase reaction in the S-adenosylmethionine to methionine salvage pathway (Ashida et al., 2003; Sekowska et al., 2004). However, we have not been able to find good matches for the other enzymes of this pathway, as it exists in *B. subtilis*. A RubisCO-like protein is also found in the *Chlorobaculum tepdium* and related green sulfur bacteria (Hanson and Tabita, 2001). Although these organisms are photoautotrophic, they use the reductive TCA cycle for CO_2 fixation instead of RubisCO. Rather, this enzyme has been suggested to have an as of yet undetermined role in sulfur metabolism or oxidative stress response (Hanson and Tabita, 2001). These RubisCO-like proteins in *B. subtilis* and *Chlorobaculum* do not have several amino acid residues that are required for carboxylase / oxygenase activity and they have been shown to lack these enzymatic activities. The RubisCO-like protein in *C. salexigens* contains non-conservative substitutions at 6 of 19 amino acid residues that are required for the binding of the carboxylase substrates and it also has a deletion of

11 residues in "loop 6", which is essential for carboxylase activity (Ashida et al., 2003). Thus, it is unlikely that this enzyme operates as a carboxylase in *C. salexigens*, but we have no insights as to its true function. We have not observed an ortholog for the small subunit of RubisCO. Interestingly, *C. salexigens* appears to have gene for phosphoribulokinase, which is a member of the Calvin-Benson-Bassham cycle in CO_2 fixing organisms. However, there is no evidence for sedoheptulose-1,7-bisphosphatase, providing a second reason why it is unlikely that *C. salexigens* can fix CO_2 with the RubisCO-like protein.

4.2. CARBON MONOXIDE DEHYDROGENASE

There were good matches to the three subunits of the aerobic type carbon monoxide dehydrogenase found in *P. aeruginosa*; as is the case with many of the our predictions, the function of this enzyme needs to be addressed experimentally.

4.3. FORMALDEHYDE DEHYDROGENASE

The genome contains information for a glutathione-dependent formaldehyde dehydrogenase and two glutathione-independent formaldehyde dehydrogenases. It is not known whether *C. salexigens* can metabolize formaldehyde, but a closely related *Halomonas elongata* strain has been observed to be able to do so (Azachi et al., 1995). Thus, we can account for the metabolism of formaldehyde in *C. salexigens*.

4.4. FORMATE DEHYDROGENASE

We found evidence for two selenocysteine-containing formate dehydrogenases, one of which may contain molybdenum and the second may contain tungsten. As *C. salexigens* cannot grow fermentatively, these formate dehydrogenases are presumably coupled to an electron acceptor (O_2, NO_3^-), but we cannot deduce the nature of the electron acceptor from the amino acid sequence. We observed eight genes that could specify the components of formate hydrogen lyase (FdhF, HycA through HycG), but because these showed marginal similarity to orthologs from *E. coli* (e-13 to e-19), we are not certain of this identification. There was no evidence of a hydrogenase, which could either form or consume H_2.

5. Summary

From the draft genomic sequence of *C. salexigens*, we were able to find good indication that the organism has all of the enzymes of glycolysis, hexose monohophoshate shunt, Entner-Doudoroff pathway, and the TCA cycle. We were able to rationalize the pathways of metabolism of many of the common sugars and mono- and dicarboxylic acids. We could account only partially for the metabolic pathway of benzoate, 4-hydroxybenzoate, and 3,4-hydroxybenzoate. The sequence information gave hints that *C. salexigens* might be able to metabolize a number of other aromatic and xenobiotic

compounds at least partially. The latter suggestion, which needs to be followed up experimentally, presents the exciting possibility that *C. salexigens* might be exploited for the biological cleanup of polluted environments that are also high in salinity.

Acknowledgments

We would like thank Dan Fraenkel and Robert Tabita for helpful discussions. Work in the Csonka laboratory on the determinants of halophilism in *C. salexigens* has been supported by the Life in Extreme Environments Program of the National Science Foundation (Grant MCB-9978253). B.W.G. would like to thank the Howard Hughes Medical Institute grant (52005125) and Hiram College for supporting undergraduate involvement in genomics research.

References

Arahal, D.R., García, M.T., Vargas, C., Cánovas, D., Nieto, J.J. and Ventosa, A. (2001) *Chromohalobacter salexigens* sp. nov., a moderately halophilic species that includes *Halomonas elongata* DSM 3043 and ATCC 33174. Int. J. Syst. Evol. Microbiol. **51,** 1457-1462.

Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N. and Yokota, A. (2003) A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. Science **302,** 286-290.

