The D-XYlose-Binding Protein, XylF, from Thermoanaerobacter ethanolicus 39E: Cloning, Molecular Analysis, and Expression of the Structural Gene†

MILUTIN ERBEZNIK,‡ HERBERT J. STROBEL,* KARL A. DAWSON, and CHRIS R. JONES

Department of Animal Sciences, University of Kentucky, Lexington, Kentucky 40546-0215

Received 3 February 1998/Accepted 15 May 1998

D-XYlose is nature’s second-most-abundant carbohydrate. The initial steps of D-xylose metabolism in bacteria typically involve transport, isomerization to D-xylulose, and phosphorylation to D-xylulose-5-phosphate. The second and third steps of this pathway are catalyzed by D-xylose isomerase and D-xylulose kinase, lies a 1,101-bp open reading frame that exhibits 61% amino acid sequence identity to the Escherichia coli D-xylose binding periplasmic receptor, XylF, a component of the high-affinity binding-protein-dependent D-xylose transport. The 25-residue N-terminal fragment of the deduced T. ethanolicus XylF has typical features of bacterial leader peptides. The C-terminal portion of this leader sequence matches the cleavage consensus for lipoproteins and is followed by a 22-residue putative linker sequence rich in serine, threonine, and asparagine. The putative mature 341-amino-acid-residue XylF (calculated molecular mass of 37,069 Da) appears to be a lipoprotein attached to the cell membrane via a lipid anchor covalently linked to the N-terminal cysteine, as demonstrated by metabolic labelling of the recombinant XylF with [35S]methionine. The induced E. coli avidly bound D-[14C]xylose, yielding additional evidence that T. ethanolicus XylF is the D-xylose-binding protein. On the basis of sequence comparison of XylFs to other monosaccharide-binding proteins, we propose that the sequence signature of binding proteins specific for hexoses and pentoses be refined as (KDQ)(LIVFAG)2X(DN)(SGP)X3(GS)X(LIVA)2X.A. Transcription of the monocistronic 1.3-kb xylF mRNA is inducible by xylose and unaffected by glucose. Primer extension analysis indicated that xylF transcription initiates from two +1 sites, both situated within the xylAB operon. Unlike in similar transport systems in other bacteria, the genes specifying the membrane components (e.g., ATP-binding protein and permease) of the high-affinity D-xylose uptake system are not located in the vicinity of xylF in T. ethanolicus. This is the first report of a gene encoding a xylose-binding protein in a gram-positive or thermophilic bacterium.

Materials and Methods

Bacterial strains, plasmids, and growth conditions. T. ethanolicus 39E (ATCC 33223) was grown as previously described (9). Plasmids pUC18 (Life Technologies, Gaithersburg, Md.) and pBluescript II KS (Stratagene, La Jolla, Calif.) were used as cloning vectors in Escherichia coli DH5α (Life Technologies, Gaithersburg, Md.). The E. coli strains were grown in Luria-Bertani medium (39) supplemented with ampicillin (100 µg/ml) or carbenicillin (50 µg/ml) as needed.

Construction and screening of genomic library. A T. ethanolicus genomic library was prepared as previously described (9). Probes for chromosome walking were prepared by random hexamer labelling of the appropriate restriction fragments with [α-32P]dATP (39). Colony hybridization and washes, as well as recombinant DNA manipulations, were carried out by standard techniques (39).
DNA sequencing and analysis. Dideoxy-termination DNA sequencing was performed by following the Sequenase 2.0 kit (Amersham) protocol with double-stranded plasmid templates isolated by a boiling minipreparation procedure (39). Oligonucleotides were synthesized by Integrated DNA Technologies (Coraville, Iowa). PCR products were sequenced on an automated Sequenase 2.0 DNA sequencer (version 2.0.1) using an applied biosystems software package (DNASTAR, Madison, Wis.). BLAST search engines were employed for sequence homology searches (2).

