

# Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator

## A CASE STUDY IN *SHEWANELLA ONEIDENSIS*<sup>\*[5]</sup>

Received for publication, June 3, 2011, and in revised form, August 16, 2011. Published, JBC Papers in Press, August 17, 2011, DOI 10.1074/jbc.M111.267963

Semen A. Leyn<sup>†§</sup>, Xiaoqing Li<sup>‡</sup>, Qingxiang Zheng<sup>¶</sup>, Pavel S. Novichkov<sup>||</sup>, Samantha Reed<sup>\*\*</sup>, Margaret F. Romine<sup>\*\*</sup>, James K. Fredrickson<sup>\*\*</sup>, Chen Yang<sup>¶1</sup>, Andrei L. Osterman<sup>‡</sup>, and Dmitry A. Rodionov<sup>†§2</sup>

From the <sup>†</sup>Sanford-Burnham Medical Research Institute, La Jolla, California 92037, the <sup>§</sup>Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia, the <sup>¶</sup>Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China, the <sup>||</sup>Lawrence Berkeley National Laboratory, Berkeley, California 94720, and the <sup>\*\*</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

**Background:** Different regulatory strategies are utilized by bacteria to control central carbohydrate metabolism.

**Results:** Transcriptional factor HexR is a global regulator of the central carbohydrate metabolism genes in various groups of proteobacterial genomes.

**Conclusion:** HexR in *Shewanella* is a repressor/activator of the glycolytic/gluconeogenic genes.

**Significance:** Integration of the comparative genomics and experimental approaches is efficient for reconstruction of transcriptional regulons in bacteria.

Bacteria exploit multiple mechanisms for controlling central carbon metabolism (CCM). Thus, a bioinformatic analysis together with some experimental data implicated the HexR transcriptional factor as a global CCM regulator in some lineages of Gammaproteobacteria operating as a functional replacement of the Cra regulator characteristic of Enterobacteriales. In this study, we combined a large scale comparative genomic reconstruction of HexR-controlled regulons in 87 species of Proteobacteria with the detailed experimental analysis of the HexR regulatory network in the *Shewanella oneidensis* model system. Although nearly all of the HexR-controlled genes are associated with CCM, remarkable variations were revealed in the scale (from 1 to 2 target operons in Enterobacteriales up to 20 operons in Aeromonadales) and gene content of HexR regulons between 11 compared lineages. A predicted 17-bp pseudo-palindrome with a consensus tTGTAATwwwATTACa was confirmed as a HexR-binding motif for 15 target operons (comprising 30 genes) by *in vitro* binding assays. The negative effect of the key CCM intermediate, 2-keto-3-deoxy-6-phosphogluconate, on the DNA-regulator complex formation was verified. A dual mode of HexR action on various target promot-

ers, repression of genes involved in catabolic pathways and activation of gluconeogenic genes, was for the first time predicted by the bioinformatic analysis and experimentally verified by changed gene expression pattern in *S. oneidensis*  $\Delta$ hexR mutant. Phenotypic profiling revealed the inability of this mutant to grow on lactate or pyruvate as a single carbon source. A comparative metabolic flux analysis of wild-type and mutant strains of *S. oneidensis* using [<sup>13</sup>C]lactate labeling and GC-MS analysis confirmed the hypothesized HexR role as a master regulator of gluconeogenic flux from pyruvate via the transcriptional activation of phosphoenolpyruvate synthase (PpsA).

Fine-tuning of the carbohydrate catabolic pathway expression is key to successful adaptation of microorganisms that occupy niches with variable carbon availability. *Escherichia coli* and related Enterobacteria utilize two global transcription factors (TFs),<sup>3</sup> Crp and Cra (FruR), to control central and peripheral carbohydrate metabolism (1). The cAMP receptor protein, Crp, mediates catabolic repression of target genes in the presence of high levels of glucose, a preferable carbon source for Enterobacteria. In the presence of cAMP, Crp positively regulates its target genes by binding the cognate DNA sites. cAMP is generated by adenylate cyclase that is activated by components of the glucose-specific phosphotransferase system in the absence of glucose (2). The LacI protein family catabolite repressor/activator, Cra, was initially characterized as the fructose repressor, FruR, because it negatively regulates transcription of the fructose utilization operon (3). Later, Cra was characterized as a pleiotropic transcriptional regulator that

\* This work was supported in part by the United States Department of Energy, Office of Science, as part of the Genomic Science Program under Contract DE-AC05-76RLO with Pacific Northwest National Laboratory (Genomic Science Program Foundational Scientific Focus Area), Sanford-Burnham Medical Research Institute, Lawrence Berkeley National Laboratory Grant DE-SC0004999, National Science Foundation Grant DBI-0850546, Russian Foundation for Basic Research Grant 10-04-01768, and the Russian Academy of Sciences Program Molecular and Cellular Biology.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables S1–S4 and Figs. S1–S4.

<sup>1</sup> Supported by National Science Foundation of China Grants 30970035 and 31070033 and Knowledge Innovation Program of Chinese Academy of Sciences Grant KSCX2-EW-G-5.

<sup>2</sup> To whom correspondence should be addressed: Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-646-3100 (Ext. 3082); Fax: 858-795-5249; E-mail: rodionov@sanfordburnham.org.

<sup>3</sup> The abbreviations used are: TF, transcription factor; EMP pathway, Embden Meyerhof-Parnas pathway; ED pathway, Entner-Doudoroff pathway; KDPG, 2-keto-3-deoxy-6-phosphogluconate; FPA, fluorescence polarization assay.

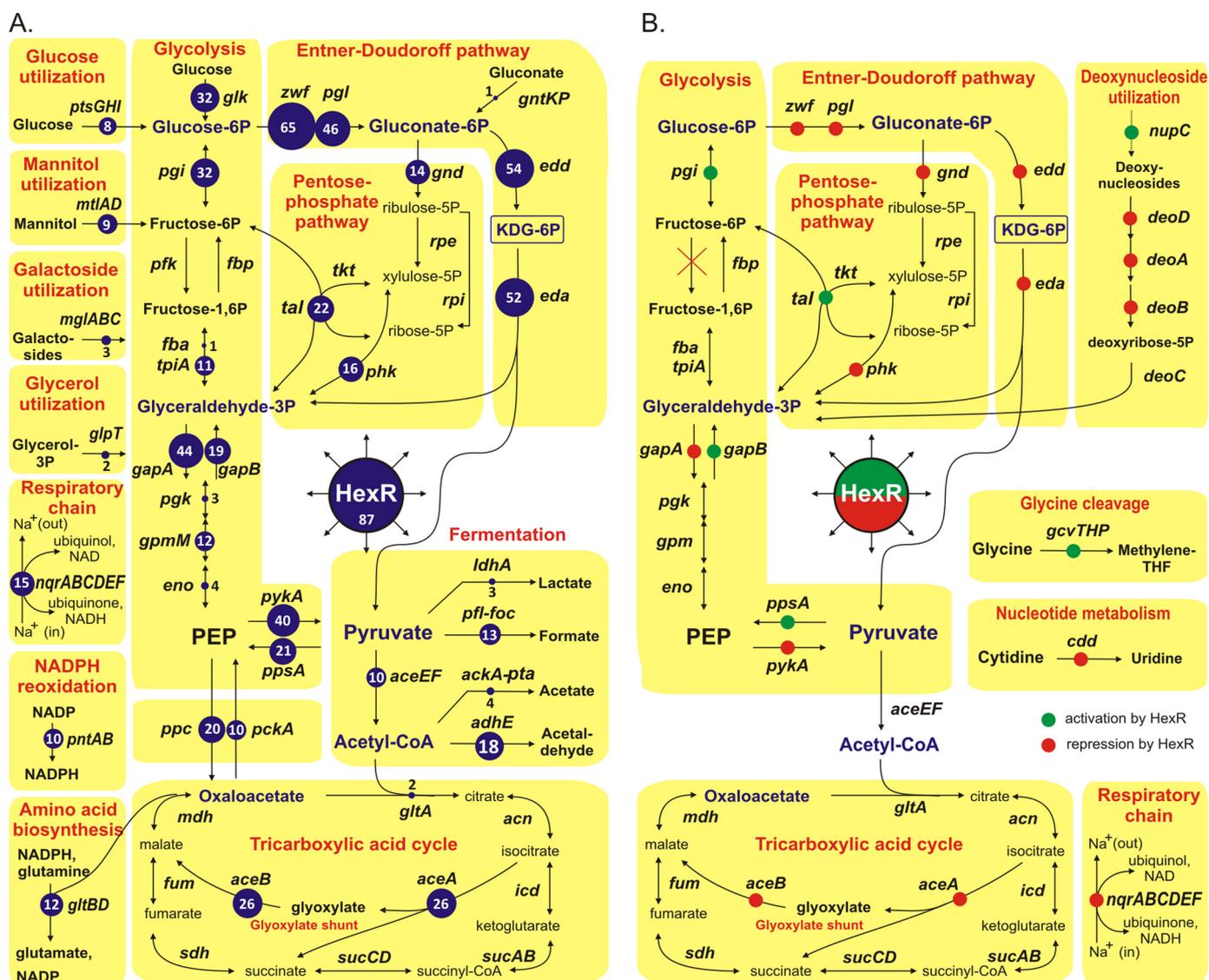


FIGURE 1. Metabolic context of the reconstructed HexR regulons in Proteobacteria (A) and *S. oneidensis* MR-1 (B). Numbers in blue circles show the number of genomes where the target gene is preceded by a candidate HexR-binding site.

represses multiple genes from the central glycolytic pathways, namely the EMP and ED pathways, and activates genes involved in gluconeogenesis and oxidative phosphorylation (4). DNA binding activity of Cra is triggered by two central intermediates of the EMP pathway, fructose 1-phosphate and fructose 1,6-biphosphate. Comparative genomic analysis of the Cra (FruR) regulon revealed substantial variability in regulon gene content among Enterobacteriales and that it acted as a local regulator of the fructose utilization operon in other  $\gamma$ -proteobacteria such as *Vibrio* and *Pseudomonas* species (5).

A different regulatory strategy is utilized to control central carbohydrate metabolism in the genus *Pseudomonas*, which favors the utilization of organic acids and amino acids over various other carbon sources (6). In contrast to Enterobacteria, the *Pseudomonas* EM glycolytic pathway is nonfunctional due to the absence of 6-phosphofructokinase (7). Degradation of glucose proceeds ultimately via the ED pathway, where 6-phosphogluconate and KDPG are key intermediates. Expression of all genes encoding the ED pathway in *Pseudomonas putida*,

including glucokinase (*glk*), glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphogluconolactonase (*pgl*), 6-phosphogluconate dehydrogenase (*edd*), KDPG aldolase (*eda*), and glyceraldehyde-3-phosphate dehydrogenase (*gap-1*), is negatively regulated by HexR (8, 9). Two monomers of HexR bind to imperfect palindromic sites with consensus sequence of TTGTN<sub>7-8</sub>ACAA in the promoter regions of the *zwf*, *edd*, and *gap-1* genes. Binding of the ED pathway intermediate KDPG to HexR releases the repressor from its target sites (8–11).

