JOURNAL OF BACTERIOLOGY, Feb. 1987, p. 612–618
0021-9193/87/020612-07$02.00/0

Vol. 169, No. 2

Sequence of the *Ampullariella* sp. Strain 3876 Gene Coding for Xylose Isomerase

GENA C. SAARI,† ANUR ASHOK KUMAR, GLENN H. KAWASAKI, MARGARET Y. INSLEY, AND PATRICK J. O'HARA*

ZymoGenetics, Inc., Seattle, Washington 98103

Received 4 August 1986/Accepted 31 October 1986

The nucleotide sequence of the gene coding for xylose isomerase from *Ampullariella* sp. strain 3876, a gram-positive bacterium, has been determined. A clone of a fragment of strain 3876 DNA coding for a xylose isomerase activity was identified by its ability to complement a xylose isomerase-defective *Escherichia coli* strain. One such complementation positive fragment, 2,922 nucleotides in length, was sequenced in its entirety. There are two open reading frames 1,182 and 1,242 nucleotides in length, on opposite strands of this fragment, each of which could code for a protein the expected size of xylose isomerase. The 1,182-nucleotide open reading frame was identified as the coding sequence for the protein from the sequence analysis of the amino-terminal region and selected internal peptides. The gene initiates with GTG and has a high guanine and cytosine content (70%) and an exceptionally strong preference (97%) for guanine or cytosine in the third position of the codons. The gene codes for a 43,210-dalton polypeptide composed of 393 amino acids. The xylose isomerase from *Ampullariella* sp. strain 3876 is similar in size to other bacterial xylose isomerases and has limited amino acid sequence homology to the available sequences from *E. coli*, *Bacillus subtilis*, and *Streptomyces violaceus-ruber*. In all cases yet studied, the bacterial gene for xylulose kinase is downstream from the gene for xylose isomerase. We present evidence suggesting that in *Ampullariella* sp. strain 3876 these genes are similarly arranged.

In a number of bacteria, utilization of D-(+)-xylose as an energy source initiates with the transport of xylose into the cell by a xylose-binding protein. Xylose is then converted by xylose isomerase to D-xylulose, which is in turn phosphorylated by xylulose kinase to D-xylulose-5-phosphate (1, 5, 25, 26). This phosphorylated intermediate is catabolized in the pentose phosphate and Embden-Meyerhoff pathways. A review of the organisms which are known to produce xylose isomerase and a description of the biochemical properties of these enzymes has been published (4).

Ampullariella sp. strain 3876 (ATCC 31351) is a gram-positive, filamentous, spore-forming bacterium classified in the order *Actinomycetales*. In *Ampullariella* sp., the gene for xylose isomerase can be induced to increased levels in the presence of xylose (S. E. Foley, P. J. Oriel, and C. C. Epstein, U.S. patent 4,308,349, 1981). In members of the related genus *Actinoplanes*, the gene for xylose isomerase is insensitive to catabolic repression (19). In *Salmonella typhimurium* the genes coding for the transport, isomerase, and kinase activities involved in xylose utilization are closely linked and under coordinate positive control (10, 25). There is evidence for similar gene organizations in *Escherichia coli* (5, 14, 15, 31) and *Bacillus subtilis* (30). The complete nucleotide sequence of the gene coding for xylose isomerase from *E. coli* (14, 24) and partial nucleotide and amino acid sequences for the *B. subtilis* (30) and *Streptomyces violaceus-ruber* (3) enzymes have been published.

Under certain conditions xylose isomerase catalyzes the conversion of D-glucose to D-fructose (32). This reaction is used industrially in the production of large quantities of high-fructose corn syrups. The xylose isomerase produced by *Ampullariella* sp. strain 3876 exhibits superior thermostability and activity over a wide range of conditions, which makes it attractive as an industrial enzyme (Foley et al., U.S. patent 4,308,349). However, *Ampullariella* sp. strain 3876 itself is difficult to use as a production organism, which makes it desirable to clone and express its xylose isomerase gene in a more convenient organism. Here, we report the complete nucleotide sequence and partial primary amino acid sequence of *Ampullariella* sp. strain 3876 xylose isomerase and present evidence that the isomerase and kinase genes are linked in this organism.