PCR. Amplifications were performed as described previously (9), with the following modifications: Pf28 DNA (5 ng per reaction) was used as a template, the total number of cycles was 30, and the annealing temperature for the xylF-specific PCR product was 59°C. The oligonucleotides used for PCR of the xylF fragment were 5′AAT GTA GAT TTA AAA ATT GTA GGA AGG AG and 5′GAT TCG CTG GCC GAT TGA CAT AAC. The DNA template for the antisense riboprobe was constructed by subcloning the 3′-terminal SpeI-EcoRI fragment from pXI-10 (Fig. 1 in reference 8; the EcoRI site was from the vector) into pBluescript II KS. The SpI-linearized template was transcribed in vitro with T3 RNA polymerase in the presence of [α-32P]CTP (39), yielding a 608-nucleotide antisense RNA.

Cloning and expression of the recombinant XylF. The xylF open reading frame (ORF) was PCR amplified by using oligonucleotides 5′ AGG AGA ATA GAC CAT GTG TAA AAA GAC CGC TGT GTG ATG GTC CTT C, which is complementary to the 3′-terminal nucleotides downstream from the xylF translation initiation site. For dot blot analysis, 1.5-μl aliquots of various dilutions of total RNA isolated from T. ethanolicus 39E cells grown in the presence of different carbon sources were spotted onto a membrane which was hybridized to the same radiolabelled xylF-specific probe used for Northern analysis. A standard hybridization procedure was used (39).

Cloning and expression of the recombinant XylF. The xylF open reading frame (ORF) was PCR amplified by using oligonucleotides 5′ AGG AGA ATA GAC CAT GTG TAA AAA GAC CGC TGT GTG ATG GTC CTT C, which is complementary to the 3′-terminal nucleotides downstream from the xylF translation initiation site. For dot blot analysis, 1.5-μl aliquots of various dilutions of total RNA isolated from T. ethanolicus 39E cells grown in the presence of different carbon sources were spotted onto a membrane which was hybridized to the same radiolabelled xylF-specific probe used for Northern analysis. A standard hybridization procedure was used (39).

Cloning and expression of the recombinant XylF. The xylF open reading frame (ORF) was PCR amplified by using oligonucleotides 5′ AGG AGA ATA GAC CAT GTG TAA AAA GAC CGC TGT GTG ATG GTC CTT C, which is complementary to the 3′-terminal nucleotides downstream from the xylF translation initiation site. For dot blot analysis, 1.5-μl aliquots of various dilutions of total RNA isolated from T. ethanolicus 39E cells grown in the presence of different carbon sources were spotted onto a membrane which was hybridized to the same radiolabelled xylF-specific probe used for Northern analysis. A standard hybridization procedure was used (39).

Cloning and expression of the recombinant XylF. The xylF open reading frame (ORF) was PCR amplified by using oligonucleotides 5′ AGG AGA ATA GAC CAT GTG TAA AAA GAC CGC TGT GTG ATG GTC CTT C, which is complementary to the 3′-terminal nucleotides downstream from the xylF translation initiation site. For dot blot analysis, 1.5-μl aliquots of various dilutions of total RNA isolated from T. ethanolicus 39E cells grown in the presence of different carbon sources were spotted onto a membrane which was hybridized to the same radiolabelled xylF-specific probe used for Northern analysis. A standard hybridization procedure was used (39).

Cloning and expression of the recombinant XylF. The xylF open reading frame (ORF) was PCR amplified by using oligonucleotides 5′ AGG AGA ATA GAC CAT GTG TAA AAA GAC CGC TGT GTG ATG GTC CTT C, which is complementary to the 3′-terminal nucleotides downstream from the xylF translation initiation site. For dot blot analysis, 1.5-μl aliquots of various dilutions of total RNA isolated from T. ethanolicus 39E cells grown in the presence of different carbon sources were spotted onto a membrane which was hybridized to the same radiolabelled xylF-specific probe used for Northern analysis. A standard hybridization procedure was used (39).