Previously, we have applied the integrated genomic and experimental approaches to predict and validate novel metabolic pathways and transcriptional regulons involved in carbohydrate utilization in the *Shewanella* genus of  $\gamma$ -proteobacteria (12, 13). The obtained genomic encyclopedia of sugar utilization included 17 distinct peripheral pathways with committed TFs (e.g. the *N*-acetylglucosamine utilization pathway controlled by NagR regulon) and a core set of central carbohydrate metabolism genes, of which 12 genes were predicted to be controlled by the HexR regulator (Fig. 1B). In this study, we expand

## Proteobacterial HexR Catabolite Repressor/Activator

this analysis toward comparative genomic reconstruction of the HexR regulon in all sequenced Proteobacteria. We report the identification of DNA-binding motifs for HexR regulators and provide a detailed description of corresponding regulons in the genomes of 62  $\gamma$ -proteobacteria and 25  $\beta$ -proteobacteria. The comparative analysis of reconstructed regulons revealed considerable variability in the regulon content between the analyzed Proteobacteria. By correlating the binding site position within promoter regions with expression patterns of downstream genes, it was possible to predict the activation mode for HexR regulation of several target genes in *Shewanella* spp. A combination of *in vivo* and *in vitro* experimental techniques was used to validate the predicted HexR-dependent regulatory network in *Shewanella oneidensis* MR-1.

### EXPERIMENTAL PROCEDURES

**Bioinformatics Techniques and Resources**—Genomic sequences were obtained from GenBank<sup>TM</sup> (14). Identification of orthologs and gene neighborhood analysis were performed in Microbes OnLine (15). Functional annotations of genes involved in central carbohydrate metabolism and related pathways were derived from the SEED comparative genomic database (16). Phylogenetic trees were built using the maximum likelihood method in the PHYLIP Package and visualized with Dendroscope (17). Sequence alignments were made with MUSCLE (18). Distant homologs were identified using pBLAST (19).

We used a well established comparative genomics method of regulon reconstruction (reviewed in Ref. 5) implemented in the RegPredict webserver (regpredict.lbl.gov) (20). We started from the genomic identification of a reference set of genomes that encodes HexR orthologs (according to the phylogenetic analysis of HexR proteins on supplemental Fig. S1). To find the conserved DNA-binding motif for HexR in each group of phylogenetically related genomes, we used initial training sets of known HexR targets from *Pseudomonas* spp., and then we updated each set by the most likely HexR-regulated genes confirmed by the comparative genomics tests as well as the functional and genome context considerations.

In each of 13 studied groups of Proteobacteria, an iterative motif detection procedure implemented in the RegPredict web tool was used to identify common regulatory DNA motifs in a set of upstream gene fragments and to construct the motif recognition profiles as described previously (21). For each clade of HexR proteins on the phylogenetic tree, we used a separate training gene set. The initial recognition profile was used to scan the genomes in this clade and to predict novel genes in the regulon. Scores of candidate sites were calculated as the sum of positional nucleotide weights. The score threshold was defined as the lowest score observed in the training set. The conserved regulatory interactions with high scored binding sites that involve target genes involved in central carbohydrate metabolism were included in the reconstructed HexR regulons. Candidate sites associated with new members of regulon were added to the training set, and the respective group profile was rebuilt to improve search accuracy. Sequence logos for the derived group-specific DNA-binding motifs were drawn using WebLogo package (22). The details of the reconstructed HexR

regulons are captured and displayed in RegPrecise, a specialized database of bacterial regulons (regprecise.lbl.gov) (23).

**Cloning of hexR and Overproduction in E. coli**—The *hexR* gene (SO2490) from *S. oneidensis* MR-1 was PCR-amplified using as a template a pET-derived vector encoding SO2490 fused to an N-terminal His<sub>6</sub> tag (24) and the following primers: 5'-CGATCATGGATCCatgaataccctagaaaagg (forward) and 5'-CTGCAGTCGAAGCTTttagcatatcggatc (reverse). The PCR products were digested with BamHI/HindIII, and the resulting 861-bp fragment encoding *hexR* was then cloned into the pSMT3 expression vector (25). The obtained pSMT3 vector encodes a fusion between the HexR protein and N-terminal hexahistidine Smt3 polypeptide (a yeast SUMO ortholog), allowing us to enhance protein solubility. The resulting construct was transformed in *E. coli* BL21/DE3 cells (Invitrogen). The cells were grown in 50 ml of LB medium to an  $A_{600} = 1.2$  at 20 °C containing 0.3 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside and harvested after 12 h. The recombinant Smt3-His<sub>6</sub>-HexR protein was purified using a rapid nickel-nitrilotriacetic acid-agarose column as described previously (26). Briefly, harvested cells were resuspended in 20 mM HEPES buffer (pH 7) containing 100 mM NaCl, 0.03% Brij 35, and 2 mM  $\beta$ -mercaptoethanol supplemented with 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor mixture (Sigma). Lysozyme was added to 1 mg/ml, and the cells were lysed by freeze-thaw cycle followed by sonication. The resulting crude extract was heated for 30 min at 70 °C. After centrifugation, the supernatant was supplemented with Tris-HCl buffer (pH 8) to achieve a final concentration of 50 mM and then was loaded onto a nickel-nitrilotriacetic acid-agarose minicolumn (0.3 ml). After washing the column with the starting buffer containing 1 M NaCl and 0.3% Brij-35, bound proteins were eluted with 0.3 ml of the starting buffer supplemented with 250 mM imidazole. The eluted protein was suspended in a solution of 20 mM HEPES (pH 7.0), 1 mM DTT, and 0.5 mM EDTA. Protein size, expression level, distribution between soluble and insoluble forms, and extent of purification were assessed with 12% SDS-polyacrylamide gels.

**DNA Binding Assays**—Interaction of purified recombinant-tagged HexR protein with its cognate DNA-binding sites in *S. oneidensis* MR-1 was assessed by two techniques, electrophoretic mobility shift assay and fluorescence polarization assay, using two sets of synthetic double-stranded oligonucleotides 3'-labeled with either biotin or 6-carboxyfluorescein, respectively. The dsDNA fragments containing the predicted HexR-binding sites were obtained by annealing custom-synthesized complementary oligonucleotides at a 1:10 ratio of labeled to unlabeled complementary oligonucleotides (supplemental Table S1).

For EMSAs, the 49-bp biotin-labeled dsDNA fragments (0.1 nM) were incubated with 500 nM purified HexR in a total volume of 20  $\mu$ l as described previously (27). The binding buffer contained 20 mM Tris-HCl (pH 8.0), 150 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM EDTA, 0.05% Nonidet P-40, and 2.5% glycerol. Poly(dI-dC) was added as a nonspecific competitor DNA at  $\sim 10^4$ -fold molar excess over labeled target DNA to reduce nonspecific binding. After 25 min of incubation at 37 °C, the reaction mixtures were separated by electrophoresis on a 5%

native polyacrylamide gel in 0.5× Tris borate/EDTA (100 min, 90 V, room temperature). The DNA was transferred by electrophoresis (30 min at 380 mA) onto a Hybond-N<sup>+</sup> membrane and fixed by UV cross-linking. Biotin-labeled DNA was detected with the LightShift chemiluminescent EMSA kit. The effect of phosphorylated intermediates of the ED pathway on HexR-DNA binding was tested by addition of 2-keto-3-deoxy-6-phosphogluconate, glucose 6-phosphate, 6-phosphogluconate, and phosphoenolpyruvate to the incubation mixture (2 mM).

For FPA assays, the 27-bp fluorescence-labeled dsDNA fragments (1 nM) were incubated with the increasing concentrations of purified HexR (10–1000 nM) in 100 μl of reaction mixture in 96-well black plates (VWR, Radnor, PA). Binding buffer contained 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.3 mg/ml BSA, and 1 μg of poly(dI-dC). Fluorescence measurements were taken on Beckman DTX 880 multimode plate reader with excitation and emission filters at 495 and 520 nm, respectively. Background fluorescence from buffer was subtracted, and the fluorescence polarization value was defined as follows:  $FP = (I_1 - G \cdot I_2)/(I_1 + G \cdot I_2)$ , where  $I_1$  and  $I_2$  are the fluorescence intensities measured in the parallel and perpendicular directions respective to the orientation of the excitation polarizer, and  $G$  is a correction factor (28).

**Phenotypic Analysis and Metabolite Measurements**—Chromosomal in-frame deletion mutants of *hexR* (SO2490) or *ppsA* (SO2644) genes were constructed using previously published methods (13). *S. oneidensis* MR-1 wild type (WT) and  $\Delta hexR$  and  $\Delta ppsA$  mutant strains were precultured on LB medium to late exponential growth phase, and washed with M1 minimal medium without any carbon sources. Duplicate 50-ml cultures were inoculated to the same absorbance ( $A_{600} = 0.01$ ) in M1 minimal medium supplemented with 10 mM *N*-acetyl-D-glucosamine or 20 mM of D-glycerate, DL-lactate, or pyruvate as the sole carbon source and incubated at 30 °C. Cell growth was monitored spectrophotometrically at 600 nm. Lactate, glycerate, acetate, and pyruvate were detected in culture supernatants by high pressure liquid chromatography with the Agilent model 1100 instrument equipped with Shodex RSpak KC-811. A mobile phase consisting of 6 mM HClO<sub>4</sub> at a flow rate of 1.0 ml/min was used, and the column was operated at 50 °C.

**RT-PCR Analysis**—Genomic RNA was isolated from *S. oneidensis* MR-1 WT and  $\Delta hexR$  cells grown in minimal medium supplied with either D-glycerate or inosine and collected at the same absorbance ( $A_{600} = 0.5$ ) using the RNA purification kit from Promega. Reverse transcription of total RNA was performed with random primers using the iScript cDNA synthesis kit from Bio-Rad. RT-PCR was performed using the SYBR<sup>®</sup> GreenER<sup>™</sup> qPCR SuperMix Universal kit from Invitrogen. Transcript levels were measured for 13 candidate regulated genes, and the results were normalized to the expression level of 16 S mRNA, and the gene expression fold change was calculated using the  $2^{-\Delta CT}$  method (29) as a ratio of normalized mRNA levels in  $\Delta hexR$  mutant and WT MR-1 strains.

**GC-MS and Metabolic Flux Analysis**—Cells were grown on a mixture containing 4 mM <sup>13</sup>C-labeled lactate, 16 mM unlabeled lactate, and 20 mM glycerate or on lactate only containing 4 mM [<sup>13</sup>C]lactate and 16 mM unlabeled lactate. Aliquots of cultures were harvested at mid-exponential growth phase ( $A_{600}$  of

0.3). After centrifugation, the cell pellet was washed with 1 ml of 0.9% (w/v) NaCl and then hydrolyzed in 200 μl of 6 M HCl at 105 °C for 24 h. The filtrate was dried in a vacuum centrifuge at room temperature and derivatized at 85 °C for 1 h in 120 μl of pyridine and 30 μl of *N*-methyl-*N*-(*tert*-butyldimethylsilyl) trifluoroacetamide. After filtration, 5 μl of derivatized sample was injected into Agilent 6890-5973 GC-MS system with a HP-5MS column (30 m × 0.25 mm × 0.25 μm). GC oven temperature was programmed from 60 to 180 °C at 5 °C per min and from 180 to 260 °C at 10 °C per min. The flow rate of carrier gas (helium) was set at 1 ml min<sup>-1</sup>. The mass spectrometer was operated in the electron impact mode at 70 eV.

