An Expression-Driven Approach to the Prediction of Carbohydrate Transport and Utilization Regulons in the Hyperthermophilic Bacterium *Thermotoga maritima*

Shannon B. Conners, Clemente I. Montero, Donald A. Comfort, Keith R. Shockley,*‡
Matthew R. Johnson, Swapnil R. Chhabra,§ and Robert M. Kelly*

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695-7905

Received 28 April 2005/Accepted 27 July 2005

Comprehensive analysis of genome-wide expression patterns during growth of the hyperthermophilic bacterium *Thermotoga maritima* on 14 monosaccharide and polysaccharide substrates was undertaken with the goal of proposing carbohydrate specificities for transport systems and putative transcriptional regulators. Saccharide-induced regulons were predicted through the complementary use of comparative genomics, mixed-model analysis of genome-wide microarray expression data, and examination of upstream sequence patterns. The results indicate that *T. maritima* relies extensively on ABC transporters for carbohydrate uptake, many of which are likely controlled by local regulators responsive to either the transport substrate or a key metabolic degradation product. Roles in uptake of specific carbohydrates were suggested for members of the expanded Opp/Dpp family of ABC transporters. In this family, phylogenetic relationships among transport systems revealed patterns of possible duplication and divergence as a strategy for the evolution of new uptake capabilities. The presence of GC-rich hairpin sequences between substrate-binding proteins and other components of Opp/Dpp family transporters offers a possible explanation for differential regulation of transporter subunit genes. Numerous improvements to *T. maritima* genome annotations were proposed, including the identification of ABC transport systems originally annotated as oligopeptide transporters as candidate transporters for rhamnose, xylose, β-xylan, and β-glucans and identification of genes likely to encode proteins missing from current annotations of the pentose phosphate pathway. Beyond the information obtained for *T. maritima*, the present study illustrates how expression-based strategies can be used for improving genome annotation in other microorganisms, especially those for which genetic systems are unavailable.

*Thermotoga maritima*, a hyperthermophilic anaerobe with an optimal growth temperature of 80°C, has been found in diverse high-temperature locations and is capable of using a wide variety of simple and complex carbohydrate substrates for growth. The complexity of its carbohydrate utilization strategies, revealed by genome sequencing (48) and through previous work (11, 12, 47, 51), is surprising, given the primitive genetic distance between homologs in this microorganism. Considerable genomic plasticity has been observed even within the *Thermotoga* genus, with respect to the gene content of carbohydrate active enzymes and transporter subunits, which may to some extent relate to lateral gene transfer events (48, 49). Despite the range of sugar-active enzymes found within *T. maritima* MSB8 genome (Table S1 in the supplemental material) (6, 9, 10, 23, 27, 34, 35, 37, 38, 39, 42, 45, 46, 54, 70, 71), a PTS (phosphoenolpyruvate-dependent phosphotransferase system) similar to those used by other species for preferential uptake of selected sugars is apparently absent (48). No homologs of the PTS components EI and HPr (phosphocarrier proteins) and no sugar-specific EIi sugar transporter subunits have been identified in the *Thermogales*. Homologs of PTS-associated transcriptional regulators are found in *T. maritima* MSB8 but have not been implicated in global transcriptional regulation of sugar uptake. Whereas catabolite repression by glucose has been demonstrated for *Thermotoga neapolitana* (78), a mechanism for the global regulation of sugar utilization remains to be identified within the *Thermotoga* genus.

The importance of carbohydrates as carbon and energy sources for *T. maritima* is reflected by the disproportionate number of ABC (for ATP-binding cassettes) transporters that are found within *T. maritima* relative to its genome size (56). These ABC transporters can be classified into large families of sugar transporters and peptide (Opp, oligopeptide; Dpp, dipeptide) transporters, although it has been suggested that both types may participate in the uptake of simple and complex sugars in *T. maritima* (11, 12, 28). Attempts to annotate the functional specificity of these transporters using computational tools have been largely unsuccessful (59) due to the phylogenetic distance between homologs in *T. maritima* and model bacteria. In fact, several sets of *T. maritima* “oligopeptide” transporters are more closely related to archaeal sugar transporters (15, 29) than characterized bacterial peptide transporters and may have arrived in the *T. maritima* lineage through lateral gene transfer (48). Presumably, subsequent duplication...
and divergence events generated paralogous sets of transporter gene subfamilies with different sugar-binding specificities. Determining the apparent specificities of each system and associated transcriptional regulators or hydrolases is a key step in testing this hypothesis. Most members of the LacI (lactose repressor) family of carbohydrate-responsive transcriptional regulators in *T. maritima* cannot be easily assigned into known functional classes using a subset of protein sites (44). Similarly, the specificities of the multiple *T. maritima* homologs of the XylR (xylose repressor) family regulators cannot be determined from sequence homology alone. The presence of these genes nearby sets of ABC transporters suggests that they may play a regulatory role in uptake and utilization of different carbohydrates. Genetic systems enabling knockouts or in vivo overexpression studies of genes are currently lacking for *T. maritima*, as well as for the majority of sequenced bacterial genomes, which now number ~180 complete and >300 in progress (4). Clearly, alternative complementary methodologies are necessary for performing large-scale functional predictions for expanded protein families in organisms such as *T. maritima*, which lack genetic tools.