Amino acid sequence analysis. The 25-residue amino-terminal domain of the T. ethanolicus XylF protein has three distinct domains: i.e., a positively charged N-terminal region, followed by a highly hydrophobic central region which ends with a stretch of short side chain neutral amino acids (data not shown). Such a sequence structure is typical of signal peptides in bacterial lipoproteins which are cleaved off at the cysteine residue (52), thus rendering Cys-26 the N terminus of the mature T. ethanolicus XylF protein, with a calculated molecular mass of 37,069 Da. The sequence LSGC, corresponding to residues 23 to 26 in T. ethanolicus XylF (data not shown), perfectly matches the four-residue sequence consensus L(S,A)(Y,A) C defined for lipoprotein leader sequences (49, 52). Taken together, these observations strongly suggested that the processed, mature form of T. ethanolicus XylF is a lipoprotein.

The putative signal peptide is followed by a 22-residue stretch, rich in polar amino acids, e.g., serine, threonine, and asparagine (7, 5, and 4 residues, respectively [Fig. 1]). To compare the T. ethanolicus XylF (Fig. 1) sequence to
those of other monosaccharide-binding proteins, their primary structures were analyzed by multiple sequence alignment. Saliently, the (S,T,N)-rich region following the amino-terminal cysteine of the processed \( T. \) ethanolicus \( XylF \) was absent in other proteins. Two more cysteine residues were found in the \( T. \) ethanolicus \( XylF \) at positions 76 and 253, unlike the mesophilic D-xylose-binding proteins which lack cysteine. The \( T. \) ethanolicus \( XylF \) sequence displayed somewhat higher homology (34% sequence identity, 52% similarity) to the multiple sugar-binding periplasmic receptor ChvE from \( A. \) tumefaciens (44), than to the \( E. \) coli \( ABC \)-transport periplasmic receptors of L-arabinose (AraF [19]), D-glucose or D-galactose (MglB [20]), and D-ribose (RbsB [16]), with 27 to 30% sequence identity and 40 to 48% similarity.

Scrutiny of the aligned sequences revealed that 4 of 14 amino acid residues implicated in binding of sugar substrates by \( E. \) coli AraF, MglB, and RbsB (53) were fully conserved in \( XylFs \) and ChvE (Fig. 1). Two of them are aromatic residues (Trp-16 in AraF and Trp-183 in MglB), which form stacking interactions with sugars and correspond to Trp-42 and Trp-194 in \( T. \) ethanolicus \( XylF \). The other two are planar polar residues creating hydrogen bonds with sugar substrates (Asp-89 in AraF, and Asp-236 in MglB), and they align with Asp-115 and Asp-247, respectively, in \( T. \) ethanolicus \( XylF \). The other 10 residues implicated in sugar binding correspond to residues that are (i) invariant in \( XylFs \) but not found in ChvE (Glu-14 and Asn-205 in AraF, matching Glu-40 and Asn-221, respectively, in \( XylF \)-Te), (ii) replaced by isofunctional (i.e., planar polar or charged) residues conserved in both \( XylFs \) and ChvE (Asp-90 in AraF and Asp-154 in MglB corresponding to Arg-116 and Asn-162, respectively, in \( T. \) ethanolicus \( XylF \)) or in \( XylFs \) only (Lys-10 in AraF and His-152 in MglB aligning with Asp-35 and Asp-160, respectively, in \( T. \) ethanolicus \( XylF \)), and (iii) not conserved (Met-108, Arg-151, Met-204, and Pro-254 in AraF).

FIG. 1. Amino acid sequence alignment of monosaccharide-binding proteins. Sequences of the mature monosaccharide-binding proteins (i.e., without their signal peptides) were aligned by the CLUSTAL method with the PAM250 residue weight table (18). Residues conserved in two or more species are boxed and shaded. Solid arrowheads indicate residues invariant in \( XylFs \) that are identical to the residues participating in protein-sugar interactions in AraF and MglB (Table 2 in reference 53). Open arrowheads denote planar polar residues invariant in \( XylFs \), which align with isofunctional residues implicated in H-bond formation in AraF and MglB. Open circles indicate internal cysteine residues in the mature \( T. \) ethanolicus \( XylF \). The sequence region containing the consensus signature for the “cluster 2” binding proteins (50) is bracketed. Te, \( T. \) ethanolicus; Ec, \( E. \) coli; Hi, \( H. \) influenzae; At, \( A. \) tumefaciens.
binding proteins (50) was also identified in D-xylode receptors (Fig. 1). Except for Glu-85 in T. ethanolicus XylF and Asn-68 in mesophilic XylFs, all other residues of this sequence region perfectly matched the previously identified consensus. Additional analysis revealed that this sequence is part of a larger domain that is shared by a diverse family of proteins (51), including transcriptional regulators such as FruR and LacI in E. coli and CcpA, involved in catabolite repression regulation in Bacillus subtilis. This domain corresponds to the region comprising residues 23 to 115 in T. ethanolicus XylF and 29 to 113 in mesophilic XylFs.