The GC-MS data were analyzed as described previously (30). Briefly, prior to analysis of cellular metabolism, mass spectra of the derivatized amino acids were corrected for the natural abundance of all stable isotopes. From the mass isotopomer distribution vectors of the phenylalanine, tyrosine, alanine, valine, aspartate, and threonine, the isotopomer distribution vectors of their respective precursor intermediates (*i.e.* phosphoenolpyruvate, pyruvate, and oxaloacetate) could be easily derived. The fractions of phosphoenolpyruvate originating from pyruvate ( $f_{PEP \leftarrow PYR}$ ) and from oxaloacetate ( $f_{PEP \leftarrow OAA}$ ) can then be determined by using Equation 1 implemented in the MATLAB program.

$$\begin{pmatrix} f_{PEP \leftarrow PYR} \\ f_{PEP \leftarrow OAA} \end{pmatrix} = \frac{PEP_{12} - glycerate_{12}}{\begin{pmatrix} PYR_{12} - glycerate_{12} \\ OAA_{12} - glycerate_{12} \end{pmatrix}^T} \quad (\text{Eq. 1})$$

In Equation 1, the division is a left-hand matrix division.

For quantification of intracellular carbon fluxes in *S. oneidensis* grown on lactate, a bioreaction network was constructed based on the Internet-accessible genome database. This network includes the reactions of the gluconeogenic pathway and the tricarboxylic acid cycle, as well as the reactions catalyzed by phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase, and malic enzyme. The glyoxylate shunt and the pyruvate kinase were identified to be inactive in this study; thus they are omitted from the bioreaction network for net flux analysis. From the bioreaction network, a stoichiometric matrix containing 21 unknown fluxes and 19 metabolite balances was constructed. Net fluxes were then calculated based on three different data sets: (i) substrate uptake and product formation rates; (ii) macromolecular biomass composition; and (iii) the calculated  $f_{PEP \leftarrow PYR}$  and  $f_{PEP \leftarrow OAA}$ . The macromolecular biomass composition of *S. oneidensis* MR-1 was taken from Ref. 31. The carbon flux distribution in the bioreaction network was determined with a MATLAB-based program by minimizing the sum of the weighted square residuals of the constraints from both metabolite balances and flux ratios as described previously (30).

## RESULTS

### Genomic Reconstruction of HexR Regulons in Proteobacteria

**Phylogenetic Distribution of hexR Orthologs**—Orthologs to the *P. putida hexR* were identified by BLAST searches against a nonredundant set of sequenced bacterial genomes and used to analyze the phylogenetic relationships between these proteins

## Proteobacterial HexR Catabolite Repressor/Activator

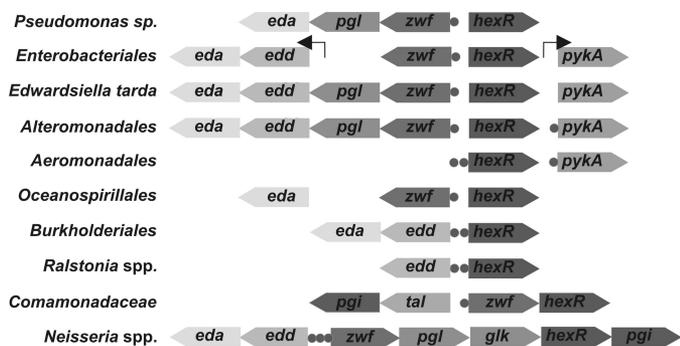


FIGURE 2. Genomic organization of the *hexR*-containing loci and HexR-regulated genes in representative bacterial genomes. Genes and candidate HexR-binding sites are shown by arrows and circles, respectively.

(supplemental Fig. S1). A single copy of *hexR* gene was detected in most lineages of  $\gamma$ - and  $\beta$ -proteobacteria but not in other taxonomic groups (supplemental Table S2). Among the  $\gamma$ -proteobacteria, *hexR* is missing in all of the Pasteurellales, Xanthomonadales, and Moraxellaceae lineages and in some Alteromonadales and Oceanospirillales genomes. Two highly diverged *hexR* paralogs were detected in Pseudomonadales. The experimentally studied *P. putida* paralog *PP1021* is similar to *hexR* genes from other  $\gamma$ -proteobacteria, whereas the second paralog (termed *hexR1*) is more related to *hexR* from  $\beta$ -proteobacteria. The *Burkholderia* genomes also possess a second copy of *hexR* gene, which is translationally fused to the glucokinase gene, *glk*. Because of the absence of other *glk* homologs in *Burkholderia*, the chimeric *hexR2-glk* gene is the only candidate to fulfill the essential glucokinase function, and the role of its HexR2 domain is to be elucidated.

**Chromosomal Co-localization of *hexR* Genes**—Functionally related genes (e.g. genes involved in the same pathway or a transcriptional regulator and its target genes) often occur close on the bacterial chromosome (32). Indeed, the previously characterized in *P. putida* *hexR* regulator is divergently transcribed with the *zwf-pgl-eda* operon (9). Similar divergent chromosomal arrangement is retained in the *Pseudomonas* and other lineages of  $\gamma$ -proteobacteria, although the content of *zwf*-containing operons is different in some genomes (Fig. 2). A monocistronic *zwf* gene previously detected in *E. coli* (33) is conserved in other Enterobacteriales with the exception of *Edwardsiella tarda*, which encodes a *zwf-pgl-edd-eda* operon. Similar four-gene operons encoding the complete ED pathway were found in the Alteromonadales lineage (e.g. in all *Shewanella* spp.). We also noted that in many  $\gamma$ -proteobacteria, *hexR* is co-localized on the chromosome with glucokinase, *glk*, and pyruvate kinase, *pykA*. Among  $\beta$ -proteobacteria, *hexR* is divergently transcribed with *edd-eda* in *Burkholderia* or it belongs to either *zwf-hexR* operon in Comamonadaceae or the *zwf-pgl-glk-hexR-pgi* operon in Neisseriales. This analysis further confirmed functional coupling between *hexR* and the central carbohydrate metabolism genes and allowed us to define a training set of upstream gene sequences for identification of HexR binding motifs.

**Identification of HexR-binding Sites and Regulons**—To reconstruct the HexR regulons in Proteobacteria, we applied the integrative comparative genomics approach (as imple-

mented in RegPredict webserver (20)) that combines identification of TFs and candidate transcription factor-binding sites with cross-genomic comparison of regulons and with the genomic and functional context analysis of candidate target genes. First, we collected training sets of prospective target genes for each out of 11 defined taxonomic groups possessing HexR orthologs. These training sets were initially defined by the analysis of chromosomal co-localization of *hexR* genes and a previous knowledge of the HexR regulon in *P. putida* and then were further expanded by identifying orthologous target genes in the analyzed genomes. For each taxonomic group, we extracted DNA upstream regions of putative operons in this training set and applied a transcription factor-binding site motif recognition program to derive conserved HexR-binding motifs. After construction of a positional weight matrix for each identified motif, we searched for additional HexR-binding sites in the analyzed genomes, and we finally performed a consistency check or cross-species comparison of the predicted regulons (reviewed in Ref. 5). For those groups of genomes where the HexR regulon was expanded by multiple novel candidate target genes, the above procedure was repeated to include these targets into the transcription factor-binding site motif model and to revise the final gene content of the regulon.

A highly conserved palindromic signal with consensus TGTARnnnnnYTACA (where R and Y denote purines and pyrimidines, respectively) was identified as a candidate HexR-binding motif in nine groups of analyzed genomes (supplemental Fig. S1). Two different motifs were detected for two groups of HexR paralogs in the *Pseudomonas* lineage. Three previously characterized HexR-binding sites in *P. putida* coincide with the binding motif for the HexR group inferred in this study (consensus sequence aTGTTGT-4–8 nucleotides-ACaAcAt). The candidate binding motif for the second group of regulators in *Pseudomonas* (HexR1) is similar to HexR-binding motifs in other Proteobacteria. Significant difference between the predicted binding motifs of HexR paralogs in *Pseudomonas* suggests that the respective regulons should not cross-talk with each other. A different conserved motif, which does not have a palindromic symmetry but which has some resemblance to a right part of the classical HexR-binding motif, was predicted for HexR regulons in *Hahella* and *Marinobacter* spp.

The content of the reconstructed HexR regulons in 11 taxonomic groups of  $\gamma$ - and  $\beta$ -proteobacteria (totally 87 species) is summarized in supplemental Table S3. Detailed information about the predicted DNA-binding sites and downstream regulated genes is provided in the RegPrecise database (regprecise.lbl.gov) (23). The reconstructed HexR regulons control the central carbohydrate metabolism in all analyzed proteobacteria (Fig. 1A). However, as we describe it below, the specific content of HexR regulons is highly variable between different lineages.

**Conserved Members of HexR Regulons**—Based on overall appearance and taxonomic distribution of HexR-regulated genes, we classified them into several groups (Table 1). The first group of genes includes the most conserved regulon members that are regulated by HexR in at least 30 species from at least six different lineages. This group includes *zwf*, *pgl*, *edd*, *eda*, *gapA*, *pykA*, *glk*, and *pgi* genes that encode key glycolytic enzymes from the ED and EMP pathways. In addition, the *hexR* gene was

**TABLE 1**

**Conserved core of HexR regulons in 11 groups of proteobacteria**

This table shows the number of genomes in each taxonomic group where the target gene is preceded by a conserved HexR-binding site. Genes without a HexR binding site are indicated by -. The absence of orthologous gene in analyzed genomes is indicated as NA.

