The hyperthermophilic bacterium *Thermotoga maritima* MSB8 was grown on a variety of carbohydrates to determine the influence of carbon and energy source on differential gene expression. Despite the fact that *T. maritima* has been phylogenetically characterized as a primitive microorganism from an evolutionary perspective, results here suggest that it has versatile and discriminating mechanisms for regulating and effecting complex carbohydrate utilization. Growth of *T. maritima* on monosaccharides was found to be slower than growth on polysaccharides, although growth to cell densities of $10^6$ to $10^9$ cells/ml was observed on all carbohydrates tested. Differential expression of genes encoding carbohydrate-active proteins encoded in the *T. maritima* genome was followed using a targeted cDNA microarray in conjunction with mixed model statistical analysis. Coordinated regulation of genes responding to specific carbohydrates was noted. Although glucose generally repressed expression of all glycoside hydrolase genes, other sugars induced or repressed these genes to varying extents. Expression profiles of most endo-acting glycoside hydrolase genes correlated well with their reported biochemical properties, although exo-acting glycoside hydrolase genes displayed less specific expression patterns. Genes encoding selected putative ABC sugar transporter genes were found to respond to specific carbohydrates, and in some cases putative oligopeptide transporter genes were also found to respond to specific sugar substrates. Several genes encoding putative transcriptional regulators were expressed during growth on specific sugars, thus suggesting functional assignments. The transcriptional response of *T. maritima* to specific carbohydrate growth substrates indicated that sugar backbone- and linkage-specific regulatory networks are operational in this organism during the uptake and utilization of carbohydrate substrates. Furthermore, the wide ranging collection of such networks in *T. maritima* suggests that this organism is capable of adapting to a variety of growth environments containing carbohydrate growth substrates.

Saccharolytic microorganisms employ a range of proteins to hydrolyze, transport, and utilize complex carbohydrates that serve as carbon and energy sources (1). In some cases, these proteins are very specific to particular carbohydrates, whereas in other situations they mediate the processing of a broader range of glycosides. For simple sugars, such as glucose, binding and transport proteins alone mediate substrate entry into specific intracellular anabolic and catabolic pathways (2). However, for complex carbohydrates, a series of glycoside hydrolases must first process the polysaccharide so that its backbone and side chain glycosidic linkages are hydrolyzed to the extent needed for binding, transport, and intracellular utilization. How specific organisms develop the capacity to utilize complex carbohydrates is not known, but this probably involves evolutionary pressures in addition to acquisition of this genetic potential through horizontal gene transfer events. In any case, a microorganism’s capacity to utilize carbohydrates presumably reflects the availability of such substrates in its habitat. Therefore, insights into the repertoire of carbohydrate-active proteins in a given organism and how the expression of these proteins is regulated would reveal much about particular metabolic features in addition to how it interacts within a given ecosystem.

*Thermotoga maritima* is an obligately anaerobic, heterotrophic, hyperthermophilic bacterium originally isolated from geothermal features associated with Vulcano Island, Italy (3). Its capacity to utilize a wide range of simple and complex carbohydrates was confirmed by the inventory of glycoside hydrolases encoded in its genome (4). In fact, the *T. maritima* genome, despite its relatively small size, encodes the largest number of glycoside hydrolases of any bacterial or archaean genome sequenced to date (see Fig. 1). From growth experiments and characterization of specific glycoside hydrolases (5), *T. maritima* is known to metabolize both polysaccharides and simple sugars, including carboxymethylcellulose, barley glucan, starch, galactomannan (5), xylan (6), pectin (7), mannose, xylose, and glucose (2). In some cases, the proteins involved in the processing, transport, and utilization of these glycosides can be inferred from their apparent organization into operons in the *T. maritima* genome sequence (4), whereas in other cases such classification is not clear. Regulation of genes encoding specific carbohydrate-active proteins in *T. maritima* has only been studied to a limited extent thus far (5, 7), and the coordinated regulation of related genes involved in polysaccharide utilization has not been examined.

Here, a targeted cDNA microarray, based on carbohydrate-active proteins from *T. maritima*, was used in conjunction with mixed model analysis (8, 9) to explore issues related to...
saccharide utilization by this organism. Despite the fact that T. maritima has been phylogenetically characterized as a primitive microorganism from an evolutionary perspective (10), results here support that it has versatile and discriminating mechanisms for regulating and effecting complex carbohydrate utilization. The relative importance of evolutionary processes and horizontal gene transfer (4) in developing its carbohydrate utilization capacity is not known, but T. maritima’s ability to respond to various substrates in its growth environment underlies its ubiquity in global geothermal settings (11).

EXPERIMENTAL PROCEDURES

Construction of the Targeted cDNA Microarray—Open reading frames (total of 269) of known and putative genes related to sugar processing and other related metabolic functions were identified through BLAST (12) comparisons of protein sequences from the T. maritima MSBS genome available on the World Wide Web at www.tigr.org/tigrscripts/CMR2/GenomePage3.spl?database/H11005 btm. DNA primers were designed with similar annealing temperatures and minimal hairpin formation using Vector NTI 7.0 (Informax, Bethesda, MD). The selected probes were PCR-amplified in a PTC-100 Thermocycler (MJ Research, Inc., Waltham, MA) using Taq polymerase (Roche Molecular Biochemicals) and T. maritima genomic DNA, isolated as described previously (5). The integrity and concentration of the PCR products were verified on 1% agarose gels. PCR products were purified to 100 ng/μl using 96-well QIAquick PCR purification kits (Qiagen, Valencia, CA), resuspended in 50% Me₂SO, and printed onto CMT-GAPS aminosilane-coated microscope slides (Corning Glass) using a 417 Arrayer (Affymetrix, Santa Clara, CA) in the North Carolina State University Genome Research Laboratory (Raleigh, NC). Eight replicates of each gene fragment were printed onto each slide. The DNA was then attached to the slides by UV cross-linking using a GS GeneLinker UV Chamber (Bio-Rad) set at 250 mJ and baked at 75 °C for 2 h.