Transcriptional analysis has proven to be a useful tool for the annotation of members of expanded gene families in a number of genomes. Such approaches have been instrumental in revealing biological pathways (41) and suggesting likely functions for individual genes, operons, or multiple members of related families of glycoside hydrolases, transporters, and regulatory proteins (3, 5, 76). Previous studies in *T. maritima* have utilized by using methodologies discussed previously (11, 24). Hybridizations were carried out for 18 h according to modified TIGR protocols described elsewhere (11, 20, 21). Hybridized slides were scanned on a Perkin-Elmer ExpressLite Scanner (Perkin-Elmer) and quantitated by using ScanArray 2.1 (Perkin-Elmer).

**MATERIALS AND METHODS**

**Growth of *T. maritima* and RNA isolation.** Cultures of *T. maritima* MS8 were grown in artificial seawater using optical density measurements and epifluorescence microscopic cell density enumeration, as described previously (12). Growth substrates glucose, mannose, arabinose, rhamnose, ribose, xylose, β-xylan (birchwood), laminarin (*Laminaria digitata*), and starch (potato) were obtained from Sigma (St. Louis, MO). Galactomannan (carob), glucomannan (konjac), and starch were obtained from Calbiochem (San Diego, CA).

**Microarray protocols.** A *T. maritima* cDNA microarray was constructed and utilized by using methodologies discussed previously (11, 24). Hybridizations were carried out for 18 h according to modified TIGR protocols described elsewhere (11, 20, 21). Hybridized slides were scanned on a Perkin-Elmer ExpressLite Scanner (Perkin-Elmer) and quantitated by using ScanArray 2.1 (Perkin-Elmer).

**Mixed model analyses of microarray data.** Replication of treatments, arrays, dyes, and cDNA spots allowed the use of analysis of variance (26, 83) models for data analysis. A loop design was constructed (Fig. 1), and reciprocal labeling utilized for all samples to estimate dye effects for each treatment. Scanarray spot intensities were imported into SAS (SAS Institute, Cary, NC) and flagged low-intensity or low-quality spots were removed before further analysis. After local background subtraction and log transformation of spot intensities, a linear normalization analysis of variance model was used to estimate global variation in the form of fixed effects (dye [D], treatment [T]), random effects (array [A], spot A [S]), and random error by using the model log2(yijklmn) = μ + A + D + \(T_k + A(S_i + e_{ijklmn})\)), A gene-specific analysis of variance model was used to partition the remaining variation into gene-specific effects using the model log2(yijklmn) = μ + A + D + \(T_k + A(S_i + e_{ijklmn})\)). Least-squares mean estimates of gene-specific treatment effects were examined by using hierarchical clustering in JMP (SAS Institute), and histograms in Excel (Microsoft) were used to visualize expression patterns for specific contiguous genomic locations. A subset of sam-

### Table 1. Carbon sources used in this studya

<table>
<thead>
<tr>
<th>Poly/monosaccharide</th>
<th>Source</th>
<th>Backbone structure</th>
<th>Side chain</th>
<th>Molecular mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabinose</td>
<td>NA</td>
<td>Ara</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>Glucose</td>
<td>NA</td>
<td>Glc</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Mannose</td>
<td>NA</td>
<td>Man</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Rhamnose</td>
<td>NA</td>
<td>Rha</td>
<td></td>
<td>182</td>
</tr>
<tr>
<td>Ribose</td>
<td>NA</td>
<td>Rib</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>Xylose</td>
<td>NA</td>
<td>Xyl</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>Galactomannan</td>
<td>Carob</td>
<td>(Manβ1→4Man)n</td>
<td>Gal(α1→6)</td>
<td>−1,000,000</td>
</tr>
<tr>
<td>Glucomannan</td>
<td>Konjac</td>
<td>(Glcβ1→4Man)n</td>
<td>−250,000</td>
<td></td>
</tr>
<tr>
<td>β-1,3/β-1,4-Glucan</td>
<td>Barley</td>
<td>(Glcβ1→3,4Glc)n</td>
<td>−90,000</td>
<td></td>
</tr>
<tr>
<td>Laminarin</td>
<td>L. digitata</td>
<td>(Glcβ1→3Glc)n</td>
<td>−1,000,000</td>
<td></td>
</tr>
<tr>
<td>Pustulan</td>
<td>U. papulosa</td>
<td>(Glcβ1→6Glc)n</td>
<td>−5,000</td>
<td></td>
</tr>
<tr>
<td>Starch</td>
<td>Potato</td>
<td>(Glcα1→4Glc)n</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>β-Xylan</td>
<td>Birchwood</td>
<td>(Xylβ1→4Xyl)n</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Chitin</td>
<td>Crab shells</td>
<td>(Nagβ1→4Nag)n</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