MATERIALS AND METHODS

Cloning of the *Ampullariella* sp. strain 3876 gene. *Ampullariella* sp. strain 3876 was provided by Dow Chemical Co., Midland, Mich. Cells were grown as described by Foley et al. (U.S. patent 4,308,349). DNA from the cells was isolated, digested with *Bam*HI, ligated to *Bam*HI-digested pUC13 (18), and used to transform a xylose-negative *E. coli* strain which contains the *xyl-5* mutation (2) as described by Kawasaki et al. (manuscript in preparation). The *xyl-5* mutation is in the gene coding for xylose isomerase in these cells. Transformants were plated on LB medium plus ampicillin and replica plated to xylose plates. Those transformants which complemented the *xyl-5* mutation were able to grow on the xylose plates. Plasmid DNA from the colonies growing on xylose medium was analyzed by restriction digests (Kawasaki et al., in preparation).

DNA sequencing. DNA sequencing was performed by the methods of Maxam and Gilbert (16) and Sanger et al. (23). Specific restriction fragments (Fig. 1) were purified by elution from acrylamide or agarose gels (28) and either sequenced directly (16) or subcloned into M13 cloning vector mp18 or mp19 (18) by established procedures (22). The ends of the subcloned fragments were sequenced in reactions

* Corresponding author.
† Present address: Zoological Institute, University of Zurich-Irchel, 8057 Zurich, Switzerland.

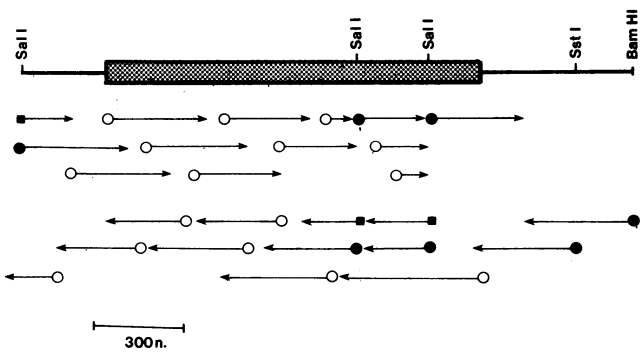


FIG. 1. Strategy used to sequence a fragment of *Ampullariella* sp. strain 3876 DNA that complemented a xylose isomerase-defective *E. coli* strain. The restriction sites used for sequencing as well as the extent of the sequence determined from each restriction site or primer are represented as follows: (■) sites sequenced by the chemical cleavage method (16), (●) sites sequenced by the dideoxy chain termination method (23) using a universal M13 primer, and (○) internal regions sequenced by the dideoxy chain termination method with synthetic oligonucleotides as primers. The boxed area is the gene coding for xylose isomerase. n, Nucleotides.

primed with a universal M13 *lac* primer. To sequence the internal portions of these fragments, the reactions were primed with specific oligonucleotides designed to hybridize with xylose isomerase sequence. The oligonucleotides (15- to 18-mers) and a universal M13 *lac* primer was synthesized on an Applied Biosystems 380A DNA synthesizer by using β-cyanoethyl phosphoramidites on CPG-LCAA solid supports. The high G+C content of the gene caused frequent compressions on the sequencing gels and necessitated such measures as the use of additional denaturants in the gels, the use of dITP in the place of dGTP in the sequencing reactions (17), and careful analysis of the sequence on both strands. Sequence data were analyzed with the IntelliGenetics (Mountain View, Calif.) software package.

CNBr digestion. *Ampullariella* sp. strain 3876 xylose isomerase (5 mg) (kindly provided by Dow Chemical Co.) was suspended in 70% formic acid (2 ml) and digested with CNBr (25 mg) in the dark at 22°C for 20 h. The digest was diluted 10-fold with water, lyophilized repeatedly, and subjected to chromatography on a Waters μ-Bondapak C-18 column (Waters Associates, Inc., Milford, Mass.).

Limited tryptic digestion. Xylose isomerase (5 mg) in 0.1 M sodium acetate (2 ml) was maleylated with maleic anhydride (6 mg) at pH 8.5 for 20 min to block the lysine groups. The maleylated protein was dialyzed against 0.2 M ammonium bicarbonate (pH 8.5) and subjected to tryptic digestion with tolylsulfonyl phenylalanyl chloromethyl ketone-treated trypsin (50:1 by weight) at 37°C for 8 h. The reaction was terminated by the addition of glacial acetic acid (30% final concentration), and the fragments were demaleylated by incubating in 30% acetic acid at 50°C for 36 h. The cyanogen bromide fragments and limited tryptic peptides were isolated and purified by high-pressure liquid chromatography on a Waters μ-Bondapak C-18 column (4.6 by 250 mm; 10 μm) with water-acetonitrile gradients containing 0.1% trifluoroacetic acid. The column fractions were monitored at 215 nm.