Growth of Thermotoga maritima and RNA Isolation—Growth of T. maritima MSBS cultures in artificial sea water was followed using optical density measurements and epifluorescence microscopic cell density enumeration, as described previously (5). Growth substrates glucose, mannose, xylose, β-xylan (birchwood), laminarin (Laminaria digitata), and starch (potato) were obtained from Sigma. Galactomannan (carob), glucomannan (konjac), carboxymethylcellulose, and β-glucan (barley) were obtained from Megazyme (Wicklow, Ireland). Growth substrates were prepared as described previously (5) and included in the medium at a final concentration of 0.25% (w/v). Substrate purities as provided by the manufacturers varied from 95 to 99%. To ensure minimum carryover between substrates, cells were grown for at least 10 passes on each carbon source using a 0.5% (v/v) starting inoculum before obtaining the growth curves. Specific growth rates on mono- and polysaccharide substrates were determined from the slopes of semilog plots of exponential cell growth versus time. Isolation of total RNA from T. maritima was performed on cells that were grown until early- to mid-exponential phase on the various growth substrates, as described in detail previously (5).

Labeling and Hybridization—First-strand cDNA was prepared from T. maritima total RNA using Stratascript (Stratagene, La Jolla, CA) and random hexamer primers (Invitrogen) by the incorporation of 5-[3-aminoallyl]-2'-deoxyuridine-5'-triphosphate (Sigma) as described elsewhere (13). The slides were scanned using a Scanarray 4000 scanner (GSI Lumonics and Billerica) in the North Carolina State University Genome Research Laboratory. Signal intensity data were obtained using Quantarray (GSI Lumonics).

Experimental Design and Data Analysis—A loop design was constructed (see Fig. 2) to ensure reciprocal labeling for all 10 different experimental conditions. Replication of treatments, arrays, dyes, and cDNA spots allowed the use of analysis of variance (ANOVA)2 models

2 The abbreviations used are: ANOVA, analysis of variance; CMC, carboxymethylcellulose; PTS, phosphotransferase system; CCR, carbon catabolite repression.
and random (array, pin within array, pin spot within array) effects. A linear normalization ANOVA model (9) of log base 2 intensities was through independent calculations in Excel (Microsoft, Seattle, WA). A set (SAS Institute, Cary, NC). The data import system was verified.

The carbon source utilization in T. maritima. The arrow ends correspond to the Cy3 and Cy5 channels as follows: Cy3 (dots), Cy5 (arrowheads).

### Table I

| Poly saccharide | Source | Backbone structure | Side chain | Mass (Da) |
|-----------------|--------|--------------------|------------|-----------|
| Glucose         | NA     | Glec               | NA         | 180       |
| Mannose         | NA     | Man                | NA         | 180       |
| Xylose          | NA     | Xyl                 | NA         | 150       |
| Galactomannan   | Carob  | (Man β1→4 Man)₅s   | Gal (α1→6) | NA        |
| Glucosan        | Konjac | (Glc β1→4 Man)₅s   | NA         | 100,000   |
| β-1,3/1,4-Glucan| NA     | (Glc β1→4 Glc)₅s   | NA         | 90,000    |
| Laminarin       | L. digitata | (Glc β1→3 Glc)₅s  | NA         | 250,000   |
| Glucan          | Potato | (Glc α1→4 Glc)₅s   | Glc (α1→16) | NA        |
| Xylose          | Birchwood | (Xyl β1→4 Xyl)₅s  | Glr (α1→6) | NA        |

* NA, not available.
* Xyl, xylose.
* Glr, glucuronic acid.

for data analysis. ANOVAs are especially appropriate for loop designs in which a large number of conditions are compared with one another, eliminating uninteresting reference samples and allowing for the collection of more information on experimental conditions (14). Mixed ANOVAs are especially appropriate for loop designs in which a large number of conditions are compared with one another, eliminating uninteresting reference samples and allowing for the collection of more information on experimental conditions (14). Mixed ANOVAs are especially appropriate for loop designs in which a large number of conditions are compared with one another, eliminating uninteresting reference samples and allowing for the collection of more information on experimental conditions (14). Mixed ANOVAs are especially appropriate for loop designs in which a large number of conditions are compared with one another, eliminating uninteresting reference samples and allowing for the collection of more information on experimental conditions (14).

RESULTS AND DISCUSSION

**Design of Targeted cDNA Microarray**—A targeted cDNA microarray for T. maritima was constructed that included 269 known and putative genes or about 15% of the total open reading frames in the T. maritima genome. This included the known set of genes related to glycoside utilization and modification (65 genes), proteolysis (40 genes), stress response, and proteolytic fermentation. Genes related to sugar transport (21 genes) or transcriptional regulation (69 genes) and 66 other genes of interest were also included.