a NA, not available; Ara, arabinose; Glr, glucuronic acid; Nag, N-acetylgalactosamine.
FIG. 1. Loop design used for the study of carbon source utilization of *T. maritima* in the present study. The arrowheads correspond to the Cy5 channel, and the dotted arrow ends correspond to the Cy3 channel. Abbreviations for sugar names used in subsequent expression histograms are shown in parentheses.

expression results for selected sets of genes in the *T. maritima* genome. These include ABC-type bacterial carbohydrate uptake transporters from the two main families, CUT1 and CUT2 (64), as well as members of the Opp/Dpp ABC transporter family. Expression data are also shown for associated hydrolases and putative transcriptional regulators. Based on these results, predictions of transporter specificities are shown in Table 2, along with a summary of specificities predicted by previous work.

CUT1 ABC transport systems. CUT1 transporters with substrate-binding proteins related to maltose-binding proteins include maltose transporter subunits (TM1836 and TM1839 and TM1202 to TM1204) recently shown to have different expression results for selected sets of genes in the *T. maritima* genome. These include ABC-type bacterial carbohydrate uptake transporters from the two main families, CUT1 and CUT2 (64), as well as members of the Opp/Dpp ABC transporter family. Expression data are also shown for associated hydrolases and putative transcriptional regulators. Based on these results, predictions of transporter specificities are shown in Table 2, along with a summary of specificities predicted by previous work.

The CUT1 permeases and substrate-binding proteins encoded by TM0810 to TM0813 are found with genes whose functions relate to breakdown of N-acetylglucosamine polysaccharides. However, growth of *T. maritima* in the presence of the β-1,4-N-acetylglucosamine polymer chitin was similar to control cultures, a finding consistent with a lack of differential expression of this locus. Sequence similarity searches suggest that *T. maritima* lacks an identifiable chitinase and might instead utilize chitin in the presence of neighboring species capable of chitin hydrolysis. Alternatively, transcription of these genes may be higher in the presence of N-acetylglucosamine or another N-acetylglucosamine-containing oligosaccharide found in the natural environment of *T. maritima*.
CUT2 transport systems. The two CUT2 transporters found in *T. maritima* are comprised of a substrate-binding protein, a single permease subunit presumed to form a homodimer in the functional transporter and a fusion protein consisting of two nucleotide-binding domains. Previously, we observed the up-regulation of the LacI family gene TM0949 and the predicted ribokinase TM0960 during growth on xylose (11). Computational analysis of LacI regulators has determined that TM0949 is most similar to RbsR, a negative regulator of ribose uptake (17). Here, several genes within the TM0949-to-TM0960 gene string were upregulated during growth on xylose, ribose and arabinose (Fig. 3A), including *rbsABCD* homologs not examined previously by Chhabra et al. (11). From expression results alone, it is unclear whether this system can import multiple pentose sugars or whether transcription of the genes is triggered by the interconversion of xylose or arabinose to ribose via the pentose phosphate pathway (Fig. 3C). Two strong matches to a LacI family consensus binding site are arrayed consecutively upstream of the ribokinase TM0960 (Table S2 in the supplemental material), and a predicted rho-independent terminator located downstream of TM0949 is the only identifiable terminator within the gene cluster (Fig. 3A) (16). Similar to observations of other transport systems of *T. maritima* discussed below, the putative binding protein of this transport set (TM0958) was more highly upregulated than other transporter components.

Expression results shown here suggest several clarifications of *T. maritima* genome annotation and the *T. maritima* pentose
phosphate pathway as predicted by sequence similarity in the KEGG database (25). The predicted KEGG pathway identifies an RpiB homolog responsible for the interconversion of ribulose-5-phosphate to ribose-5-phosphate as TM1080, which was detected at similar levels on all substrates (data not shown). Given the lack of additional glycerol utilization genes nearby TM0952, a role for the encoded protein as an inducible xylulokinase should be considered.