Amino acid sequence analysis. The automated Edman degradation of the intact *Ampullariella* sp. strain 3876 xylose isomerase and selected cyanogen bromide fragments and limited tryptic peptides was carried out with a gas-phase protein sequenator, model 470A (Applied Biosystems). Phenylthiohydantoin amino acids were separated on a Rainin C-18 column with a Varian model 5500 high-pressure liquid chromatography system. The phenylthiohydrantoin amino acids were identified by absorbance at 254 nm with an isocratic solvent system consisting of 60% 0.01 M sodium acetate (pH 4.5) and 40% acetonitrile.

RESULTS AND DISCUSSION

A plasmid containing a 2,922-base-pair *Bam*HI fragment of *Ampullariella* sp. strain 3876 DNA complemented the xylose isomerase defect in a xylose isomerase-defective *E. coli* strain (Kawasaki et al., in preparation). Within the 2,922-base-pair *Bam*HI fragment there are two open reading frames, both contained within a 1,925-base-pair *Sal*I-*Bam*HI fragment, which could code for a protein with the size expected for xylose isomerase. The strategy used to sequence this 1,925-base-pair fragment is summarized in Fig. 1. Of the large open reading frames, one is 1,182 nucleotides in length, initiates with a GTG 253 bases from the *Sal*I site, and terminates at position +1183. The second is on the opposite strand, initiates with a GTG at position +1167, 1,419 bases from the *Sal*I site, terminates at position −76, and is 1,242 nucleotides in length.

Intact purified xylose isomerase and three selected internal peptides were subjected to amino acid analysis. The amino acid sequence data (Table 1) were used to identify an open reading frame 1,182 nucleotides in length as the *Ampullariella* sp. strain 3876 gene that codes for the xylose isomerase enzyme. The translated nucleotide sequence of the gene is shown in Fig. 2. The identification of the GTG at position +1 as the initiation codon for the gene coding for xylose isomerase is consistent with the amino acid sequence of the amino terminus of the protein (Table 1). It is not uncommon for the genes of gram-positive bacteria to initiate with GTG (11).

The base composition within the 1,925-base-pair fragment is 201 (A) (17.0%), 165 (T) (13.9%), 385 (G) (32.5%), and 434 (C) (36.0%). The high G+C content (69.1%) is similar to the reported G+C content (73%) of total DNA from *Ampullariella digitalis* and correlates with the high G+C contents of DNA from other mesophilic actinomycetes, which range

TABLE 1. Amino acid sequence of xylose isomerase amino-terminal region and selected internal peptides

Sample[a]	Sequence[b]	Region[c]
Xylose isomerase	S L Q A T P D D K F S F G L W T V G W Q A R D A F G D A T – P V L	1–33
LT-20	S A F E D Y D A D A V G A K G Y G F V K L N Q L A I D – L L G	361–391
CB-2	V T T N L F T – P V F K D G G F T S N D	88–107
CB-5	Y L L L K E – A K A F – A D P E V Q A A	314–333

[a] LT, Limited tryptic digest peptide; CB, CNBr cleavage peptide.
[b] –, Unidentified residue.
[c] Numbering refers to Fig. 2.