Characterization of transcription. In order to perform Northern analysis, equal amounts of RNA from cells cultivated with xylose or glucose were hybridized against a 323-bp fragment from xylF. A prominent band, approximately 1.3 kb in size, was seen only in RNA from xylose-cultured cells (Fig. 2A). Two other faint bands (approximately 5 and 4 kb long) were detected when xylose was the energy source. DNA sequence analysis (reference 8 and this study) showed that an inverted repeat \( \Delta G = -10.5 \text{kcal/mol} \), followed by a poly(dT) stretch, is located in the xylB-xylF intercistronic spacer (Fig. 3B). Northern analysis suggested that this sequence is likely to function as a p-independent termination hairpin of xylAB transcription; therefore, the majority of xylF mRNA appeared to be monocistronic.

This observation was also supported by results of an RNase protection assay with total RNA from xylose-grown T. ethanolicus and a riboprobe overlapping the 3' and 5' termini of xylB and xylF, respectively. Two strong signals, corresponding in size to fragments expected from protection of xylAB and xylF mRNAs, were obtained (Fig. 2B). Also, a rather faint band, similar in length to the riboprobe, could be seen in the same lane on the original autoradiograph, which represented the protected fragment of the xylABF read-through transcript. The additional band, approximately 140 nucleotides long, that was obtained in the reaction was probably a nonspecific product of RNase cleavage, due to the AU-rich sequence. Since the RNase protection assay has higher sensitivity relative to that of Northern blotting, these observations provide evidence that the majority of xylF transcripts initiate from the xylF promoter, proximal to the xylF coding sequence.

To precisely map the 5' terminus of the xylF transcript, primer extension analysis was performed by using an oligonucleotide complementary to the sequence region downstream of the xylF ORF 5' end. Two major cDNA products, the longer being somewhat more prominent than the shorter (Fig. 3A),
were identified in RNA samples from *T. ethanolicus* grown with xylose. These extension products signified two possible *xylF* transcription initiation sites situated within the *xylAB* operon (Fig. 3B). The multiple initiation sites of the longer transcript (where the major 5′ terminus is marked with S1) lie within the *xylB* coding region, while the downstream site, S2, is located 17 bp downstream of the *xylB* termination codon. Both S1 and S2 sites are preceded by properly spaced sequences (5′ TTA CTA-17 bp-AATTGT 3′ and 5′ TGTTGA-16 bp-TGG TAT 3′ for S1 and S2, respectively) resembling the promoter consensus in gram-positive bacteria (15). Therefore, if one assumes that the shorter cDNA product was not the result of a premature termination of primer extension by the reverse transcriptase, it seems likely that the *xylF* transcription could initiate at either site. Alternatively, the transcript initiated at S2 may be a product of RNase processing, rather than of transcription initiation. The fact that the 5′ end of this transcript is adjacent to the palindromic sequence (Fig. 3B) supports this interpretation.

To ascertain the effect of sugar substrate on *xylF* expression, dot blot analysis was carried out with total RNA samples isolated from *T. ethanolicus* 39E grown with various carbon sources (Fig. 4). Similarly to the Northern blotting result, only a faint signal was observed in the absence of xylose. Interestingly, the presence of glucose did not have a repression effect on the xylose-induced *xylF* mRNA synthesis.