Two hypothetical proteins of unknown function within the TM0949–to-TM0960 locus are also differentially expressed. The functions of these proteins remain unclear, but their up-regulation during growth on multiple pentoses suggests a plausible role in pentose uptake or catabolism. TM0950, which is located in a gene cluster with similar composition to the T. maritima pentose-responsive locus, including a putative sugar kinase (TTE0216) classified into COG2971 in the clusters of relatives (COG) (74). A related T. maritima pentose-responsive locus, including a putative sugar isomerase (LJ1064), a LacI family regulator (LJ1265), N- and C-terminal transketolase subunits (LJ1266–1267), and an FGGY family sugar kinase should be considered.

Homologs to the RbsABC ABC transporter subunits bearing 40 to 52% identity to the T. maritima homologs are found together in the genome of the hyperthermophile Thermotoga maritima (LJ1257), contains no known domains. However, LJ1257 is located in a gene cluster with similar composition to the T. maritima pentose-responsive locus, including a putative sugar isomerase (LJ1064), a LacI family regulator (LJ1265), N- and C-terminal transketolase subunits (LJ1266–1267), and an FGGY family sugar kinase.
xylose polymer xylan. The annotation for sugar specificity of this putative kinase is apparently drawn from a distantly related human N-acetylglucosamine kinase (22), since the specificities of closely related microbial homologs have not yet been determined.

The second *T. maritima* CUT2 transport set is found with the XylR family regulator TM0110. Despite the lack of an *rbsD* cytoplasmic sugar-binding homolog, genes homologous to *rbsABC* are all present. Transcripts of TM0110 were detected at higher levels during growth on xylose compared to all other substrates tested here except laminarin, although other genes within this gene string were not significantly differentially expressed between xylose and any other sugar. Unlike the xylose catabolic genes of many model organisms, the characterized *T. maritima* xylose isomerase (TM1667) (1) is not found with the predicted xylulokinase (TM0116). This separation might reflect a broader physiological specificity of the TM1667 enzyme, which has also been used in the conversion of glucose to fructose (1) or may reflect differential regulation of the two activities in response to different xylose-containing substrates.

In addition to genes within the TM0949-to-TM0960 locus described above, other *T. maritima* genes also respond to the simple sugar L-arabinose, including a characterized L-arabinose isomerase (TM0276) (33), an α-l-arabinofuranosidase (TM0281), an uncharacterized conserved protein (TM0280), and a homolog to the protein araM from the *B. subtilis* arabinoce utilization operon (66) (Fig. 3B). Located upstream of these genes is a LacI family regulator, TM0275, which is most
similar to AraR from a *Geobacillus stearothermophilus* arabinose cluster (Table S2 in the supplemental material). Sugar ABC permease subunits TM0278 and TM0279 do not show strong differential regulation and, together with a frameshifted substrate-binding protein (TM0277), suggest a nonfunctional transporter.

**Sequence analysis of Opp/Dpp transporters subunits in *T. maritima***. Taken together, the well-documented ability of *T. maritima* to use complex carbohydrates and the lack of annotated polysaccharide transporters suggested novel oligosaccharide transporters yet to be identified in the *T. maritima* genome. The high degree of identity between the Dpp/Opp family cellobiose transporter of *P. furiosus* and a likely cellobiose transporter of *T. maritima* (29) has raised the possibility that additional related transporters of *T. maritima* might transport oligosaccharides. The phylogeny of Opp/Dpp transport subunits in the COG database (74) and BLAST homology searches (Table S2 in the supplemental material) suggested three different lineage-specific gene expansions likely to have taken place after the divergence of *T. maritima* from the next closest sequenced organism. A consensus tree based on substrate-binding protein relationships with operon organizations superimposed is shown in Fig. 4. Duplication or acquisition of fully intact Opp/Dpp ABC transport operons (one substrate-binding protein, two permeases, and two ATP-binding subunits) can be inferred, although three solitary substrate-binding proteins are also apparent (Fig. 4 and Table 2). In two instances, these proteins display high levels of homology (>60% identity) to substrate-binding proteins of full transport systems, perhaps suggesting interaction with subunits of other transport systems. Duplication of Opp/Dpp substrate-binding proteins in *T. maritima* might accomplish expansion of sugar binding capabilities for related substrates, since the peptide specificities of two Opp/Dpp family transporters of *Lactococcus lactis* IL1403 have been largely attributed to features of substrate-binding proteins (14, 67).