Genes apparently related to glycoside utilization and modification in T. maritima include 41 glycoside hydrolases, 17 glycosyl transferases, 6 carbohydrate esters, and 1 polysaccharide lyase. The corresponding encoded proteins have been classified into several families, based on amino acid sequence homology (15) (available on the World Wide Web at afmb.cnrsmrs.fr/CAZY). There are over 130 T. maritima proteins with sufficient BLAST homology to be classified into transcriptional regulatory or signal transduction COG categories (16). These regulatory proteins have been assigned to families based on sequence homology; however, different proteins in the same families may have different DNA and substrate-binding specificities (17). Also, proteins placed in different families may share the same name because of their regulon composition, as in the case of the Escherichia coli and Bacillus subtilis xylR protein (18, 19). Of the 69 transcription/transduction genes on the array, six share similarity with the
ROK (receptor, open reading frame, kinase) family of transcriptional regulators, which include glucokinases, *B. subtilis* XylR, and *E. coli* NagC (COG1940) (20). Six members of the PurR/LacI superfamily (COG1609) were included (21) along with the *T. maritima* IclR transcriptional regulator, whose structure was recently solved (22). Several pairs of sensor histidine kinases and response regulators of putative two-component regulatory systems were included, as were regulators from the MarR (23), AraC (24), TroR (25), LytR (26), ArsR (27), and CspC (28) families. The *T. maritima* genome contains ~120 genes involved in oligopeptide/sugar transport. In the targeted microarray used here, 21 genes related
to sugar transport were included on the basis of their proximity to the genes involved in glycoside utilization.

This targeted microarray was used to examine the differential response of *T. maritima* grown on a range of mono- and polysaccharides at its optimal growth temperature of 80 °C. Growth conditions were analyzed based on an incomplete loop design (Fig. 2). Treatments in the loop design were balanced with respect to dyes so that treatment effects were not confounded with dye effects.

**Growth of *T. maritima* on Monosaccharides and Polysaccharides**—*T. maritima* cultures were grown on a variety of saccharides, including the monosaccharides glucose, mannose, and xylose. The polysaccharides investigated differed in backbone sugar type (glucose, mannose, and xylose), backbone linkage type (β-1,3; β-1,4; or α-1,4), and side chain residue type (galactose, glucuronic acid, or glucose) (see Table I). Included in these were a mixed backbone (konjac glucomannan: glucose/mannose) and a mixed linkage (barley glucan: β-1,4/1,3) polysaccharide. Final cell densities were in the range of 10⁸ to 10⁹ cells/ml in all cases. Doubling times (min) for galactomannan (carob), β-glucan (barley), laminarin (*L. digitata*), β-xylan (birchwood), starch (potato), glucomannan (konjac), and carboxymethylcellulose were estimated to be 85, 72, 143, 61, 117, 74, and 78, respectively. On the monosaccharides, the doubling

---

**Fig. 4. Substrate-dependent regulation.** Sample Clusters constructed using standardized least squares means. Known or putative functions as reported in the genome sequence are indicated.
times (min) were 162, 253, and 188, for glucose, mannose, and xylose, respectively. Under identical conditions, the average doubling time for growth on monosaccharides (201 min) was observed to be substantially higher than that on the corresponding polysaccharide substrates (90 min).

Substrate-dependent Differential Expression of Genes—Two hierarchical clusters are shown in Fig. 3 to summarize the expression patterns of 269 *T. maritima* genes during growth on 10 saccharides. The first cluster is based on least squares means and compares the normalized expression levels of all

### Table 1: Substrate-dependent Differential Expression of Genes

| Locus       | Cluster 4.4 | Function                                                                 |
|-------------|-------------|--------------------------------------------------------------------------|
| TM1192      |             | alpha-galactosidase, Gal36                                               |
| TM1193      |             | beta-galactosidase, Gal42A                                              |
| TM1201      |             | arabinogalactan endo-1,4-beta-galactosidase, pat.                        |
| TM1195      |             | beta-galactosidase, Gal42B                                              |
| TM1200      |             | transcriptional regulator, Lel family                                    |
| TM1224      |             | transcriptional regulator, XylR-related                                 |
| TM1227      |             | endo-1,4-beta-mannosidase, Man5                                           |
| TM1274      |             | oligopeptide ABC transporter, permease protein                           |
| TM1275      |             | oligopeptide ABC transporter, permease protein                           |
| TM1279      |             | oligopeptide ABC transporter, ATP-binding protein                        |
| TM1280      |             | endoglucanase, Cel5A                                                    |
| TM1281      |             | endoglucanase, Cel6B                                                    |
| TM1624      |             | transcriptional regulator, RppR family                                   |
| TM1628      |             | oligopeptide ABC transporter                                             |
| TM1629      |             | beta-mannosidase, Man2                                                  |