Azachi, M., Henis, Y., Oren, A., Gurevich, P. and Sarig, S. (1995) Transformation of formaldehyde by *Halomonas* sp. Can. J. Microbiol. **41,** 548-553.

Barbe, V., Vallenet, D., Fonknechten, N., Kreimeyer, A., Oztas, S., Labarre, L., Cruveiller, S., Robert, C., Duprat, S., Wincker, P., Ornston, L.N., Weissenbach, J., Marlière, P., Cohen, G.N. and Medigue, C. (2004) Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res. **32,** 5766-5779.

Chen, L. and Roberts, M.F. (2000) Overexpression, purification, and analysis of complementation behavior of the *E. coli* SuhB protein: comparison with bacterial and archaeal inositol monophosphatases. Biochemistry **39,** 4145-4153.

Cronan, J.E. and LaPorte, D. (1996) Tricarboxylic acid cycle and glyoxylate bypass, In: F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (eds.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* ASM Press, Washington D.C., pp. 206-216.

Dimroth, P. and Hilbi, H. (1997). Enzymatic and genetic basis for bacterial growth on malonate. Mol. Microbiol. **25,** 3-10.

Donahue, J.L., Bownas, J.L., Niehaus, W.G. and Larson, T.J. (2000) Purification and characterization of *glpX*-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*. J. Bacteriol. **182,** 5624-5627.

Duncan, M.J. and Fraenkel, D.G. (1979) α-Ketoglutarate dehydrogenase mutant of *Rhizobium meliloti*. J. Bacteriol. **137,** 415-419.

Fraenkel, D.G. (1996). Glycolysis, In: F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (eds.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* ASM Press, Washington D.C., pp. 189-198.

Fujita, Y., Yoshida, K.-I., Miwa, Y., Yanai, N., Nagakawa, E. and Kasahara, Y. (1998) Identification and expression of the *Bacillus subtilis* fructose-1,6-bisphosphatase gene (fbp). J. Bacteriol. **180,** 4309-4313.

Gottschalk, G. (1985) *Bacterial Metabolism*. Springer-Verlag, New York.

Hanson, T.E. and Tabita, F.R. (2001) A ribulose-1,5-bisphosphate carboxylase / oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. Proc. Natl. Acad. Sci. USA **98,** 4397-4402.

Horn, J.M., Harayama, S. and Timmis, K.N. (1991) DNA sequence determination of the TOL plasmid (pWWO) *xylGFJ* genes of *Pseudomonas putida*: implications for the evolution of aromatic catabolism. Mol. Microbiol. **5,** 2459-2474.

Lin, E.C.C. (1996) Dissimilatory pathways for sugars, polyols, and carboxylates, In: F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (eds.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. ASM Press, Washington D.C., pp. 307-342.

Movahedzadeh, F., Rison, S.C.G., Wheeler, P.R., Kendall, S.L., Larson, T.J. and Stokcer, N.G. (2004) The *Mycobacterium tuberculosis* Rv1099c gene encodes a GlpX-like class II fructose 1,6-bisphosphatase. Microbiology **150,** 3499-3505.

Nakazawa, T. and Yokota, T. (1973) Benzoate metabolism in *Pseudomonas putida* (*arvilla*) mt-2: demonstration of two benzoate pathways. J. Bacteriol. **115,** 262-267.

Pedrosa, F.O. and Zancan, G.T. (1974) L-Arabinose metabolism in *Rhizobium japonicum*. J. Bacteriol. **119,** 336-338.

Sangari, F.J., Agüero, J. and García-Lobo, J.M. (2000) The genes for erythritol catabolism are organized as an inducible operon in *Brucella abortus*. Microbiology **146,** 487-495.

Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H. and Imanaka, T. (2004) Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other thermophiles. J. Bacteriol. **186,** 5799-5807.

Sekowska, A., Denervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A. and Danchin, A. (2004) Bacterial variations on the methionine salvage pathway. BMC Microbiol **4,** 9-25.

Stec, B., Yang, H., Johnson, K.A., Chen, L. and Roberts, M.F. (2000) MJ0109 is an enzyme that is both an inositol monophosphatase and the "missing" archaeal fructose-1,6-bisphosphatase. Nature Struct. Biol. **7,** 1046-1050.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K.S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature **406,** 959-964.

Yoshida, K., Yamaguchi, M., Ikeda, H., Omae, K., Tsurusaki, K. and Fujita, Y. (2004) The fifth gene of the *iol* operon of *Bacillus subtilis*, *iolE*, encodes 2-keto-myo-inositol dehydratase. Microbiology **150,** 571-580.