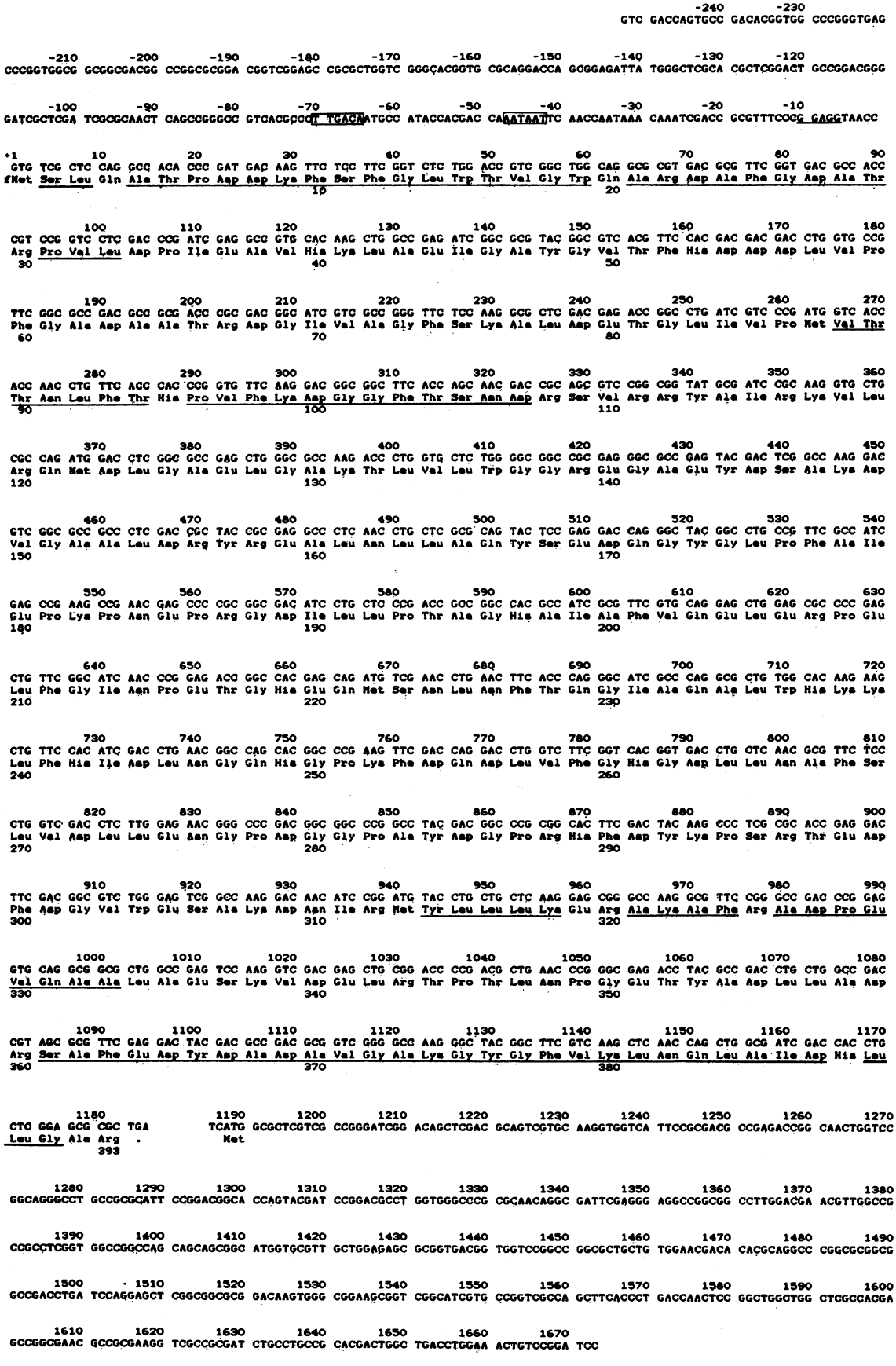


FIG. 2. DNA sequence of the *Ampullariella* sp. strain 3876 gene coding for xylose isomerase and flanking regions. The fragment shown here is a 1,925-base-pair *Sal*I-*Bam*HI subclone of a 2,922-base-pair *Bam*HI fragment obtained by complementation of a xylose isomerase-defective *E. coli* strain. Underlined amino acid residues were confirmed by amino acid sequence analysis. Boxed sequences are similar to promoter elements common to bacteria. The underlined sequence is the location of a potential Shine-Dalgarno sequence. The ATG at position 1188 may be the initiation codon for the gene coding for xylulose kinase.