| Operon       | Taxonomic group(s) <sup>1</sup> |                     |   |                        |                  |   |                     |                       |                    |                    |                  | Functional role                         |
|--------------|---------------------------------|---------------------|---|------------------------|------------------|---|---------------------|-----------------------|--------------------|--------------------|------------------|---|
|              | Shewanellaceae (16)             | Alteromonadales (5) | Oceanospirillales/<br>Alteromonadales (6) | Enterobacteriales (11) | Vibrionales (10) | Psychromonadaceae/<br>Aeromonadales (6) | Pseudomonadales (8) | Burkholderia spp. (8) | Ralstonia spp. (6) | Comamonadaceae (9) | Neisseriales (2) |   |
| <i>hexR</i>  | 16                              | 5                   | 5   | 11                     | 4                | 5                                       | 8                   | 8                     | 4                  | 5                  | 2                | HexR regulator, RpiR family             |
| <i>zwf</i>   | 16                              | 5                   | 5   | 11                     | -                | -                                       | 8                   | 8                     | 2                  | 8                  | 2                | phosphoglucose dehydrogenase            |
| <i>pgl</i>   | 16                              | 5                   | 4   | 1                      | n/a              | -                                       | 8                   | 8                     | 2                  | n/a                | 2                | 6-phosphogluconolactonase               |
| <i>edd</i>   | 16                              | 5                   | 5   | 1                      | -                | -                                       | 7                   | 8                     | 6                  | 4                  | 2                | phosphogluconate dehydratase            |
| <i>eda</i>   | 16                              | 5                   | 5   | 3                      | 1                | -                                       | 8                   | 8                     | -                  | 4                  | 2                | 2-dehydro-3-deoxygluconate-6P aldolase  |
| <i>glk</i>   | n/a                             | 4                   | 5   | -                      | n/a              | 4                                       | 7                   | 8                     | 2                  | -                  | 2                | glucokinase                             |
| <i>pgi</i>   | 16                              | 1                   | 1   | -                      | 7                | -                                       | -                   | -                     | -                  | 5                  | 2                | glucose-6-phosphate isomerase           |
| <i>pykA</i>  | 16                              | 4                   | 6   | -                      | 1                | 5                                       | 1                   | -                     | -                  | 7                  | -                | pyruvate kinase                         |
| <i>gapA</i>  | 16                              | 5                   | 5   | -                      | 8                | 4                                       | 6                   | -                     | -                  | -                  | -                | glyceraldehyde-3P dehydrogenase [NAD]   |
| <i>tal</i>   | 16                              | 1                   | -   | -                      | -                | 2                                       | -                   | -                     | -                  | 3                  | -                | transaldolase                           |
| <i>ppsA</i>  | 16                              | 3                   | 2   | -                      | -                | -                                       | -                   | -                     | -                  | -                  | -                | phosphoenolpyruvate synthase            |
| <i>gapB</i>  | 16                              | 1                   | 2   | n/a                    | -                | -                                       | -                   | n/a                   | n/a                | n/a                | n/a              | glyceraldehyde-3P dehydrogenase [NADPH] |
| <i>adhE</i>  | 15                              | -                   | 1   | -                      | -                | 2                                       | n/a                 | n/a                   | -                  | n/a                | -                | alcohol dehydrogenase                   |
| <i>aceBA</i> | 15                              | n/a                 | 1   | -                      | 8                | 2                                       | -                   | -                     | -                  | -                  | -                | malate synthase, isocitrate lyase       |
| <i>aceEF</i> | -                               | -                   | 2   | -                      | -                | -                                       | 8                   | -                     | -                  | -                  | -                | pyruvate dehydrogenase                  |
| <i>ppc</i>   | -                               | 1                   | -   | 4                      | 1-               | 5                                       | -                   | -                     | -                  | -                  | -                | phosphoenolpyruvate carboxylase         |
| <i>pckA</i>  | -                               | 1                   | 2   | 4                      | 3                | -                                       | -                   | n/a                   | n/a                | n/a                | n/a              | phosphoenolpyruvate carboxykinase       |
| <i>tpiA</i>  | -                               | -                   | -   | -                      | 8                | 3                                       | -                   | -                     | -                  | -                  | -                | triosephosphate isomerase               |
| <i>gpmM</i>  | -                               | -                   | 2   | -                      | 10               | -                                       | -                   | n/a                   | n/a                | -                  | -                | phosphoglycerate mutase                 |
| <i>pflBA</i> | -                               | n/a                 | 1   | -                      | 6                | 6                                       | n/a                 | n/a                   | n/a                | n/a                | -                | formate fermentation                    |
| <i>grcA</i>  | n/a                             | n/a                 | n/a                                       | -                      | 9                | 3                                       | n/a                 | n/a                   | n/a                | n/a                | n/a              | formate fermentation                    |
| <i>gltBD</i> | -                               | -                   | -   | -                      | 9                | 3                                       | -                   | -                     | -                  | -                  | -                | glutamate synthase                      |
| <i>pntAB</i> | -                               | -                   | -   | -                      | 8                | 2                                       | -                   | -                     | -                  | -                  | -                | NAD(P) transhydrogenase                 |
| <i>aldE</i>  | -                               | 3                   | 1   | -                      | -                | 2                                       | 1                   | -                     | n/a                | -                  | -                | aldose-1-epimerase                      |
| <i>eno</i>   | -                               | -                   | 1   | -                      | -                | 2                                       | 1                   | -                     | -                  | -                  | -                | enolase                                 |
| <i>pgk</i>   | -                               | -                   | 1   | -                      | -                | -                                       | -                   | -                     | -                  | 2                  | -                | phosphoglycerate kinase                 |
| <i>ptsHI</i> | -                               | n/a                 | n/a                                       | -                      | 2                | 5                                       | -                   | -                     | -                  | -                  | -                | sugar utilization (PTS system)          |
| <i>mtlAD</i> | n/a                             | n/a                 | -   | -                      | 7                | 2                                       | n/a                 | n/a                   | n/a                | n/a                | n/a              | mannitol utilization                    |
| <i>focA</i>  | -                               | n/a                 | -   | -                      | 2                | 5                                       | n/a                 | n/a                   | n/a                | n/a                | n/a              | formate fermentation                    |
| <i>ptsG</i>  | -                               | n/a                 | -   | -                      | 7                | -                                       | n/a                 | n/a                   | n/a                | n/a                | 1                | glucose utilization                     |

<sup>1</sup> The number of genomes in the analyzed taxonomic group is indicated in parentheses.

predicted to be autoregulated in most genomes. The second group includes genes that belong to the HexR regulons in at least 10 species from at least two lineages. Genes in the second group are involved in glycolysis (*gpmM* and *tpiA*), gluconeogenesis (*ppsA*, *gapB*, and *pckA*), the pentose phosphate pathway (*tal*), pyruvate metabolism (*aceEF* and *ppc*), fermentation (*adhE*, *pflBA*, and *grcA*), glyoxylate metabolism (*aceBA*), amino acid biosynthesis (*gltBD*), and NADPH re-oxidation (*pntAB*). The third group includes genes that were found within the HexR regulons in two or more lineages but appeared in less than 10 species. In this group, there are genes involved in gly-

colysis (*aldE*, *pgk*, and *eno*), formate fermentation (*focA*), and glucose and mannitol utilization (*ptsG*, *ptsHI-crr*, and *mtlADR*).

*Lineage-specific HexR-regulated Genes*—The rest of the predicted target genes belong to a group of lineage-specific regulon members (Table 2). At most, this group includes genes with candidate HexR-binding sites that are conserved in at least two species within the same lineage but that are not conserved outside of this lineage. This group also includes the *Colwellia psychrerythraea* fructose-biphosphate aldolase gene, *fba*, which is preceded by a strong HexR-binding site. The largest set of lin-

**TABLE 2**  
Lineage-specific predicted members of HexR regulons

| Taxonomic group(s) <sup>a</sup>       | Target gene(s)                      | No. <sup>b</sup> | Functional role  |  |
|---------------------------------------|-------------------------------------|------------------|--|--|
| Shewanellaceae (16)                   | <i>phk</i>                          | 16               | Phosphoketolase  |  |
|                                       | <i>gnd</i>                          | 14               | 6-Phosphogluconate dehydrogenase                           |  |
|                                       | <i>nqrABCDEF</i>                    | 15               | NADH:ubiquinone oxidoreductase                             |  |
|                                       | <i>deoABD</i>                       | 14               | Deoxynucleoside utilization                                |  |
|                                       | <i>nupC</i>                         | 15               | Nucleoside transporter                                     |  |
|                                       | <i>cdd</i>                          | 10               | Cytidine deaminase   |  |
|                                       | <i>gcvTHP</i>                       | 12               | Glycine cleavage   |  |
|                                       | <i>adhB</i>                         | 8                | Alcohol dehydrogenase II                                   |  |
|                                       | <i>mcp</i>                          | 9                | Chemotaxis protein   |  |
|                                       | <i>SO1118</i>                       | 10               | Hypothetical protein                                       |  |
|                                       | Psychromonadaceae/Aeromonadales (6) | <i>mglBAC</i>    | 3  | Galactoside utilization                  |
|                                       |                                     | <i>glpT</i>      | 2  | Glycerol utilization                     |
|                                       |                                     | <i>ldhA</i>      | 3  | Lactate dehydrogenase                    |
|                                       |                                     | <i>ackA-pta</i>  | 4  | Acetate kinase                           |
| Oceanospirillales/Alteromonadales (6) | <i>gltA</i>                         | 2                | Citrate synthase   |  |
|                                       | <i>mgsA</i>                         | 2                | Methylglyoxal synthase                                     |  |
|                                       | <i>PF00248</i>                      | 2                | Putative aldo/keto reductase                               |  |
|                                       | <i>fbA</i>                          | 1                | Fructose-bisphosphate aldolase                             |  |
| Alteromonadales (5)                   | <i>glgP</i>                         | 2                | Glycogen phosphorylase                                     |  |
|                                       | <i>cpsA</i>                         | 2                | Capsular polysaccharide synthesis enzyme                   |  |
| Enterobacteriales (11)                | <i>ybfA</i>                         | 10               | Putative exported protein                                  |  |
|                                       | <i>pepD</i>                         | 9                | Aminoacyl-histidine dipeptidase                            |  |
| Vibrionales (10)                      | <i>nirBD</i>                        | 7                | Nitrite reductase (NAD(P)H)                                |  |
|                                       | <i>glgX</i>                         | 6                | Glycogen debranching enzyme                                |  |
|                                       | <i>glgCA</i>                        | 5                | Glycogen synthase, glucose-1-phosphate adenylyltransferase |  |
|                                       | <i>lctP</i>                         | 2                | L-Lactate permease   |  |
|                                       | <i>ygaW</i>                         | 8                | Putative inner membrane protein                            |  |
|                                       | Pseudomonadales (8)                 | <i>gapN</i>      | 2  | Putative glyceraldehyde-3P dehydrogenase |
|                                       |                                     | <i>gltRS</i>     | 7  | Glucose uptake regulatory system         |

<sup>a</sup> The number of genomes in the analyzed taxonomic group is indicated in parentheses.

<sup>b</sup> Number of genomes in taxonomic group where target gene has a conserved HexR-binding site.

age-specific HexR targets was detected in the *Shewanella* species, and it includes genes from the central carbohydrate metabolism (*gnd*, *phk*, and *adhB*), nucleoside/deoxynucleoside utilization (*deoABD*, *nupC*, and *cdd*), respiratory chain (*nqr* operon), and glycine utilization (*gcvTHP*). The second largest pool of predicted lineage-specific HexR-regulated genes was detected in *Vibrio* species; it includes glycogen metabolism genes, a nitrite reductase, a peptidase, and a lactate permease. Extended HexR regulons in other lineages contain genes involved in utilization of galactosides and glycerol (*mgl* and *glpT*), glycolysis (*glpN*), the tricarboxylic acid cycle (*gltA*), lactate, and acetate fermentation (*ldhA* and *ackA-pta*).