**Expression of group 1 Opp/Dpp transporters is elevated during growth on β-linked gluco-oligosaccharide substrates**. Three related substrate-binding proteins detected at higher levels during growth with β-linked sugars can be classified into group 1 of Opp/Dpp family transport operons (Fig. 4). These proteins share considerable similarity with a *P. furiosus* transporter implicated in the uptake of β-1,4 linked glucose oligomers, including cellobiose, cellotriose, and laminaribose (29). Sequence similarity patterns suggest that this group likely arose from lateral gene transfer of one or two transport sys-
tems from archaea, followed by duplication of the sugar-binding protein and divergence of regulatory strategies and expression specificity (74). We have previously noted the location of a tightly conserved palindromic sequence motif similar to the LacI family consensus upstream of selected genes responsive to growth on carboxymethyl cellulose, barley glucan, glucomannan (11), and cellobiose (C. I. Montero and K. R. Kelly, unpublished observations). The putative regulator binding sites are situated between the $\beta$-1,4-linked glucanase binding protein (TM1224), a putative glycosylase (TM1225), a second CbtA homolog (TM1226, 60% identity with TM1223), and ManB (TM1227), a characterized $\beta$-mannanase (55). TM1224 to TM1227 are upregulated on mannose-containing carbon sources (Fig. 5A), a finding consistent with the ability of the carbohydrate-binding domain of TM1227 to accommodate manno-oligosaccharides, galactomannan, and glucomannan degradation products (8). The specificity of the upregulation of TM1224 (here designated ManR) and TM1226 (here designated MbtA) on mannose, glucomannan, and galactomannan is especially striking. It appears likely that TM1226 might interact with the ATP-binding and permease subunits of the cellobiose transporter. In agreement with their close phylogenetic grouping, TM1223 and TM1226 are reciprocal best BLAST homologs; this suggests possible past duplication and specialization of the binding protein for the cellobiose transporter to accommodate mannan oligosaccharides. It also appears likely that expression of TM1226 could be under the transcriptional control of TM1224. Other candidates for regulation by TM1224 include genes previously observed to be
upregulated on glucomannan and galactomannan within the TM1745-to-TM1752 gene string and the β-mannosidase TM1624 (11) (Fig. 5B). Although the OppA-family binding protein TM1746 is more highly upregulated during growth on xylose, other components of the transporter (TM1747 to TM1750), two endoglucanases (Cel5A TM1751 and Cel5B TM1752), and a β-mannosidase (TM1624) are highest during growth on mannans. We propose to designate the transporter components TM1746 to TM1750 as MtpABCDF (Table 2).

The third substrate-binding protein of group 1 (TM0031) is located within a gene string encoding the laminarinase BglB/Cel3 (TM0025) (50, 89) and laminarinasme TM0024 (7, 88), as well as components of an ABC transport complex homologous to CbtABCDF not examined by Chhabra et al. (11). A XylR family regulator (TM0032) is located upstream of the ABC transporter components. Higher transcript levels for components of this transporter during growth on the β-1,3-linked glucose polymer laminarin, the mixed β-1,3-β-1,4-linkage glucose polysaccharide barley, and the β-1,6-linked glucose polymer pestulan may suggest a general role in the uptake of β-linked sugars (Fig. 5C). We suggest the designations BgtpABCDF and Bgkle for the transporter and regulator, respectively (Table 2). Similarly to CbtA of the cellulose transporter, TM0031 (BgtpA) was detected at higher levels in the presence of β-glucans than other transporter components, and a GC-rich inverted repeat was found in the intervening sequence between BgtpA and TM0030 (BgtpB) (Table S2 in the supplemental material). A proposed pathway for the uptake and utilization of β-glucan and β-mannan oligosaccharides is shown in Fig. 5D.

**Higher transcript levels of group 2 and 3 Opp/Dpp transporters during growth on xylose and xylose-containing oligosaccharides.** Components of two distinct Opp/Dpp family transporters were detected at higher levels in the presence of the simple sugar xylose and the polysaccharide xylan (Table 1). The two sets of transport proteins are located nearby one another, separated by a set of genes predicted to encode enzymes for the catabolism of uronic acids. We have previously noted the similarities in functional composition of this gene cluster (11) to the xyulose utilization cluster of *Geobacillus stearothermophilus* T-6 (68). Both sets of *T. maritima* transporters are divergently transcribed from family 10 xyylanases (xylA/xyl10A, TM0061; xylB/xyl10B, TM0070) (Fig. 6), both reported previously to be active on xylan polysaccharides (13, 43, 79, 81, 82, 87). The similarities in expression profiles and gene content of the two gene sets do not appear to be the result of a recent duplication, as reflected in the consensus phylogenetic tree of Opp/Dpp family transport components (Fig. 4). Comparison with sequences from other sequenced organisms reveals that the TM0071-to-TM0075 gene set clusters with two other *T. maritima* ABC transporter sets in a grouping which apparently arose from a lateral gene transfer event with archaea (74), likely followed by duplication and divergence within an ancestral lineage (group 2, Fig. 4). In contrast, TM0056 to TM0060 cluster within a group of bacterial transporter proteins (74). The genomic arrangement of the two xylose and xylan-responsive transporter gene sets also differs (Fig. 6).