### Table 2: Substrate-dependent Differential Expression of Genes

| Locus       | Cluster 4.5 | Function                                                                 |
|-------------|-------------|--------------------------------------------------------------------------|
| TM0055      |             | alpha-glucuronidase                                                     |
| TM0310      |             | beta-D-galactosidase, Gal2                                               |
| TM0062      |             | hypothetical protein                                                    |
| TM0070      |             | endo-1,4-beta-xylanase, Xyl10A                                           |
| TM0076      |             | hypothetical protein                                                    |
| TM0077      |             | acetyl xylan esterase                                                    |
| TM0078      |             | alpha-glucosidase, putative                                              |
| TM0132      |             | Xyl13                                                                    |
| TM0136      |             | xylosidase, Xyl2                                                        |
| TM0137      |             | arginine repressor                                                      |
| TM0139      |             | sugar ABC transporter, ATP-binding protein                               |
| TM0140      |             | L-lactate dehydrogenase                                                  |
| TM0141      |             | transcriptional regulator, crr family                                    |
| TM0142      |             | membrane bound protein LrR, putative                                    |
| TM0143      |             | nitrogen regulatory protein P-II                                        |
| TM0144      |             | heat shock response protein, periplasmal                                 |
| TM0145      |             | N-acetylglucosaminyl-1-phosphodiynositol                                 |
| TM0146      |             | transcriptional regulator, Lel family                                    |

### Table 3: Substrate-dependent Differential Expression of Genes

| Locus       | Cluster 4.6 | Function                                                                 |
|-------------|-------------|--------------------------------------------------------------------------|
| TM0053      |             | esterase, putative                                                       |
| TM0433      |             | exo-poly-alpha-D-galacturonanase, putative                               |
| TM0437      |             | transcriptional regulator, TefR family                                   |
| TM0438      |             | uronic isomerase, putative                                               |
| TM0439      |             | D-mannonoate hydrolase                                                  |
| TM0440      |             | transcriptional regulator, ICR family                                    |
| TM0441      |             | 2-deoxy-D-3-deoxy-d-glycero-6-P phosphate aldolase                       |
| TM0442      |             | 2-deoxy-3-deoxy-D-glycero-6-P kinase                                      |
| TM0443      |             | D-mannonoate oxidoreductase, putative                                    |
| TM0444      |             | sugar ABC transporter, permease protein                                  |
| TM0445      |             | gluconate kinase                                                         |
| TM0446      |             | conserved hypothetical protein                                           |
| TM0447      |             | oxidoreductase, short chain dehydrogen                                  |
| TM0448      |             | conserved hypothetical protein                                           |
| TM0449      |             | sugar ABC transporter, permease protein                                  |
| TM0450      |             | sugar ABC transporter, periplasmal sugar-binding                         |
| TM0451      |             | pecube lyase                                                             |

**Fig. 4—continued**
genes within each treatment condition. The second cluster is based on standardized least squares means for a single gene across all 10 treatments to show the effect of different treatments on the relative expression of a particular gene. The hierarchical clustering based on standardized least squares means revealed many cases of apparent co-regulation of genes within potential operons (29). Several sets of spatially distant gene strings were observed to cluster with similar expression profiles, suggesting the presence of regulons in the T. maritima genome. Representative clusters are displayed in Fig. 4. Overall expression levels of a number of genes remained consistently high or low regardless of the growth condition. These included constitutively expressed genes like TM0017 (pyruvate ferredoxin oxidoreductase) and TM0688 (glyceraldehyde-3-phosphate dehydrogenase) (30) as well as genes related to proteolytic activity. Both sets of genes with the corresponding known or putative functions are displayed in Fig. 5. Individual genes with high overall expression levels on only a single carbon source are indicated in Table II. Least squares means for all genes included in this study for all growth conditions are shown in Supplemental Table IV, along with the corresponding standardized values in Supplemental Table V. Below, gene regulation patterns within each functional category are examined for each monosaccharide and corresponding polysaccharide growth substrate.

Glucose and Glucan Polysaccharides—Backbone- and linkage-specific gene regulation was observed in the case of endoglycoside hydrolase genes for growth on α- and β-specific glucans. Growth on carboxymethylcellulose (CMC) (see cluster 4.1), a β-1,4-linked glucose polymer, induced genes encoding extracellular endoglucanases TM1525 (cel12B) and TM0305 (cel74), as well as the intracellular endoglucanase TM1524 (cel12A) and the intracellular cellobiosyl phosphorylase, TM1848. Examination of cluster I (Fig. 3) reveals that expression levels of cel74 were substantially lower than those of cel12A on glucan polysaccharides. Although the presence of a β-1,4-glucosidase gene (bglA) (accession number CAA52276) in T. maritima MSB8 has been reported (31), the corresponding protein sequence does not show homology to deduced sequences identified in the T. maritima MSB8 genome (4). Recently, bglA was reported to be present in T. maritima RQ2 (11).

A switch in sugar backbone linkage from β-1,4 to β-1,3 (cluster 4.2) resulted in specific expression of genes encoding a laminarinase, TM0024 (lam16), as well as the corresponding exoglucosidase, TM0025 (cel3), with comparable overall expression levels. The ROK family protein TM0032 is located upstream of lam16 and cel3 and displays a similar expression pattern. These expression patterns suggest specific regulatory networks for each glucan that differ only by backbone linkage. The mixed linkage β-glucan from barley induced genes encoding β-1,4- as well as β-1,3-glucanases in addition to a related glycosyl transferase, cellobiosyl phosphorylase; however, over-
all expression levels of lam16 were lower on this substrate than on laminarin.