### Experimental Validation of HexR Regulon in *S. oneidensis* MR-1

**Dual Mode of HexR Regulation**—The reconstructed HexR regulon in *S. oneidensis* MR-1 contains 30 genes (organized in 15 operons) that are involved in various metabolic pathways (Fig. 1B). These include two opposite central carbon metabolism pathways, the glycolytic (ED) pathway for hexose catabolism (*zwf-pgl-edd-eda*, *gapA*, and *pykA*) and the gluconeogenic pathway for glucose biosynthesis (*ppsA* and *gapB*). This observation suggests that HexR may have an opposite regulatory effect on the target gene expression, similar to the Cra dual repressor/activator control of carbon metabolism in *E. coli* (4). To reveal a negative or positive mode of HexR regulation in *Shewanella* we used several complementary approaches as follows: (i) the comparative analysis of upstream promoter sequences in multiple *Shewanella* genomes; (ii) the analysis of correlations of gene expression profiles in the compendium of microarray data available for *S. oneidensis* MR-1; and (iii) RT-PCR analysis of regulon gene expression in *hexR* knock-out mutant (see below).

For each predicted HexR-regulated gene in *S. oneidensis* MR-1, DNA upstream regions of orthologous genes from other *Shewanella* genomes were aligned, and candidate HexR-binding sites and potential promoter elements (the -35 and -10 promoter boxes) were located (supplemental Fig. S2). For 9 genes, namely *zwf*, *pykA*, *gapA*, *gnd*, *nqrA*, *deoA*, *cdd*, *phk*, and *mcp*, HexR operators either overlap with or are located downstream of the putative conserved promoter elements, suggesting that HexR is a repressor of these genes. In contrast, five other genes (*ppsA*, *gapB*, *gcvT*, *tal*, and *nupC*) have HexR operators located upstream of predicted promoters (distance between the center of HexR site and the 5' position of the -35 promoter box equal to either 29, 21, or 14 nucleotides), suggesting that HexR may be acting as an activator.

To evaluate pairwise correlations in expression of all HexR-regulated genes, we have computed the Pearson correlation between each pair of genes in the *S. oneidensis* MR-1 genome based on the expression of each gene in ~200 microarray experiments available in the MicrobesOnline database (supplemental Fig. S3). This analysis allowed us to identify subregulons that have different expression patterns. There is a strong correlation within the first group of *zwf*, *gapA*, *pykA*, and *tal* genes and the second group of *phk*, *deoA*, *cdd*, and *nqrA* genes, although correlation between these two groups is also significant. The third group of highly correlated genes *gapB*, *ppsA*, and *gcvT* shows no correlation with the genes from the first two groups. Indeed, the expression data previously obtained for *S. oneidensis* MR-1 grown on various carbon sources (34) confirm up-regulation of the *deoA*, *zwf*, *phk*, *nupC*, *gapA* genes during growth on inosine, whereas *ppsA*, *gapB*, and *gcvT* genes were down-regulated in these conditions compared with the growth on lactate. The observed different patterns of gene expression

TABLE 3

*In vitro* and *in vivo* assessment of *S. oneidensis* HexR regulon

EMSA and FPA testing of the purified HexR protein was for its ability to bind its target sites in the *S. oneidensis* genome. For FPA, the apparent  $K_d$  values for HexR protein interacting with the tested DNA fragments are indicated. Differential expressions of HexR target genes in *S. oneidensis* MR-1 WT and  $\Delta hexR$  mutant strains grown on various carbon sources are shown.

| Gene   | Operon                 | HexR-binding site  | EMSA shift <sup>a</sup> | FPA, $K_d$ | $\Delta hexR$ /WT glycerate | $\Delta hexR$ /WT inosine |
|--------|------------------------|--------------------|-------------------------|------------|-----------------------------|---------------------------|
| SO0779 | <i>gvcTHP</i>          | ATGTAAAAAATTACAT   | No shift                | 52         | 0.32                        | 1.87                      |
| SO0902 | <i>nqrABCDEF</i>       | TTGTTATAAACTACAT   | Shift                   | NA         | 4.12                        | 6.20                      |
| SO1214 | <i>nupC</i>            | TTGAAATTAACTCCAA   | Shift                   | NA         | 0.48                        | 0.45                      |
| SO1218 | <i>deoABD</i>          | TTGTAATATGCTACAG   | Shift                   | NA         | 5.08                        | 1.85                      |
| SO1483 | <i>aceBA</i>           | TGGTAATTTAATTACGT  | NA                      | 62         | NA                          | NA                        |
| SO1902 | <i>gnd</i>             | TGGTAATTAAGTTACAA  | Weak shift              | 49         | 0.90                        | 1.91                      |
| SO2345 | <i>gapA</i>            | TGGTAATTTTACTACCA  | Shift                   | 71         | 1.87                        | 1.64                      |
| SO2347 | <i>gapB</i>            | CTCTAATAAAAATTACAG | Shift                   | NA         | 0.14                        | 0.36                      |
| SO2489 | <i>zwf-pgl-edd-eda</i> | CTGTGTTATATTTACAA  | Shift                   | 39         | 7.33                        | 4.93                      |
| SO2491 | <i>pykA</i>            | TTGTAACTTTACTACAT  | Shift                   | NA         | 9.81                        | 2.33                      |
| SO2644 | <i>ppsA</i>            | TAGTAATAAAAATTACAA | Shift                   | 37         | 0.06                        | 0.11                      |
| SO2791 | <i>cdd</i>             | CTGTAACAAAATTACTA  | Shift                   | NA         | 3.18                        | 1.38                      |
| SO3542 | <i>phk</i>             | TTGTAAGAAAATTACAA  | Shift                   | 38         | 12.79                       | 4.71                      |
| SO3546 | <i>tal-pgi</i>         | TTGTAATTTTTTTTACAT | Shift                   | 28         | 1.31                        | 1.00                      |
| SO4557 | <i>mcp</i>             | CTGTAAATTTTTTACAG  | Shift                   | NA         | 2.07                        | 0.82                      |

<sup>a</sup> Genes that were not tested in any experiment are indicated as NA.

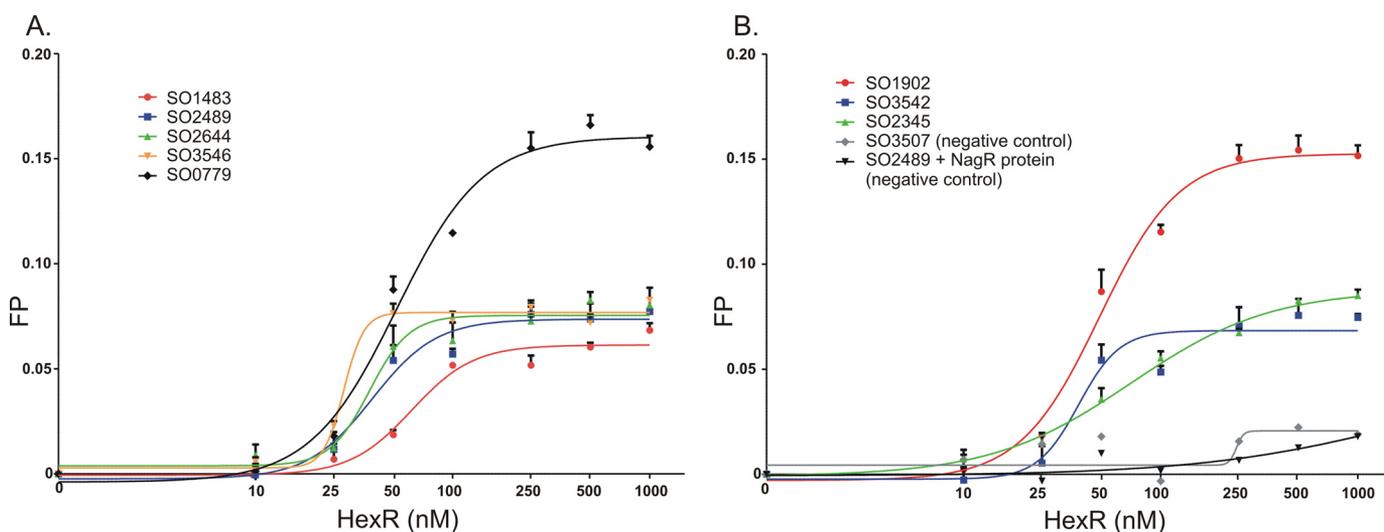


FIGURE 3. Fluorescence polarization binding assay of HexR with target DNA operators in *S. oneidensis* MR-1. Increasing HexR protein concentrations (0.01–1  $\mu$ M) were mixed with 27-bp fluorescence-labeled DNA fragments (1 nM) of gene regions containing HexR-binding sites. A, SO1483, SO2489, SO2644, SO3546, and SO0779; B, SO1902, SO3542, and SO2345. DNA-protein complex formation was monitored by an increase in the fluorescence polarization value. As a negative control, the *S. oneidensis* NagR regulator and its target DNA fragment (SO3507) were used.

in the HexR regulon and the sequence analysis of promoter gene regions in *S. oneidensis* MR-1 suggest that HexR is a dual mode regulator that can be either repressor or activator of genes expression depending on the relative position of its operators and promoters.

**HexR Binds Its Cognate DNA Sites *in Vitro***—To test the ability of HexR to specifically bind to the predicted DNA sites and to assess effectors, *hexR* (SO2490) from *S. oneidensis*, MR-1 was cloned and overexpressed in *E. coli*. The Smt3-His<sub>6</sub>-tagged recombinant protein HexR was purified by Ni<sup>2+</sup>-chelating chromatography, followed by gel filtration as described under “Experimental Procedures.” Electrophoretic mobility shift assay (EMSA) and FPA were used to test specific binding of the purified HexR protein to its predicted operator sites in upstream regions of 15 genes in *S. oneidensis* MR-1 (Table 3).

Binding of purified HexR to the SO2489 dsDNA fragment containing the predicted HexR-binding site resulted in the reduced electrophoretic mobility of the DNA fragment in a concentration-dependent manner (supplemental Fig. S4A). Of

the tested intermediary metabolites associated with the ED pathway (2 mM KDPG, 6-phosphoglucose, phosphoenolpyruvate, or 6-phosphogluconate), only KDPG demonstrated significant suppression of the HexR-DNA binding-mediated shift. These findings are in agreement with the previous results obtained for the *P. putida* HexR regulator, which binds DNA in the absence of an effector, whereas KDPG induces the HexR-DNA complex dissociation (8). Specific binding at 0.5  $\mu$ M HexR was also confirmed for all other tested DNA fragments (supplemental Table S1) with the single exception of the SO0779 dsDNA fragment (supplemental Fig. S4B). The latter fragment did not show a clear shift in the EMSA experiments but was confirmed to bind HexR using the FPA assay (see below).