**Expression of the recombinant XylF.** SDS-PAGE was carried out to analyze the expression of the *T. ethanolicus* XylF in *E. coli* (Fig. 5). A highly abundant protein band migrating at approximately 37 kDa was observed in cell lysates of induced cells, but not in uninduced cells or cells carrying only the expression vector (pET15b) or the *T. ethanolicus* *xylF* construct (pXFe) were harvested 90 min postinduction with IPTG, and aliquots of cell lysates were fractionated by SDS-PAGE (approximately 10 μg of protein per lane), S, molecular mass protein standards; – and +, absence and presence, respectively, of IPTG in the culture.

**FIG. 4.** RNA dot blot analysis. *T. ethanolicus* 39E was cultivated with various sugar substrates, and total RNA was purified from logarithmically grown cells. Equal aliquots of serial RNA dilutions were applied to a membrane and hybridized to [α-32P]dATP-labelled 323-bp XylF-specific PCR product, which had also served as the probe in Northern analysis. X, 0.4% xylose; G, 0.4% glucose, X + G, xylose and glucose, each at 0.2%.

**FIG. 5.** Expression of the recombinant XylF (rXylF). *E. coli* cells carrying the expression vector (pET15b) or the *T. ethanolicus* xylF construct (pXFe) were harvested 90 min postinduction with IPTG, and aliquots of cell lysates were fractionated by SDS-PAGE (approximately 10 μg of protein per lane), S, molecular mass protein standards; – and +, absence and presence, respectively, of IPTG in the culture.

**DISCUSSION**

To our knowledge, the molecular analysis of the *T. ethanolicus* *xylF* presented here is the first report of a gene encoding D-xylose-binding protein in a gram-positive bacterium and in a thermophilic organism. We have provided evidence strongly suggesting that the *T. ethanolicus* XylF is a lipoprotein: (i) the sequence begins with a putative signal peptide that contains the lipoprotein-scission consensus site (52); (ii) metabolic labelling with [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

Metabolic radiolabelling of the recombinant XylF. The amino acid sequence analysis suggested that the native *T. ethanolicus* XylF is a lipoprotein. Biochemical evidence to support this hypothesis could be provided by covalently linking the recombinant XylF to a radiolabelled fatty acid. A similar approach has been used to identify lipoproteins in other gram-positive bacteria (49). In an attempt to accomplish this, the host cells harboring the recombinant clone were grown in the presence of [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

**FIG. 4.** RNA dot blot analysis. *T. ethanolicus* 39E was cultivated with various sugar substrates, and total RNA was purified from logarithmically grown cells. Equal aliquots of serial RNA dilutions were applied to a membrane and hybridized to [α-32P]dATP-labelled 323-bp XylF-specific PCR product, which had also served as the probe in Northern analysis. X, 0.4% xylose; G, 0.4% glucose, X + G, xylose and glucose, each at 0.2%.

**FIG. 5.** Expression of the recombinant XylF (rXylF). *E. coli* cells carrying the expression vector (pET15b) or the *T. ethanolicus* xylF construct (pXFe) were harvested 90 min postinduction with IPTG, and aliquots of cell lysates were fractionated by SDS-PAGE (approximately 10 μg of protein per lane), S, molecular mass protein standards; – and +, absence and presence, respectively, of IPTG in the culture.

**DISCUSSION**

To our knowledge, the molecular analysis of the *T. ethanolicus* *xylF* presented here is the first report of a gene encoding D-xylose-binding protein in a gram-positive bacterium and in a thermophilic organism. We have provided evidence strongly suggesting that the *T. ethanolicus* XylF is a lipoprotein: (i) the sequence begins with a putative signal peptide that contains the lipoprotein-scission consensus site (52); (ii) metabolic labelling with [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

Metabolic radiolabelling of the recombinant XylF. The amino acid sequence analysis suggested that the native *T. ethanolicus* XylF is a lipoprotein. Biochemical evidence to support this hypothesis could be provided by covalently linking the recombinant XylF to a radiolabelled fatty acid. A similar approach has been used to identify lipoproteins in other gram-positive bacteria (49). In an attempt to accomplish this, the host cells harboring the recombinant clone were grown in the presence of [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