Starch, an α-glucan, induced expression of genes encoding the extracellular endo-acting α-amylase TM1840 (amy13A), a debranching pullulanase, TM1845 (pull13), and the intracellular exo-acting α-glucosidase, TM1068 (amy4A) (cluster 4.3). Expression of the α-glucuronidase TM0434 (32), previously classified as a putative α-glucosidase, also increased on starch. The cyclomaltodextrinase TM1835 was significantly induced in the presence of starch. This intracellular protein has been recently shown to have exoglycosidase activity at the reducing end in addition to maltodextrin decycling activity (33). The genes encoding a 4-α-glucanotransferase (TM0364) and maltodextrin glycosyl transferase (TM0767) were also up-regulated during growth on starch, the latter also being expressed during growth on glucose and laminarin. Interestingly, other genes putatively assigned for starch degradation, including an α-amylase, TM1650 (amy13B), as well as the putative α-glucosidases, TM0752 and TM1834 (amy4AC), remained unaffected by this substrate, suggesting a different intracellular role perhaps related to utilization of storage polysaccharides like glycogen.

The relative expression of the AraC family regulator TM1005 increased specifically on starch. AraC proteins have been shown to act as transcriptional activators for a variety of substrates (24). A related MarR regulator TM0710 showed an increase in relative expression on all glucose-containing polymers, suggesting that transcriptional activators also play a role in T. maritima response to glucans. The highly expressed ROK family protein TM0393 showed its highest expression on laminarin and starch. Although expressed at a low level, a DeoR family regulator TM1069 showed its highest relative expression on starch, like its B. subtilis and E. coli homologs, which play a major role in the regulation of sugar metabolism genes (34).

All endoglycoside hydrolases were repressed during growth
on glucose. Although most endoglycoside hydrolases and certain exoglycosidases exhibited biochemically specific regulation in response to backbone and sugar linkage type, several other exoglycosidases displayed unexpected patterns. For instance, growth on CMC induced expression of genes encoding a fucosidase (TM0306), a \( \alpha \)-xilosidase (TM0308), three \( \beta \)-galactosidases (TM0310, TM1192, and TM1195), a \( \alpha \)-galactosidase, a \( \alpha \)-mannosidase (TM1851), and an \( \alpha \)-arabinofuranosidase (TM0281). Some of these genes (TM0306, TM0310, and TM1851) occur in the vicinity of genes responsible for the uptake and utilization of CMC (TM0305 and TM1848) and might be expected to co-regulated. In other cases, up-regulation of these genes presumably reflects the complex nature of polysaccharide structures present in the natural environs of T. maritima. In any case, it is apparent that annotations of some of these exo-acting glycosidase genes in the T. maritima genome need to be reevaluated.

Overall expression levels of genes encoding the glycosyl transferases TM0392, TM0744, and TM0895 remained high on glucose as well as on all \( \alpha \)- and \( \beta \)-glucan polysaccharides. Whereas TM0392 and TM0744 encode hypothetical proteins placed in family 4, TM0895 encodes a glycerogen synthase. Growth on glucose also resulted in high overall expression levels of the gene encoding a lipopolysaccharide biosynthesis protein, TM0631. Relative expression levels of genes encoding other glycosyl transferases TM0757 (hypothetical protein), TM1405 (lipopolysaccharide biosynthesis-related), and TM0756 (galactosyltransferase) remained high on glucose as compared with the \( \alpha \)- and \( \beta \)-glucan polysaccharides, despite their low overall expression levels on glucose. Growth on laminarin also resulted in high overall expression levels of lipopolysaccharide synthesis genes, TM0624 and TM0627.

Growth on CMC affected expression patterns of TM0300–TM0304 and TM1219–TM1223, two sets of genes annotated as oligopeptide ABC transporters. The former set of genes lies directly upstream of the gene encoding the extracellular endo-glucanase TM0305 and displayed low overall expression levels compared with the latter, which is present in a gene string encoding the extracellular \( \beta \)-mannanase, TM1227. The latter set was also overexpressed during growth on barley glucan. Whereas it could be hypothesized that TM0300–TM0304 are involved in the transport of \( \beta \)-1,4-linked glucose oligomers and TM1219–TM1223 are involved in the transport of mixed linkage (\( \beta \)-1,4/\( \beta \)-1,3) glucose oligomers, the latter set remained unaffected during growth on the \( \beta \)-1,3-linked glucose polymer, laminarin.

The motif discovery software AlignACE (35) was used to examine the upstream sequences of genes co-regulated on CMC and barley glucan, revealing a strongly conserved palindromic element upstream of TM1848, TM1223, TM0309, TM0310, and TM1524 with consensus NTGAAACATTCCATCGAATATT (see Table III). This motif shows strong similarity to the LacI family protein CcpA consensus CRE site within the Bacillus/Clostridium group of bacteria (36). Thus, the LacI family regulator TM1218, located downstream of the TM1219–TM1223 gene string, might be involved in local regulation of these ABC transporters and other members of the CMC degradation pathway proposed below. However, matches similar to this consensus are found upstream of genes and gene strings expressed on other sugars, raising the intriguing possibility that a global regulatory mechanism similar to CcpA-mediated carbon catabolite repression (CCR) in B. subtilis might be operating in T. maritima (37). Further work will assess this possibility.

Based on specific gene expression patterns observed during growth on CMC, the biochemical characteristics of TM1524 (Cel12A) (38), TM1525 (38), TM1848 (39), and TM0305 (40), the cellular localization of these proteins (5), and the motifs described, a mechanism for the uptake and utilization of CMC by T. maritima can be proposed (see Fig. 6A).