The FPA binding assay was used to assess the specific binding of eight predicted HexR-binding sites in *S. oneidensis* with increasing concentrations of the HexR protein (Fig. 3). All tested DNA fragments demonstrated the HexR concentration-dependent increase of fluorescence polarization, confirming specific interactions between HexR and DNA. The apparent  $K_d$

values for HexR protein interacting with the tested DNA fragments were in the range of 28–71 nM (Table 3). As a negative control, we assessed interaction between a different *S. oneidensis* transcription factor (NagR or SO3516) with the SO2489 DNA fragment, and the HexR regulator with the SO3507 DNA fragment containing the previously confirmed NagR-binding site (12), and in both experiments no significant change of fluorescence polarization was detected. These results confirm that HexR is a KDPG-responsive regulator that binds specifically to the predicted operator sites in *S. oneidensis* MR-1.

**HexR Affects Expression of Its Target Genes in Vivo**—To validate the predicted mode of HexR regulation on gene expression *in vivo*, a  $\Delta hexR$  targeted deletion mutant strain of *S. oneidensis* MR-1 was constructed, and the relative transcript levels of the predicted HexR target genes analyzed by quantitative RT-PCR (Table 3). Relative mRNA levels of nine genes were elevated more than 1.5-fold in the  $\Delta hexR$  mutant compared with the WT strain when grown in the minimal medium supplied with either glycerate or inosine. The most prominent effect of *hexR* mutation was observed for the *zwf*, *pykA*, *nqrA*, *phk*, *deoA*, and *gapA* genes, suggesting that HexR represses expression of these genes. The *cdd* and *mcp* genes were up-regulated 2–3-fold in  $\Delta hexR$  strain in the media supplied with glycerate, whereas *gnd* was up-regulated near 2-fold in the cells grown on inosine. Expression of the *tal* gene was not significantly affected by *hexR* mutation in both conditions. In contrast, the *ppsA*, *gapB*, and *nupC* genes were significantly down-regulated in the  $\Delta hexR$  strain. For instance, expression of the *ppsA* gene was decreased 9- and 17-fold in the cells grown on inosine and glycerate, respectively. Finally, expression of *gcvT* was down-regulated 3-fold in the  $\Delta hexR$  strain when the cells were grown on glycerate but was up-regulated 2-fold in the inosine-supplied cells. In our previous study of metabolic regulons in *Shewanella* spp. (12), we have identified several regulatory motifs in the upstream region of the *gcvTHP* operon (supplemental Fig. S2), which is presumably regulated by several factors, including the HexR activator, the glycine-responsive activator GcvA, and the previously unknown regulator of purine metabolism operated by binding to PUR-boxes. These overlapping regulatory interactions could possibly explain the observed differences in the expression patterns of the *gcvH* genes in  $\Delta hexR$  strain grown on either glycerate or inosine, as the purine nucleoside could potentially repress the *gcv* operon via a PUR-box operator. The observed differences in the regulation of *gnd* and *mcp* promoters in strains grown on different carbon sources could be explained by possible involvement in their regulation of other still unknown regulatory mechanisms that can differentiate the tested carbon sources. These observations confirm the predicted negative mode of HexR regulation for the *zwf*, *pykA*, *nqrA*, *phk*, *deoA*, *gapA*, *gnd*, *cdd*, and *mcp* genes and its positive mode of action on the *ppsA*, *gapB*, *nupC*, and *gcvT* genes.

**Phenotype Characterization of *S. oneidensis*  $\Delta hexR$  Mutant**—The previous phenotypic and genomic analyses suggest that *S. oneidensis* can use GlcNAc, glycerate, inosine, lactate, and pyruvate as the sole carbon source (13, 34, 35). Because of the absence of phosphofructokinase in *Shewanella* spp., utilization of GlcNAc proceeds through glucose 6-phosphate, which is

catabolized via the HexR-regulated ED pathway (Fig. 1B). Utilization of inosine requires the HexR-regulated *nupC* and *deoABD* genes that mediate nucleoside uptake, release of ribose base, and its further utilization via the pentose phosphate pathway. The catabolic pathways produce pyruvate, which is further utilized via the fermentation/respiration routes to gain energy and the gluconeogenesis for biosynthetic needs. To determine whether the loss of HexR function impairs the utilization of carbon sources, the *S. oneidensis* MR-1 WT and  $\Delta hexR$  mutant strains were tested for the ability to grow on the above five carbon sources. The *hexR* mutant was unable to grow on lactate or pyruvate but grew similar to the WT on GlcNAc (Fig. 4) or inosine (data not shown). An enhanced growth on glycerate was observed for the  $\Delta hexR$  mutant compared with the WT strain (Fig. 4). The inability of the *S. oneidensis*  $\Delta hexR$  strain to grow on lactate as a sole carbon source confirms that HexR is an important transcriptional activator of the gluconeogenic gene *ppsA*, whose activity is known to be essential for the growth of *E. coli* on lactate (36). Therefore, based on a combination of regulon analysis, metabolic reconstruction (31), and physiological data, we hypothesized that HexR controls the main gluconeogenic flux from lactate via activation of PpsA. Disruption of this flux in  $\Delta hexR$  mutant might be the main cause of its inability to grow on lactate as a single carbon source. However, the flux from lactate through the tricarboxylic acid cycle appears to be HexR-independent, and thus, it might continue in its role in energy production even in the  $\Delta hexR$  mutant (in the presence of an additional carbon source, see below).

**Metabolic Flux Analysis of *S. oneidensis*  $\Delta hexR$  Mutant**—To test this hypothesis, the intracellular carbon fluxes in the WT and the  $\Delta hexR$  mutant strains were quantified by using the  $^{13}\text{C}$ -based metabolic flux analysis that relies on the [U- $^{13}\text{C}$ ]lactate labeling and GC-MS analysis of the mass isotopomer pattern in proteinogenic amino acids (supplemental Table S4). From the labeling patterns of the respective amino acids, the origins of phosphoenolpyruvate were quantitatively determined (Fig. 5). Phosphoenolpyruvate can be derived from pyruvate via phosphoenolpyruvate synthase encoded by *ppsA*, from oxaloacetate via phosphoenolpyruvate carboxykinase, or from glycerate through the glycolytic conversion of 3-phosphoglycerate to phosphoenolpyruvate. In the *S. oneidensis* MR-1 WT strain grown on the mixture of lactate and glycerate, more than 40% of phosphoenolpyruvate originated from the phosphoenolpyruvate synthase reaction (Fig. 5). In contrast, the phosphoenolpyruvate synthase flux was absent, and the majority of phosphoenolpyruvate was derived from glycerate in the  $\Delta hexR$  mutant strain. A similar result was also found for the  $\Delta ppsA$  mutant strain (supplemental Table S4). These results indicate that the *ppsA*-encoding phosphoenolpyruvate synthase is largely inactive *in vivo* in the  $\Delta hexR$  mutant. Moreover, the HexR deficiency also resulted in significant changes in central metabolic fluxes. Compared with the WT, a significant increase in the uptake fluxes of lactate and glycerate was observed for the  $\Delta hexR$  mutant. This mutant also exhibited much higher fluxes of secretion of acetate and pyruvate.

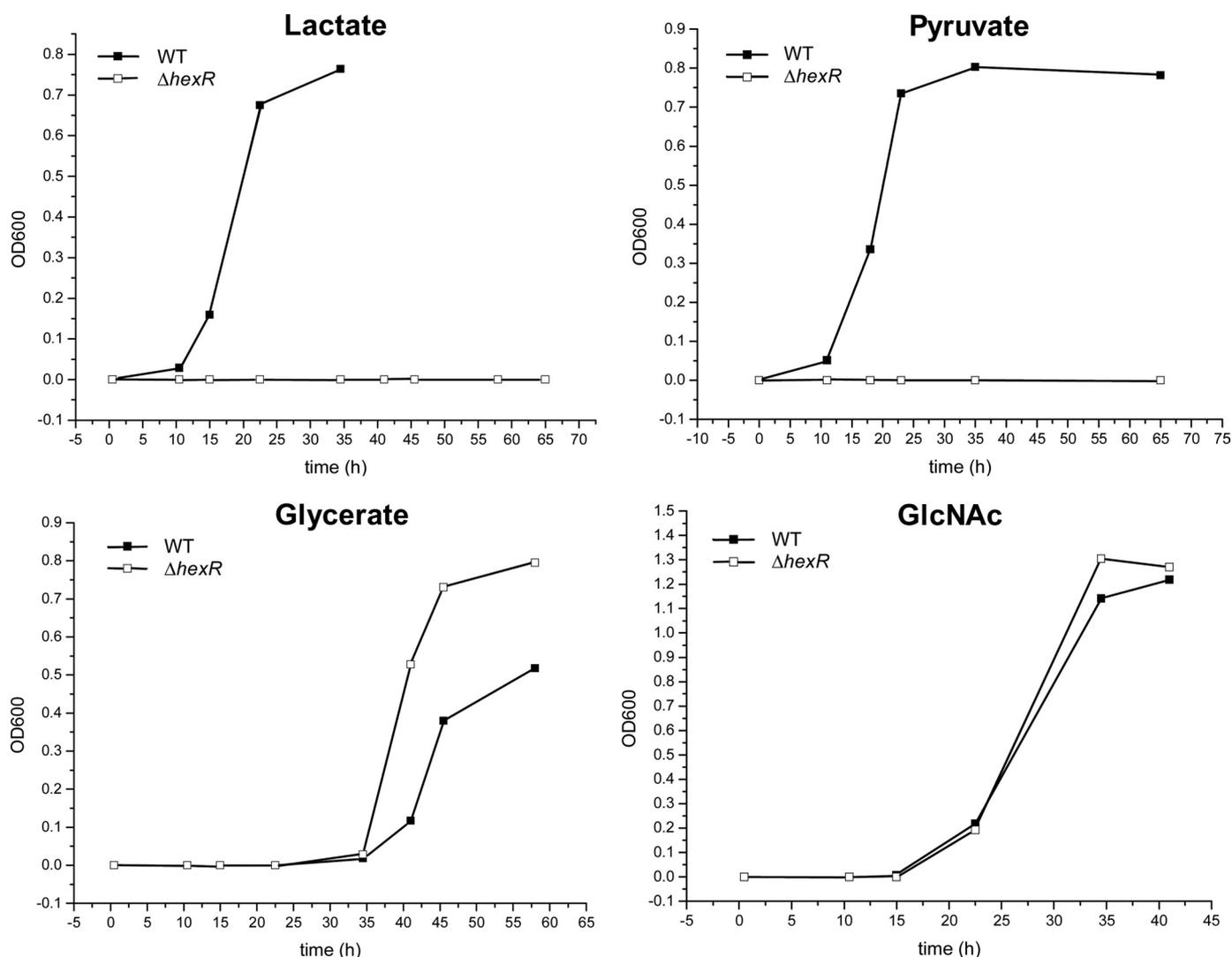


FIGURE 4. Growth phenotypes of *S. oneidensis* MR-1  $\Delta hexR$  mutant on various carbon sources. Cells were grown in the minimal medium containing 10 mM *N*-acetyl-D-glucosamine (GlcNAc) or 20 mM glycerate, lactate, and pyruvate as the sole carbon source. The cell growth was monitored by measuring the absorbance at 600 nm.