We propose to designate TM0071 to TM0075 as XtpABCDF in keeping with the names assigned to orthologous proteins found in an unpublished cluster of xylan utilization genes from *T. neapolitana* (GI:23270356). Transcript levels of TM0071 were slightly higher on xylose than xylan, whereas other members of the transport operon showed various degrees of preference for xylose over xylan (Fig. 6B). In contrast, the substrate-binding protein TM0056 was detected at much higher levels during growth on xylan than xylose (Fig. 6A). Transcripts from the remaining transporter subunits (TM0057 to TM0060) and the α-glucuronidase AguA (TM0055) (62) were detected at higher levels during growth on xylose and xylan compared to nonxylose sugars. We propose to designate the transport proteins encoded by TM0056 to TM0060 as XtpGHJLM (Table 2). The DppA substrate-binding protein encoded by TM0309 (proposed designation XtpN) is closely related to TM0056 (Table S2) and is found nearby a predicted α-xyllosidase (TM0308). The slight upregulation of both TM0309 (proposed designation XtpN) and TM0310 on xylan may suggest a role for these proteins (and possibly TM0308) in uptake and hydrolysis of an undetermined xylose-containing polysaccharide (Fig. 2).

The variation in expression patterns for the xylanase-associated ABC transporters may relate to differences in the carbohydrate binding specificity for *T. maritima* Xyl10A and Xyl110B. XylA contains four carbohydrate binding domains absent in XylB: the A1 and A2 domains of XylA have been shown to bind xylan while the C1 and C2 domains bind cellulose and a number of other monosaccharides and polysaccharides (7). The hydrolase content of the two gene strings also differs, suggesting likely specialization of the transporters for differently substituted xylan degradation products. A β-xyllosidase (86) and acetyl xylan esterase colocalize with TM0071 to TM0075, and an α-glucuronidase colocalizes with TM0056 to TM0060. Although no regulatory proteins are located within either xylanase-transporter gene string, similar inverted repeat sequences found upstream of Xyl10A (TM0061), Xyl110B (TM0070), and a putative α-xyllosidase of glycosyl hydrolase family 31 (TM0308) share similarity with the consensus for a XylR family regulator (Table S2 in the supplemental material). The XylR family regulator (TM0110) is expressed more highly on xylose and laminarin than any other substrate. The observation of similar expression profiles on xylose and laminarin may relate to the co-occurrence of carbohydrate binding domains for binding xylan and mixed linkage glucan carbohydrates by distinct domains in Xyl10A (8) or reflect sequence similarity between the XylR family regulators TM0110 and TM0032 (BglcR).

**Higher transcript levels of a group 3 Opp/Dpp transporter during growth on the simple sugar rhamnose.** Growth of *T. maritima* on L-rhamnose (a methyl pentose also known as deoxy-1-mannose) had not been previously demonstrated. Within group 3 of the Opp/Dpp family transporters of *T. maritima* is a set of ABC transporter components which colocalize with predicted rhamnose catabolic genes (31, 48) (Fig. 7A). Here, the majority of genes that showed higher transcript levels during growth on rhamnose are found in this locus (Fig. 7A). Transcripts of nearly all genes encoding subunits of the transporter (TM1063 to TM1067) were observed at higher levels during growth on rhamnose compared to all other sugars examined here. We suggest the designation RtpABCDF for these transport components (Table 2). Similar to related transport systems, a GC-rich inverted repeat was found in the intervening sequence between RtpA (TM1067) and RtpB.
FIG. 5. Expression results for transcripts detected at higher levels on β-linked polysaccharides. Small hairpin symbols represent locations of GC-rich inverted repeats, while large hairpin symbols represent locations of predicted rho-independent terminators (http://www.tigr.org/software/TransTermResults/btm.html). Predicted ρ^+ promoters are represented by arrows, and an asterisk denotes the position of a putative cellobiose regulator operator. Substrate-binding proteins are outlined in bold and boxed. Spacing between genes is less than 30 bases unless indicated otherwise. (A) Genes within a putative cellobiose transport operon (proposed designation CbTABCDF), including a likely regulator of cellobiose uptake and utilization (proposed designation CelR) and a colocalized mannan-responsive locus (including TM1224 [proposed designation ManR] and TM1226 [proposed designation MbtA]). (B) Genes within a glucomannan and galactomannan responsive locus include the Opp transporter components TM1746 to TM1752 (proposed designation MtpABCDF), Cel5A, and Cel5B and the β-mannosidase TM1624. (C) Genes within a β-glucan responsive locus (proposed designation BgtpABCDF), including a putative regulator of β-glucan uptake (proposed designation BgICR). (D) Predicted pathway for the utilization of β-glucans and glucomannan by T. maritima. Extracellular enzymes responsible for polysaccharide hydrolysis are shown, as well as periplasmic binding proteins, membrane-embedded permeases, associated ATP-binding subunits, and intracellular hydrolases. References for hydrolases shown in the pathway are listed in Table S1 in the supplemental material.
The presence of an α-glucuronidase (TM1056, Agu4C) and β-glucuronidase (TM1062) within this locus suggest that the ABC transporter encoded by TM1063 to TM1067 might also be involved in the uptake of rhamnose-containing disaccharides or oligosaccharides that include glucuronic acid residues. A second candidate rhamnose transporter is encoded by TM1060, which shares sequence similarity with major facilitator superfamily sugar-proton symporters. Although the likely L-rhamnulose aldolase RhaD (TM1072) (53) and predicted rhamnulokinase RhaB (TM1073) are homologous to E. coli K-12 rhamnose catabolic genes, an RhaA rhamnose isomerase homolog is missing. A likely substitute is TM1071, annotated as a putative sugar isomerase, which is homologous to rhamnose isomerase RhaI of Rhizobium leguminosarum bv. trifolii (52) and Bacteroides thetaiotaomicron VPI-5482 (85). Several hypothetical proteins within the rhamnose locus present interesting targets for further work (Table S2 in the supplemental material).