**Mannose and Mannan Polysaccharides—** Regulation patterns relating the monosaccharide mannose to mannan polysaccharides differed considerably from those observed for glucose and glucan polysaccharides. A strong correlation in
FIG. 6. Proposed pathways for polysaccharide uptake and utilization. A, carboxymethylcellulose uptake and utilization. TM1524, TM1525, and TM0305 encode for endo-β-1–4-glucanases. TM1219–TM1223 encode for ABC transporters. TM1848 encodes for a cellobiosyl phosphorylase. B, galactomannan uptake and utilization. TM1227 encodes for an endo-β-1,4-mannanase. TM1746–TM1750 encode for putative...
gene regulation was observed between cells grown on mannose and the complex polysaccharide galactomannan as well as the mixed polysaccharide glucomannan (cluster 4.4). Thus, unlike glucose, growth on mannose as well as mannann polysaccharides induced expression of the genes encoding the endoglycoside hydrolases, TM1227 (man5), TM1751 (cel5A), and TM1752 (cel5B). Whereas extracellular Man5 strictly degrades β-1,4 linkages between mannose residues, the latter two have been recently characterized as intracellular glucosaminanases (5). The mixed polymer glucomannan induced expression of all of these genes in addition to genes encoding the endoglucanases TM0805 (cel74) and TM1524 (cel12A) and the gene encoding the cellobioso phosphorylase TM1848. Growth on galactomannan also induced the putative α-amylase, TM1650 (amy13B) and an arabinoogalactan endogalactosidase, TM1201. The presence of galactose side chains in this substrate induced the expression of genes encoding the exo-acting debranching enzymes, TM1192 (gal36), TM1193 (gal42A), and TM1195 (gal42B). Whereas Gal36 hydrolyzes α-linked galactose residues attached to the mannann backbone, Gal42A-B cleave β-linked galactose end residues. These genes remained unaffected on the monosaccharide mannose as well as the mixed backbone polymer glucomannan, which lacks galactose side chain residues. The expression pattern of the Laci family protein TM1200 is similar to those of TM1191, TM1192, and TM1195, consistent with the location of TM1200 upstream of the other gene string (4). Thus, TM1200 may act as a local regulator of galactose genes in *T. maritima*.

The degradation of β-linked mannose oligosaccharides appears to be affected by the intracellular exo-acting glycosidase Man2 (TM1624), the gene that was induced on both mannann polysaccharides and on mannose. Overall expression levels of genes encoding a hypothetical glycosyl transferase (TM0392) and glycogen synthase (TM0895) were comparable with those observed for glucose and glucon polysaccharides. In contrast, the gene encoding the hypothetical glycosyl transferase TM0744 remained repressed during growth on mannose. Very high overall expression levels of pectate lyase (TM0433) were observed on glucomannan. Growth on mannose as well as on mannann polysaccharides resulted in high overall expression levels of the genes encoding the ABC transporters TM1746–TM1750, putatively annotated as oligopeptide transporters and located upstream of genes encoding the glucomannanases, TM1751 and TM1752. Overall expression levels of genes annotated as sugar ABC transporters (TM1232–TM1234) located downstream of *man5* (TM1227) remained low on all three substrates. Growth on glucomannan induced the transporter genes TM1220–TM1223, also co-regulated during growth on CMC and barley glucan. A proposed pathway for galactomannan utilization is shown in Fig. 6B, based on the specific gene expression patterns described above, as well as the biochemical properties and cellular localization of TM1227 (41), TM1624 (42), TM1192 (42), TM1751 (5), and TM1752 (5).

Among the transcriptional regulators that showed mannose-specific up-regulation were two ROK family proteins, TM1010 and TM0411 (cluster 4.7). Mlc, a ROK family member, is known to negatively regulate three phosphotransferase system (PTS) operons, including the mannose PTS *manXYZ* (43). However, PTS transporters have not been observed in *T. maritima*. The Laci family regulator TM1856 was also observed to have its highest expression on mannose, although its overall expression levels and those of TM0411 were low when compared with TM0110. TM0510, a putative manganese-dependent transcriptional regulator, and the response regulator TM0842 had high overall expression consistently across all conditions but showed their highest relative expression levels on mannose. Another response regulator, TM0126, displayed lower overall expression levels but also showed up-regulation on mannose. Whereas the transcriptional regulators described above were largely mannose-specific, the ROK family protein TM1224 and the ribose phosphate isomerase family protein TM1228 were expressed on mannose and mannann polysaccharides. Both TM1224 and TM1228 lie in the same gene string as the endoglycoside hydrolase gene, TM1227 (man5).

**Xylose and β-Xylan**—Similar to regulation patterns observed for mannose and mannann polysaccharides, growth on xylose and β-xylan revealed several sets of co-regulated genes (cluster 4.5). Genes encoding the extracellular endo-acting xylanases TM0061 (xyl10A) and TM0070 (xyl10B) and the intracellular exo-acting xyllosidase TM0076 (xyl13) displayed comparable expression levels on both substrates and were also up-regulated on the monosaccharide mannose. Very high overall expression levels of the gene encoding the α-glucuronidase (45) TM0055 were observed exclusively during growth on β-xylan, whereas overall expression levels of the gene encoding the acetyl xylan esterase TM0077 were similar across several polysaccharides (CMC, β-1,3/4 glucan, starch) and monosaccharides (mannose and xylose). Relative expression levels of the family 4 α-glucuronidase gene TM0434 (32) were high on both xylose and β-xylan, whereas its overall expression levels were lower than those of other related genes. As mentioned above, TM0434 was previously classified as a putative α-glucosidase and showed high relative expression levels on starch. The other acetyl xylan esterase, TM0435, had high relative expression levels on xylose, β-xylan, and glucomannan, although its overall expression levels were low on these substrates. Other genes encoding exo-acting glycosidases that were expressed on both xylose and β-xylan included a putative α-glucosidase (TM0752) and a α-mannosidase (TM1231). Whereas TM0300–TM0303 were highly expressed on xylose, TM0430–TM0432 were expressed on β-xylan. The proposed pathway for the utilization of β-xylan is shown in Fig. 6C.