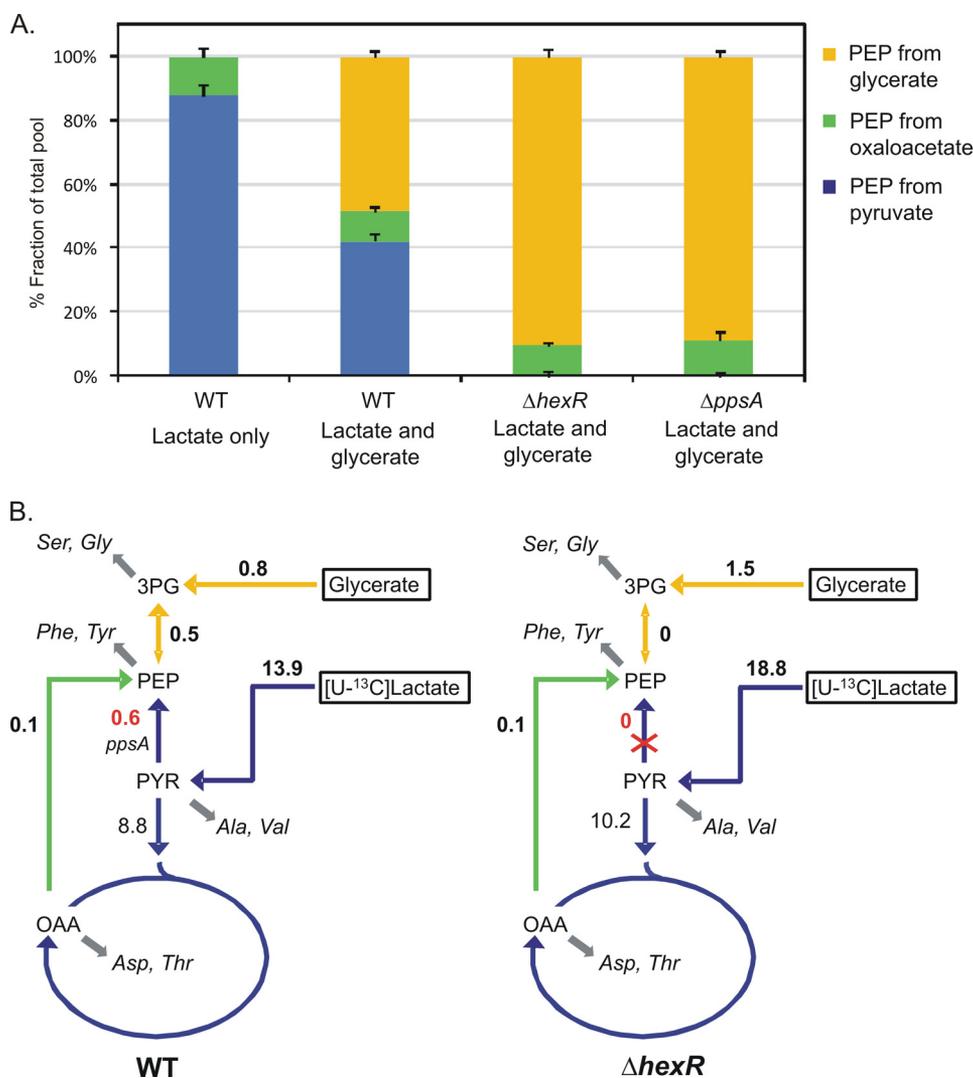
## DISCUSSION

Transcriptional control of metabolic pathways for carbohydrate utilization in bacteria is mediated by a variety of TFs from different protein families (e.g. AraC, LacI, GntR, DeoR, and ROK). Peripheral sugar catabolic pathways are usually controlled by committed transcription factors with presumably local regulons (e.g. the GlcNAc utilization pathway is regulated by NagR repressor in *Shewanella*) (13), whereas central glycolytic pathways are regulated by global TFs (e.g. the EMP pathway is under control of Crp and FruR in *E. coli*) (1). Previous studies show that *P. putida* HexR is a KDPG-responsive transcriptional regulator that negatively controls the *edd-glk-gltRS*, *zwf-pgl-eda*, and *gap* genes involved in the ED and EMP pathways of glucose utilization (8–11). Here, we performed the comparative genomics reconstruction of HexR regulons in 87 species of  $\gamma$ - and  $\beta$ -proteobacteria by combining the identification candidate HexR-binding sites with the analysis of functional genomic context. We report a high variability in gene content of the reconstructed HexR regulons in different taxa, although the cognate DNA motifs of HexR orthologs share a

common palindromic consensus sequence TGTA-N<sub>7</sub>-TACA. The combined bioinformatics, *in vitro* and *in vivo* characterization of the extended HexR regulon in *Shewanella* spp., revealed that HexR functions as both a repressor and activator on different promoters involved in the catabolic and gluconeogenic pathways, respectively. The discovered dual mode of action and the functional context of target genes indicate that HexR is a novel pleiotropic regulator of the central carbon metabolism in Proteobacteria. In *Shewanella*, HexR usually controls various genes from the central carbon metabolism, as well as the deoxynucleoside and glycine utilization.

HexR belongs to the RpiR family of TFs that are characterized by an N-terminal DNA binding helix-turn-helix domain (HTH\_6 or PF01418 in Pfam database) and a C-terminal sugar phosphate binding domain (SIS or PF01380) (37). The presence of the SIS effector binding domain in regulators from the RpiR family correlates with their known ability to bind various sugar phosphates, including KDPG for HexR in *P. putida* (8, 10), allse 6-phosphate for RpiR/AlsR in *E. coli* (38), *N*-acetylmuramic acid 6-phosphate for MurR in *E. coli* (39), and glucos-

## Proteobacterial HexR Catabolite Repressor/Activator



**FIGURE 5. Origins of phosphoenolpyruvate (A) and *in vivo* phosphoenolpyruvate-synthesizing fluxes (B) in *S. oneidensis* MR-1 WT and  $\Delta$ hexR mutant.** Results were obtained from the labeling experiments with 4 mM [U-<sup>13</sup>C]lactate and 16 mM unlabeled lactate. Unlabeled glycerate (20 mM) was added to the medium as indicated. The numbers in B are the molar net fluxes (mmol g<sup>-1</sup> h<sup>-1</sup>) determined. Double-headed arrows indicate reactions assumed to be reversible, and the arrows with bigger heads represent the direction of the determined net fluxes. Gray arrows indicate precursor withdrawal for the amino acid analyzed by GC-MS. Metabolite abbreviations: 3PG, 3-phosphoglycerate; PYR, pyruvate; OAA, oxaloacetate. Net fluxes are in supplemental Table S4.

amine 6-phosphate for SiaR in *Haemophilus influenzae* (40). Multiple alignment of HexR proteins revealed that among the residues that were previously predicted to be more important for the recognition of DNA in *P. putida* HexR (according to a homology-based three-dimensional model of HexR (8)), only Arg-54 is absolutely conserved in orthologs; Glu-49 and Arg-57 are present in more than 90% of orthologs, whereas Gln-43 and Lys-46 are present only in HexR orthologs in *Pseudomonas* spp. These observations correlate with the relative conservation of the identified HexR-binding DNA motifs. Interestingly, the same residues are retained in a HexR paralog translationally fused to a glucokinase domain in *Burkholderia* spp., suggesting that this chimeric protein could function both as a kinase and a transcriptional regulator. Two residues likely involved in effector recognition in *P. putida* HexR, Ser-140 and Ser-184 (8), were found absolutely conserved in all HexR orthologs, suggesting that its effector molecule KDPG validated in *P. putida* (8) and *S. oneidensis* (this study) is conserved in all HexR orthologs.

Because of the absence of 6-phosphofruktokinase, the ED pathway is an essential sugar catabolic route, and KDPG is its key intermediate in *Shewanella*, as well as in *Pseudomonas*, *Burkholderia*, *Ralstonia* and some other Proteobacteria. In this work, we demonstrated that the HexR protein in *Shewanella* is a KDPG-responsive repressor/activator of the central carbon metabolism (Fig. 1). The proposed model of HexR-dependent control in *Shewanella* is presented below. In the presence of carbohydrates (hexoses or pentoses), KDPG levels in the cell increase due to the consecutive action of peripheral and central catabolic pathways. Binding of KDPG to the HexR SIS domain promotes dissociation of HexR-DNA complexes, leading to de-repression of transcription of the central glycolytic genes (*zwf-pgl-edd-eda*, *gapA*, *pykA*, *gnd*, and *phk*) and de-activation of the *ppsA* and *gapB* genes involved in the gluconeogenesis. In contrast, when carbohydrates are limited in the medium (e.g. when the cell is growing on lactate or pyruvate as the sole carbon source), the KDPG level is decreased; and HexR binding to its operator sites promotes

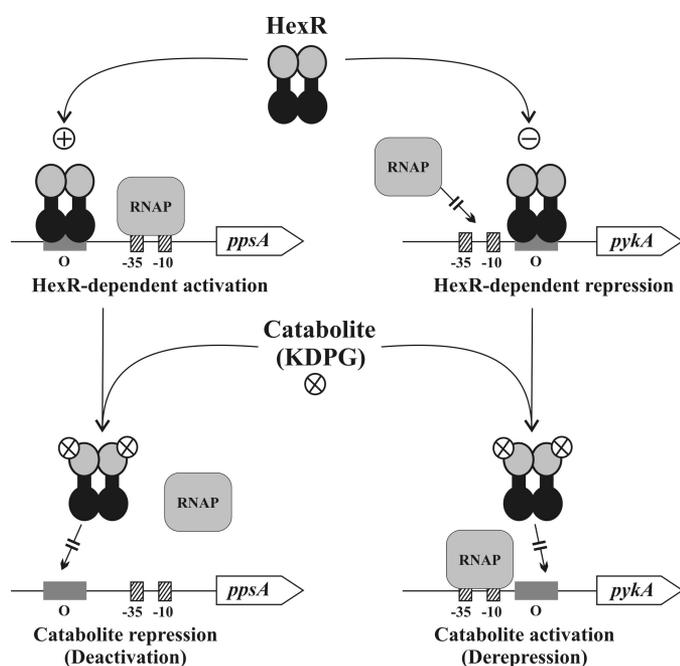


FIGURE 6. **Model of HexR-dependent regulation in *Shewanella*.** In the unliganded form, HexR binds to the operator (O) regions in the target operons. When HexR operator is upstream of the RNA polymerase (RNAP)-binding site, activation of transcription is observed. When HexR operator overlaps or is downstream of the RNA polymerase-binding site, inhibition of transcription is observed. In the presence of exogenous sugars, their utilization results in accumulation of glycolytic intermediates. One of these catabolites, KDPG, bind to HexR, causing it to dissociate from the DNA. Dissociation reverses the activating effect of HexR, as in the case of *ppsA* gene of *S. oneidensis*, which encodes phosphoenolpyruvate synthase (catabolite repression, left), and reverses the inhibiting effect of HexR, as in the case of the *pykA* gene, which encodes pyruvate kinase (catabolite activation, right).

repression of the glycolytic genes and activation of the gluconeogenic genes (Fig. 6).