Based on this analysis, a predicted pathway for rhamnose utilization in *T. maritima* is shown in Fig. 7B. Expression data from this locus also suggest a potential mechanism for transcriptional regulation. A DeoR/GlpR family transcriptional regulator (TM1069, COG1349) found within the rhamnose transport and catabolism cluster shares sequence identity with proteins found within rhamnose catabolic clusters of *Bacillus halodurans* (72) and *Oceanobacillus iheyensis* (73). Therefore, we propose to designate TM1069 as RhaR.

**Opp/Dpp transporters of unknown specificity.** Expression data and genomic neighborhood analysis did not reveal specific substrate preferences for several sets of Opp/Dpp family transporter components (Table 2). Further work will be necessary to clarify whether these proteins are involved in uptake of untested sugars or alternative substrates transported by other members of the Opp/Dpp transporter family, such as metal ions (84) or peptides (19, 57).

**DISCUSSION**

The combination of microarray data with gene neighborhood and sequence analysis represents a powerful high-throughput approach for examining gene regulation and predicting functional roles of genes for microorganisms which lack genetic systems. Here, the genomic contexts and transcriptional responses of *T. maritima* genes to 14 monosaccharide and polysaccharide substrates were examined to improve upon previous annotations of ABC transporter proteins (48, 59). However, similar approaches could be used to perform substrate-by-substrate analysis of transcriptional responses to additional carbohydrates, peptides, metals, antibiotics, or other elements in microbial species for which microarrays are avail-
able. As demonstrated here, a loop experimental design allows efficient collection of large microarray datasets. Mixed model analysis of these datasets then enables comparisons between transcript levels for all pairs of substrates, rather than limiting comparisons to an arbitrary reference condition. In this case, this approach allowed greater flexibility in comparing responses to both changes in carbohydrate composition and branch type.

The transcriptional data presented here support the hypothesis that many members of the Opp/Dpp ABC transporter family of *T. maritima* are involved in carbohydrate transport, and explain the observation that glycoside hydrolases often colocalize with these genes. Given the differential regulation of related Opp/Dpp transport systems in response to carbohydrates, this strategy has likely allowed the acquisition of new uptake capabilities, perhaps assisting in the adaptation of *Thermotoga* species to specific environments. Transcriptional information was especially helpful in suggesting candidate substrates for several Opp/Dpp gene sets resulting from apparent lateral gene transfer followed by duplication and divergence (Fig. 4). In two cases, the next closest related transporter gene sets are found in archaea. In total, carbohydrate specificities were proposed for six full or partial operons of Opp/Dpp transporter subunits, and expression results were confirmed for two operons previously examined (11). The results obtained will assist in streamlining biochemical characterizations of sub-