Genes up-regulated on xylan and konjac glucomannan are located in at least two main genomic regions (cluster 4.6). The region located between TM0055 and TM0077 revealed some striking compositional similarities to the glucuronic acid utilization cluster of *Bacillus stearothermophilus* 7-6 (46), consistent with the identification of side chain substitutions in the form of acetyl, arabinosyl, ferulic acid and 4-O-methyl-α-gluco- ronic acid in the β-xylan backbone (46). Common features include the presence of an IclR/KdgR family regulatory protein upstream of the glucuronic acid metabolism genes *kgdAK* and *uxaBA*, and divergently transcribed from *uxaC*, and the co-localization of extracellular and intracellular xylanases within the cluster (4, 46). The second genomic region between TM0430 and TM0439 contains gene strings sharing sequence homology with genes involved in the related process of pectin degradation.** The ABC transporter subunits TM0430–TM0432 are located upstream of TM0433 (pectate lyase), whereas TM0434–TM0443 lie downstream within a gene string whose members are all separated by 20 bases or less. TM0439 is annotated as a GntR family regulator and is 35% identical and 51% similar to the gluconate repressor GntR of *B. subtilis* (47).

Other putative regulators showed expression patterns sug-
gesting involvement in the regulation of xylose and xylan-related genes. TM0049 displays 30% identity with E. coli XyIR, the positive regulator of the xylFGH ABC transporter operon, and showed a particularly strong induction on xylose. ABC transporters located downstream of TM0049 show similarity to xylFGH, and it remains to be seen whether these genes, annotated as ribose ABC transporters, are induced during growth on xylose as observed in E. coli (18, 48). The CRP-like transcriptional regulator TM1171 was up-regulated primarily on xylose but to a lesser extent on CMC. Experimental evidence in Thermotoga neapolitana suggests that CRP in this species is not dependent on cAMP (49). If this is also the case in T. maritima, the function of the CRP protein may differ from that of E. coli CRP, whose role in global catabolite repression has been well documented (50). TM0080, the single ROK family protein that showed increased expression on xylan and xylose, is located upstream of a putative hydrolase (TM0089), which showed expression preferences for xylan and starch.

Summary—The genome sequence of T. maritima encodes for the highest number of known glycoside hydrolase genes among hyperthermophilic bacterial and archaeal genome sequences reported to date (Fig. 1) (51). The data set resulting from this study provides a rich resource for characterizing putative regulatory mechanisms in T. maritima that might have been proposed from genome sequence data (4). Previous studies on T. neapolitana have reported the classical lactose-glucose-diauxie, operating independently of cAMP (49). In the case of β-galactoside transport and metabolism in T. neapolitana, transport system-based regulation through inducer exclusion or expulsion has been ruled out (52). The results of this study suggest sugar-specific regulation patterns for T. maritima Glycoside Hydrolases systems that operating in Thermotoga neapolitana raises even more questions about available mechanisms for CRP in T. maritima.

Further studies using full genome microarrays and biochemical studies to unravel the pathways for intracellular metabolism in this organism with presumably primitive traits are currently under way.

Acknowledgments—We acknowledge Dr. Bryon Sosinski and Dr. Len van Zyl for helpful discussions.