**DISCUSSION**

To our knowledge, the molecular analysis of the *T. ethanolicus* *xylF* presented here is the first report of a gene encoding D-xylose-binding protein in a gram-positive bacterium and in a thermophilic organism. We have provided evidence strongly suggesting that the *T. ethanolicus* XylF is a lipoprotein: (i) the sequence begins with a putative signal peptide that contains the lipoprotein-scission consensus site (52); (ii) metabolic labelling with [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

Metabolic radiolabelling of the recombinant XylF. The amino acid sequence analysis suggested that the native *T. ethanolicus* XylF is a lipoprotein. Biochemical evidence to support this hypothesis could be provided by covalently linking the recombinant XylF to a radiolabelled fatty acid. A similar approach has been used to identify lipoproteins in other gram-positive bacteria (49). In an attempt to accomplish this, the host cells harboring the recombinant clone were grown in the presence of [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

**DISCUSSION**

To our knowledge, the molecular analysis of the *T. ethanolicus* *xylF* presented here is the first report of a gene encoding D-xylose-binding protein in a gram-positive bacterium and in a thermophilic organism. We have provided evidence strongly suggesting that the *T. ethanolicus* XylF is a lipoprotein: (i) the sequence begins with a putative signal peptide that contains the lipoprotein-scission consensus site (52); (ii) metabolic labelling with [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.

Metabolic radiolabelling of the recombinant XylF. The amino acid sequence analysis suggested that the native *T. ethanolicus* XylF is a lipoprotein. Biochemical evidence to support this hypothesis could be provided by covalently linking the recombinant XylF to a radiolabelled fatty acid. A similar approach has been used to identify lipoproteins in other gram-positive bacteria (49). In an attempt to accomplish this, the host cells harboring the recombinant clone were grown in the presence of [1-14C]palmitic acid prior to induction. Cell lysates of radiolabelled cells were then subjected to SDS-PAGE, followed by autoradiography (Fig. 6). The result clearly showed the presence of a radioactively labelled protein corresponding in size to the recombinant XylF, and this protein was absent in the control (uninduced) cells. Thus, *E. coli* performed posttranslational fatty acyl modification of the recombinant XylF.
enzae, D-xylose transport genes also constitute an operon, H. influ-
with ATP-binding protein. In xylAB, these proteins are clustered within the operon. The question as to how the expression of these remote cistrons is coordinated. (38) are not situated within 1 site (8) and was postulated to serve as a T. ethanolicus xylA promoter. However, without experimental evidence, it is difficult to explain how an XylR homolog in T. ethanolicus would achieve simultaneous binding of the xylF palindrome spaced by the 23-bp loop. Future studies should show whether a looping mechanism, analogous to that observed in t-arabinose operon regulation by AraC (40), may be involved in regulation of xylF expression in T. ethanolicus. Nevertheless, the presence of conserved palindromes adjacent to both xylAB and xylF promoters and similar expression patterns of the two operons with respect to the carbon sources suggest that both xylAB and xylF operons may be controlled by a similar mechanism of negative regulation. Interestingly, the inverted repeat implicated in the xylAB transcription termination coincides with the putative xylF operator sequence. Hence, this might serve as an example of versatility of cis-acting elements in transcriptional regulation of bacterial carbon metabolism.

A recent study showed that XylFs from E. coli and H. influ-
enzae are more closely related to the A. tumefaciens multiple-sugar-binding protein ChvE than to the E. coli periplasmic receptors of t-arabinose, D-glucose or D-galactose, or D-ribose (22). Our inclusion of T. ethanolicus XylF in the multiple sequence comparison confirmed this observation. Significantly, our sequence analysis indicated that most residues involved in protein-sugar interactions in E. coli monosaccharide-binding transporters (i.e., AraF, MglB, and RbsB) appear to be either fully conserved or replaced by isofunctional residues in D-xylose-binding proteins. In E. coli AraF and MglB, they include residues that sandwich sugar molecules inside the protein cleft by stacking interactions and those forming hydrogen bonds with sugars, which confer substrate specificity (34).