---

**FIG. 6.** β-Xylan and xylose-responsive operons from groups 2 and 3 of Opp/Dpp family transporters. Small hairpin symbols represent locations of GC-rich inverted repeats found between substrate-binding proteins and other transporter subunits, and large hairpin symbols represent locations of predicted rho-independent terminators (http://www.tigr.org/software/TransTermResults/btm.html). Predicted σ^54^ promoters are represented by arrows, and an asterisk denotes the positions of putative xylan/xylose regulator operator. Substrate-binding proteins are outlined in bold and boxed. Spacing between genes is less than 30 bases unless indicated otherwise. (A) XtpGHJLM, a predicted xylooligosaccharide transport system, is divergently transcribed from xylanase Xyl10A. (B) XtpABCDF, a predicted xylose/xyloside transport system, is divergently transcribed from xylanase Xyl10B.
FIG. 7. Rhamnose responsive locus containing Opp/Dpp family transporter from group 3 of Opp/Dpp family transporters. Small hairpin symbols represent locations of GC-rich inverted repeats found between substrate-binding proteins and other transporter subunits. Predicted ρ^− promoters are represented by arrows. Substrate-binding proteins are outlined in bold and boxed. Spacing between genes is less than 30 bases unless indicated otherwise. (A) A rhamnose-responsive locus contains candidate genes likely to encode enzymes responsible for the transport and hydrolysis of rhamnose-containing di- or oligosaccharides, and the complete catabolism of the simple sugar L-rhamnose. (B) A predicted pathway for the utilization of L-rhamnose by *T. maritima*. A periplasmic binding protein, membrane-embedded permeases, associated ATP-binding subunits, and rhamnose catabolic enzymes are shown.
strate-binding protein specificities for *T. maritima* in progress in our laboratory and others. Although *T. maritima* does not grow on peptides as a sole carbon source, it is still unclear whether any of its Opp/Dpp transport systems are involved in peptide import. However, transcripts of an Opp/Dpp family transporter operon (TM0500 to TM0503) which lacks a substrate-binding protein are detected at higher levels in high density cocultures of *T. maritima* and *Methanococcus jannaschii* (24). This transporter may be involved in export of a small peptide (TM0504) located downstream of the transporter which has been implicated in quorum sensing and biofilm formation (24).

Differential expression information for predicted carbohydrate-responsive transcriptional regulators of *T. maritima* has now assisted in the prediction of putative functions for previously unannotated members of the LacI (three proteins), XylR (three proteins), and GlpR/DeoR (one protein) families. These include candidates for the control of the uptake of β-glucans (TM1218 and TM0032), xylose/xylan (TM0110 and TM0949), arabinose (TM0275), and rhamnose (TM1069). Sequences resembling binding sites can be detected upstream of selected carbohydrate-responsive genes (Table S2 in the supplemental material), further supporting the hypothesis that some or all of these proteins are involved in the regulation of carbohydrate import. The specificities of most of these regulators would have been impossible to determine from sequence analysis alone, but transcriptional data now offer insights into plausible substrates for further characterization efforts.

For several Opp/Dpp transporter operons, substrate-binding proteins showed greater transcriptional responses to changes in carbon source than did other transporter subunits. If transcript levels correlate well with protein levels, this might indicate that transporter subunits are present in the absence of substrate. Increased transcription of substrate-binding proteins could allow maximal capture of available carbohydrates to be transported by existing permease and ATPase subunits. A partial explanation for the differential regulation of Opp/Dpp substrate-binding proteins relative to other transporter subunits is suggested by the presence of GC-rich hairpin structures in the intervening sequence between the subunits. The possibility that these hairpin sequences act as partial transcriptional terminators should be explored further.

In contrast to the Opp/Dpp family transporters, most members of known carbohydrate transporter families were not differentially expressed here. A notable exception was the CUS2 transporter which showed transcriptional responses to ribose, arabinose, and xylose. It is possible that other predicted sugar transporters respond to substrates not tested here. For example, maltose and lactose were not examined, but CUS1 transporters of *T. maritima* do respond to the presence of these sugars (51) (Table 2). The possibility remains that some *T. maritima* sugar transport operons are transcribed constitutively, perhaps independent of the control of local transcriptional regulators.

The lack of PTS system components in the *T. maritima* genome argues against mechanisms of global catabolite repression identical to those operating in gram-negative and gram-positive model organisms. Future work using microarrays to examine data from growth experiments with combinations of substrates will be needed to explore alternative mechanisms of preferred substrate utilization *T. maritima*. Although an alternative mechanism cannot be ruled out, the proximity of regulators and differentially expressed genes involved in sugar utilization provides evidence that local transcriptional regulators play important roles in regulating uptake of individual sugars through inducible ABC transport systems. Inducible and independent transcriptional control of transport systems of varied specificities may assist *T. maritima* in discriminating between and responding to complex polysaccharides found in its natural environment.

ACKNOWLEDGMENTS
This study was supported in part by grants from the National Science Foundation (grants BES-0317886 and BES-0000487) and the Department of Energy, Energy Biosciences Program (grant DE-FG02-96ER20219).

REFERENCES