REFERENCES
1. de Vos, W. M., Kengen, S. W. M., Voorhorst, W. G. B., and van der Oost, J. (1998) Extremophiles 2, 201–205
2. Galperin, M. Y., Noll, K. M., and Romano, A. H. (1996) Appl. Environ. Microbiol. 62, 2915–2918
3. Huber, L., Langworthy, T. A., Konig, H., Thomm, M., Weise, C. R., Sleytr, U. B., and Stetter, K. O. (1986) Arch. Microbiol. 144, 324–333
4. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Jia, D. N., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., Fraser, C. M., et al. (1999) Nature 399, 323–329
5. Chhabra, S. R., Shockley, K. R., Ward, D. E., and Kelly, R. M. (2002) Appl. Environ. Microbiol. 68, 555–561
6. Bronnenmeier, K., Kern, A., Liebl, W., and Staudenbauer, W. L. (1995) Appl. Environ. Microbiol. 61, 1399–1407
7. Nguyen, T. N., Borges, K. M., Romano, A. H., and Noll, K. M. (2001) FEMS Microbiol. Lett. 195, 79–83
8. Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001) Nat. Genet. 29, 389–395
9. Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C., and Paules, R. S. (2001) J. Comput. Biol. 8, 625–637
10. Achenbach-Richter, L., Gupta, R., Stetter, K. O., and Weise, C. R. (1987) Syst. Appl. Microbiol. 9, 34–39
11. Nesbo, C. L., Nelson, K. E., and Dokulil, W. F. (2002) J. Bacteriol. 184, 4475–4488
12. Asai, S. F., Gich, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
13. Hasseman, J. (2001) TIGR Microarray Protocols, http://www.tigr.org/dbh/microarray/protocolsTIGR.shtml
14. Kerr, M. K., and Churchill, G. A. (2001) Genet. Res. 77, 123–128
15. Henryssat, B., and Bairoch, A. (1996) Biochem. J. 316, 695–696
16. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, R. S., Kyrysiuc, R., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V. (2001) Nucl. Acids Res. 29, 22–28
17. Mirny, L. A., and Gelfand, M. S. (2002) J. Mol. Biol. 321, 7–20
18. Scharf, S., and Park, C. (2000) J. Bacteriol. 182, 785–825
19. Kreuzer, P., Gartner, D., Almannersberg, R., and Hillen, W. (1989) J. Bacteriol. 171, 3840–3845
20. Tiggesmeier, F. (1993) J. Coll. Biochem. 51, 69–74
21. Mirny, L. A., and Gelfand, M. S. (2002) Nucl. Acids Res. 30, 1704–1711
22. Zhang, R. G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimak, A., and Savchenkov, A. (2002) J. Biol. Chem. 277, 19183–19190
23. Cohen, S. P., Hachler, H., and Levy, S. B. (1993) J. Bacteriol. 175, 1484–1492
24. Martin, R. G., and Rosner, J. L. (2001) Curr. Opin. Microbiol. 4, 132–137
25. Hardham, J. M., Stamm, L. V., Porcella, S. F., Frye, J. G., Barnes, N. Y., Howell, J. K., Mueller, S. L., Radolf, J. D., Weinstock, G. M., and Norris, S. J. (1997) Gene (Amst.) 197, 47–64
26. Nikolaevskaya, A. N., and Galperin, M. Y. (2002) Nucl. Acids Res. 30, 2453–2459
27. Diore, C., Cai, J., Marmur, J., Shinder, R., and DuBrow, M. S. (1995) J. Bacteriol. 177, 2050–2053
28. Phadtare, S., Alsina, J., and Inouye, M. (1999) Curr. Opin. Microbiol. 2, 175–180
29. McGuire, A. M., Hughes, D. J., and Church, G. M. (2000) Genome Res. 10, 744–757
30. Blamey, J. M., and Adams, M. W. (1994) Biochemistry 33, 1000–1007
31. Liebl, W. (2001) Methods Enzymol. 330, 290–309
32. Suresh, C., Rus’d, A. A., Kotaoka, M., and Hayashi, K. (2002) FEBS Lett. 517, 159–162
33. Lee, M. H., Kim, Y. W., Kim, T. J., Park, C. S., Kim, J. M., Moon, T. W., and Song, S. J. (2000) Biochim. Biophys. Res. Commun. 295, 818–825
34. Xing, S., Saxild, H. H., and Switzer, B. L. (2000) J. Bacteriol. 182, 1916–1922
35. Hughes, J. D., Estep, P. W., Tavazoei, S., and Church, G. M. (2000) J. Mol. Biol. 296, 1205–1214
36. Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2001) FEMS Microbiol. Lett. 205, 305–314
37. Bruckner, R., and Tiggesmeier, F. (2002) FEMS Microbiol. Lett. 209, 141–148
38. Liebl, W., Ruile, P., Bronnenmeier, K., Riedel, K., Lettscheg, F., and Greif, I. (1996) Microbiology (Reading) 142, 2533–2542
39. Verron, D. A., McCarthy, J. K., Evelidge, D. E., and Bok, J. D. (2000) J. 3

3 S. B. Conners and R. M. Kelly, unpublished data.
Regulation of T. maritima Glycoside Hydrolases

40. Chhabra, S. R., and Kelly, R. M. (2002) FEBS Lett. 531, 375–380
41. Chhabra, S., Parker, K. N., Lam, D., Callen, W., Sneed, M. A., Mathur, E. J., Short, J. M., and Kelly, R. M. (2001) Methods Enzymol. 330, 224–238
42. Parker, K. N., Chhabra, S. R., Lam, D., Callen, W., Duffaud, G. B., Sneed, M. A., Short, J. M., Mathur, E. J., and Kelly, R. M. (2001) Biotechnol. Bioeng. 75, 322–333
43. Plumbridge, J. (2001) J. Mol. Microbiol. Biotechnol. 3, 371–380
44. Sorensen, K. I., and Hove-Jensen, B. (1996) J. Bacteriol. 178, 1003–1011
45. Ruile, P., Winterhalter, C., and Liebl, W. (1997) Mol. Microbiol. 23, 267–279
46. Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999) J. Bacteriol. 181, 3695–3704
47. Miwa, Y., and Fujita, Y. (1988) J. Biol. Chem. 263, 13252–13257
48. Song, S., and Park, C. (1998) FEMS Microbiol. Lett. 163, 255–261
49. Vargas, M., and Null, K. M. (1996) Microbiology 142, 139–144
50. Bushy, S., and Ehrlich, R. H. (1999) J. Mol. Biol. 283, 199–213
51. Henrissat, B., and Coutinho, P. M. (2001) Methods Enzymol. 330, 183–201
52. Galperin, M. Y., Null, K. M., and Romano, A. H. (1997) Appl. Environ. Microbiol. 63, 969–972
53. Fujita, Y., Miwa, Y., Galmier, A., and Deutscher, J. (1995) Mol. Microbiol. 17, 953–960