Lactate is a major source of carbon and energy for *Shewanellae*, a diverse group of dissimilatory metal-reducing bacteria commonly found in aquatic and sedimentary environments (35). Based on the combination of regulatory and phenotype analysis of *S. oneidensis* wild-type and  $\Delta hexR$  mutant (Fig. 4), we concluded that HexR is a master regulator of the gluconeogenic flux from pyruvate (and lactate). Thus, when lactate is present as a single carbon source, it is used for both energy production (via the tricarboxylic acid cycle) and biosynthesis (via gluconeogenesis). This is achieved via a low level of KDPG leading to transcriptional activation of PpsA and down-regulation of PykA (to avoid futile cycling). When lactate is present simultaneously with other carbon sources such as GlcNAc, most of the lactate would route to the tricarboxylic acid cycle for energy, whereas biosynthetic needs are supported by GlcNAc. This is achieved via accumulation of KDPG, which interacts with HexR, resulting in de-activation of PpsA and de-repression of PykA. According to this model, the pyruvate kinase PykA up-regulation would promote disposal of excess glycolytic substrate via an elevated flux from phosphoenolpyruvate to pyruvate. This situation is modeled by deletion of HexR or by deletion of PpsA (Fig. 5).

The HexR regulon in *Shewanella* can be considered as a partial functional replacement of the Enterobacterial Cra (FruR) regulon that plays a pleiotropic role, modulating the direction

of carbon flow through the different metabolic pathways of energy metabolism. The catabolite repressor/activator Cra in *E. coli* binds DNA operators in the absence of fructose 1-phosphate or fructose 1,6-bisphosphate and is inactivated upon the interaction with the effectors (3). When bound to its targets, it acts as an activator of gluconeogenic (*ppsA* and *pckA*) and glyoxylate shunt (*aceBA*) genes and a repressor of glycolytic genes (*pykF*, *gapA*, *edd-eda*, *eno*, *pgk*, and *pfkA*) (41). Thus, both HexR and Cra are dual transcriptional regulators that mediate cAMP-independent catabolite control in response to key catabolite intermediates. *Shewanella* spp. does not have Cra and are not able to grow on fructose (13). The Cra (FruR) regulons reconstructed in other  $\gamma$ -proteobacteria (Aeromonadales, Pseudomonadales, Psychromonadaceae, and Vibrionales) are implicated in the local control of the fructose utilization operon (5). The HexR regulons are local (regulate 1–2 operons) in some lineages (e.g. in Enterobacteriales, *Ralstonia*, and *Burkholderia*) and global (control 15–20 operons) in other lineages (e.g. in *Shewanella*, Aeromonadales, Psychromonadaceae, and Vibrionales). For instance, in *E. coli* HexR has only two predicted target genes (*zwf* and *ybfA*), whereas in *Photobacterium profundum* it controls near 20 operons implicated in several central metabolic pathways. These observations suggest that regulatory networks for the central carbon metabolism are very flexible in Proteobacteria. Global rewiring of these regulatory networks includes multiple regulon expansion and reduction events.

Interestingly, the KDG-responsive KdgR regulon negatively controlling the pectin catabolism in plant pathogenic *Erwinia* is largely extended in comparison with other enterobacteria and include the positively regulated gene *ppsA* (42). Therefore, the *ppsA* gene is activated by at least three dual-mode catabolic regulators (HexR, FruR, and KdgR) in three lineages of  $\gamma$ -proteobacteria (*Shewanella*, *Escherichia*, and *Erwinia*, respectively).

**Concluding Remarks**—This study provided a comprehensive bioinformatic analysis of the HexR regulon that constitutes a pleiotropic system of transcriptional regulation of central carbohydrate metabolism in several groups of Proteobacteria. The key conjectures delivered by this analysis, the global regulatory effect and dual mode of HexR-dependent regulation, were experimentally confirmed in the *S. oneidensis* model providing additional support for the reconstruction of the entire HexR regulon in Proteobacteria. This study demonstrates the value of comparative genomics supported by focused validation experiments for the *ab initio* reconstruction regulatory networks in large sets of previously unexplored biological species.

## REFERENCES

1. Saier, M. H., Jr. (1998) *Biotechnol. Bioeng.* **58**, 170–174
2. Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W. A., and Saier, M. H., Jr. (2004) *J. Bacteriol.* **186**, 3516–3524
3. Saier, M. H., Jr., and Ramseier, T. M. (1996) *J. Bacteriol.* **178**, 3411–3417
4. Ramseier, T. M., Bledig, S., Michotey, V., Feghali, R., and Saier, M. H., Jr. (1995) *Mol. Microbiol.* **16**, 1157–1169
5. Rodionov, D. A. (2007) *Chem. Rev.* **107**, 3467–3497
6. Collier, D. N., Hager, P. W., and Phibbs, P. V., Jr. (1996) *Res. Microbiol.* **147**, 551–561
7. Lessie, T. G., and Phibbs, P. V., Jr. (1984) *Annu. Rev. Microbiol.* **38**, 359–388

8. Daddaoua, A., Krell, T., and Ramos, J. L. (2009) *J. Biol. Chem.* **284**, 21360–21368
9. del Castillo, T., Duque, E., and Ramos, J. L. (2008) *J. Bacteriol.* **190**, 2331–2339
10. Kim, J., Jeon, C. O., and Park, W. (2008) *Microbiology* **154**, 3905–3916
11. Petruschka, L., Adolf, K., Burchhardt, G., Dervede, J., Jürgensen, J., and Herrmann, H. (2002) *FEMS Microbiol. Lett.* **215**, 89–95
12. Rodionov, D. A., Novichkov, P. S., Stavrovskaya, E. D., Rodionova, I. A., Li, X., Kazakov, M. D., Ravcheev, D. A., Gerasimova, A. V., Kazakov, A. E., Kovaleva, G. Y., Permina, E. A., Laikova, O. N., Overbeek, R., Romine, M. F., Fredrickson, J. K., Arkin, A. P., Dubchak, I., Osterman, A. L., and Gelfand, M. S. (2011) *BMC Genomics* **12**, S3
13. Rodionov, D. A., Yang, C., Li, X., Rodionova, I. A., Wang, Y., Obratsova, A. Y., Zagnitko, O. P., Overbeek, R., Romine, M. F., Reed, S., Fredrickson, J. K., Neelson, K. H., and Osterman, A. L. (2010) *BMC Genomics* **11**, 494
14. Benson, D. A., Karsch-Mizrachi, L., Lipman, D. J., Ostell, J., and Sayers, E. W. (2010) *Nucleic Acids Res.* **38**, D46–D51
15. Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J., and Arkin, A. P. (2010) *Nucleic Acids Res.* **38**, D396–D400
16. Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) *Nucleic Acids Res.* **33**, 5691–5702
17. Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., and Rupp, R. (2007) *BMC Bioinformatics* **8**, 460
18. Edgar, R. C. (2004) *Nucleic Acids Res.* **32**, 1792–1797
19. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
20. Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., Arkin, A. P., Mironov, A. A., and Dubchak, I. (2010) *Nucleic Acids Res.* **38**, W299–W307
21. Mironov, A. A., Koonin, E. V., Roytberg, M. A., and Gelfand, M. S. (1999) *Nucleic Acids Res.* **27**, 2981–2989
22. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) *Genome Res.* **14**, 1188–1190
23. Novichkov, P. S., Laikova, O. N., Novichkova, E. S., Gelfand, M. S., Arkin, A. P., Dubchak, I., and Rodionov, D. A. (2010) *Nucleic Acids Res.* **38**, D111–D118
24. Lin, C. T., Moore, P. A., Auberry, D. L., Landorf, E. V., Peppler, T., Victry, K. D., Collart, F. R., and Kery, V. (2006) *Protein Expr. Purif.* **47**, 16–24
25. Mossessova, E., and Lima, C. D. (2000) *Mol. Cell* **5**, 865–876
26. Osterman, A. L., Lueder, D. V., Quick, M., Myers, D., Canagarajah, B. J., and Phillips, M. A. (1995) *Biochemistry* **34**, 13431–13436
27. Rodionov, D. A., De Ingeniis, J., Mancini, C., Cimadamore, F., Zhang, H., Osterman, A. L., and Raffaelli, N. (2008) *Nucleic Acids Res.* **36**, 2047–2059
28. Titolo, S., Brault, K., Majewski, J., White, P. W., and Archambault, J. (2003) *J. Virol.* **77**, 5178–5191
29. Livak, K. J., and Schmittgen, T. D. (2001) *Methods* **25**, 402–408
30. Zamboni, N., Fendt, S. M., Rühl, M., and Sauer, U. (2009) *Nat. Protoc.* **4**, 878–892
31. Pinchuk, G. E., Hill, E. A., Geydebrekht, O. V., De Ingeniis, J., Zhang, X., Osterman, A., Scott, J. H., Reed, S. B., Romine, M. F., Konopka, A. E., Beliaev, A. S., Fredrickson, J. K., and Reed, J. L. (2010) *PLoS Comput. Biol.* **6**, e1000822
32. Osterman, A., and Overbeek, R. (2003) *Curr. Opin. Chem. Biol.* **7**, 238–251
33. Rowley, D. L., and Wolf, R. E., Jr. (1991) *J. Bacteriol.* **173**, 968–977
34. Driscoll, M. E., Romine, M. F., Juhn, F. S., Serres, M. H., McCue, L. A., Beliaev, A. S., Fredrickson, J. K., and Gardner, T. S. (2007) *Genome Inform.* **18**, 287–298
35. Pinchuk, G. E., Rodionov, D. A., Yang, C., Li, X., Osterman, A. L., Dervyn, E., Geydebrekht, O. V., Reed, S. B., Romine, M. F., Collart, F. R., Scott, J. H., Fredrickson, J. K., and Beliaev, A. S. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2874–2879
36. Niersbach, M., Kreuzaler, F., Geerse, R. H., Postma, P. W., and Hirsch, H. J. (1992) *Mol. Gen. Genet.* **231**, 332–336
37. Bateman, A. (1999) *Trends Biochem. Sci.* **24**, 94–95
38. Sørensen, K. I., and Hove-Jensen, B. (1996) *J. Bacteriol.* **178**, 1003–1011
39. Jaeger, T., and Mayer, C. (2008) *J. Bacteriol.* **190**, 6598–6608
40. Johnston, J. W., Shamsulddin, H., Miller, A. F., and Apicella, M. A. (2010) *BMC Microbiol.* **10**, 240
41. Shimada, T., Yamamoto, K., and Ishihama, A. (2011) *J. Bacteriol.* **193**, 649–659
42. Rodionov, D. A., Gelfand, M. S., and Hugouvieux-Cotte-Pattat, N. (2004) *Microbiology* **150**, 3571–3590