We also noted, for the first time, that XylF sequences possess a region closely matching a sequence pattern of binding proteins specific for hexoses and pentoses, the so-called cluster 2 signature. This consensus was defined as K(LIVFAG)3 

LxΩ(D(SGP)X3(GS)X(LIVA)X-A, where boldface letters indicate invariant residues in all of the relevant sequences available at the time (50). Our inclusion of XylFs in multiple sequence alignment revealed two residues in XylFs that deviate from this consensus (Fig. 1). Thus, we propose that the cluster 2 sequence signature be refined as (KDQ)(LIVFAG)3 

DN(SGP)X3(GS)X(LIVA)X-A, where boldface letters reflect the changes introduced. This may facilitate future sequence analyses of sugar-binding proteins.

At present, it is not clear which residues and/or interactions

Thermotoga maritima (29). These sequences, dubbed “linkers,” are also found in immunoglobulin 1A, a pullulanase, xylanases, and cellulases, where they join different domains within an enzyme (13, 26). It has been postulated that a linker in a cell membrane-anchored lipoprotein enhances enzyme performance by serving as an extended spacer relative to the cell surface (29). The noticeable lack of linkers in periplasmic monosaccharide-binding proteins (Fig. 1) supports this view. Further work is needed to determine whether the linker facilitates interactions of XylF with membrane-associated transport proteins during D-xylose transport.

The marked binding of D-xylose by E. coli carrying the re-
combinant XylF suggested that the protein also functions in vivo in T. ethanolicus. Our preliminary transport studies indicate that a high-affinity D-xylose uptake system does operate in T. ethanolicus (10). These observations imply the existence of T. ethanolicus transport of D-xylose in vivo in T. ethanolicus and cellulases, where they join different domains within an enzyme. (29). These sequences, dubbed “linkers,” may play a role in the interaction of proteins during D-xylose transport.

Furthermore, the XylFs identified in various bacterial species may be involved in the specific function of XylF. However, without experimental evidence, it is difficult to explain how an XylR homolog in T. ethanolicus would achieve simultaneous binding of the xylF palindrome spaced by the 23-bp loop. Future studies should show whether a looping mechanism, analogous to that observed in the t-arabinose operon regulation by AraC (40), may be involved in regulation of xylF expression in T. ethanolicus. Nevertheless, the presence of conserved palindromes adjacent to both xylAB and xylF promoters and similar expression patterns of the two operons with respect to the carbon sources suggests that both xylAB and xylF operons may be controlled by a similar mechanism of negative regulation. Interestingly, the inverted repeat implicated in the xylAB transcription termination coincides with the putative xylF operator sequence. Hence, this might serve as an example of versatility of cis-acting elements in transcriptional regulation of bacterial carbon metabolism.

A recent study showed that XylFs from E. coli and H. influ-
enzae are more closely related to the A. tumefaciens multiple-sugar-binding protein ChvE than to the E. coli periplasmic receptors of t-arabinose, D-glucose or D-galactose, or D-ribose (22). Our inclusion of T. ethanolicus XylF in the multiple sequence comparison confirmed this observation. Significantly, our sequence analysis indicated that most residues involved in protein-sugar interactions in E. coli monosaccharide-binding transporters (i.e., AraF, MglB, and RbsB) appear to be either fully conserved or replaced by isofunctional residues in D-xylose-binding proteins. In E. coli AraF and MglB, they include residues that sandwich sugar molecules inside the protein cleft by stacking interactions and those forming hydrogen bonds with sugars, which confer substrate specificity (34).
are responsible for the presumed thermostability of XylF in the thermophilic bacterium *T. ethanolicus*. The amino acid composition analysis of its predicted sequence (data not shown) compared to that of mesophilic XyLFs did not reveal any obvious differences, with one exception: there are three cysteines in the peptide chain of the *T. ethanolicus* XylF, unlike in the other two XyLFs, which completely lack cysteine residues. Now that the recombinant *T. ethanolicus* XylF is available, it will be possible to examine if a disulfide bridge is present in the mature, properly folded protein, and, if so, whether such a conformation is required for the protein thermostability, as has been documented for numerous other thermostable enzymes (14).

**ACKNOWLEDGMENT**

This work was supported by the U.S. Department of Agriculture under agreement 95-37500-1793.

**REFERENCES**