TABLE 2. Codon usage in the *Ampullariella* sp. strain 3876 gene

Codon-amino acid	n	Codon-amino acid	n	Codon-amino acid	n	Codon-amino acid	n
UUU-Phe	0	UCU-Ser	0	UAU-Tyr	1	UGU-Cys	0
UUC-Phe	22	UCC-Ser	5	UAC-Tyr	11	UGC-Cys	0
UUA-Leu	0	UCA-Ser	0				
UUG-Leu	1	UCG-Ser	5			UGG-Trp	5
CUU-Leu	0	CCU-Pro	0	CAU-His	0	CGU-Arg	3
CUC-Leu	15	CCC-Pro	5	CAC-His	11	CGC-Arg	11
CUA-Leu	0	CCA-Pro	0	CAA-Gln	0	CGA-Arg	0
CUG-Leu	28	CCG-Pro	16	CAG-Gln	13	CGG-Arg	7
AUU-Ile	0	ACU-Thr	0	AAU-Asn	0	AGU-Ser	0
AUC-Ile	13	ACC-Thr	15	AAC-Asn	13	AGC-Ser	3
AUA-Ile	0	ACA-Thr	1	AAA-Lys	0	AGA-Arg	0
AUG-Met	4	ACG-Thr	2	AAG-Lys	18	AGG-Arg	0
GUU-Val	0	GCU-Ala	0	GAU-Asp	1	GGU-Gly	4
GUC-Val	14	GCC-Ala	27	GAC-Asp	37	GGC-Gly	30
GUA-Val	0	GCA-Ala	0	GAA-Glu	0	GGA-Gly	1
GUG-Val	8	GCG-Ala	17	GAG-Glu	24	GGG-Gly	3

from 60 to 75% (8). A high G+C content is also a characteristic of the DNA of extreme thermophiles. This characteristic is thought to contribute to the thermostability of nucleic acids and is important for the processes of replication, transcription, and translation at extremely high temperatures (13). In contrast to thermophiles which can proliferate at 90°C, the mesophilic actinomycetes in the vegetative stage can tolerate temperatures up to only 45°C and grow optimally at 30°C (19). However, a heat shock activates germination of the spores of actinomycetes (7) and may constitute a selective pressure reflected in the high G+C content of the DNA.

The codon usage of the *Ampullariella* sp. strain 3876 gene for xylose isomerase is shown in Table 2. There is an exceptionally strong preference for G or C (97%) in the third position of the codons (Table 3). Third-position preferences for G or C have been reported in other genes with a high G+C content, such as genes from *Streptomyces plicatus*, an actinomycete (20); *Thermus thermophilus*, a thermophilic organism (13); and *Halobacterium halobium*, an archaebacterium (6) (Table 4). This third-position preference for G or C probably reflects the flexibility of the third codon position in attaining an overall high G+C content rather than selective codon usage for optimizing translation (13).

Upstream of the initiation codon are sequences that resemble elements common to the promoters and ribosome binding sites of bacteria. At positions −10 to −6, the sequence GGAGG appears, which is similar to a Shine-Delgarno ribosome binding site consensus sequence (27). The consensus sequences of the RNA polymerase binding sites in bacteria are TTGACA near position −35 from the start of transcription and TATAAT near position −10 (21). Upstream from the *Ampullariella* sp. strain 3876 gene coding for xylose isomerase, at positions −70 to −65 from the start of translation, is the sequence TTGACA, and at positions −47 to −42 is the sequence AATAAT. The similarity of the

TABLE 3. Nucleotide preferences in the three codon positions in the *Ampullariella* sp. strain 3876 xylose isomerase codons

Codon position	No. (%) of nucleotides in codon position				
	U	C	A	G	G + C
1	50 (12.6)	109 (27.6)	69 (17.5)	166 (32.5)	275 (60.1)
2	105 (26.6)	93 (23.6)	129 (32.7)	67 (17.0)	160 (40.6)
3	9 (2.2)	232 (58.8)	2 (0.5)	151 (38.3)	383 (97.1)

TABLE 4. Comparison of third codon position usage

Species	% G + C in:		
	Total DNA	Gene	Third codon
Ampullariella sp. strain 3876[a]	73	70	97
S. plicatus[b]	73	67	92
T. thermophilus[c]	69	70	89
H. halobium[d]	67	61	82

[a] Xylose isomerase gene.
[b] Endo-β-*N*-acetyglucosaminidase gene (20).
[c] 3-Isopropylmalate dehydrogenase gene (13).
[d] Bacteriorhodopsin gene (6).

sequence and spatial arrangement of these regions to procaryotic promoter elements suggests that they may be functionally significant in transcription of the gene and that transcription might start at around positions −35 to −30. Upstream of the coding region of some procaryotic G+C-rich genes, such as the bacteriorhodopsin gene of *H. halobium* (6), there is a shift to relative A+T richness, and elements similar to procaryotic promoter consensus sequences are present, as is observed here. In other cases, such as the *leuB* gene of *T. thermophilus* (13) and the endo-β-*N*-acetylglucosaminidase H gene of *S. plicatus* (20), the overall G+C richness extends upstream of the coding region, and TTGACA and TATAAT elements are lacking.

The molecular weights and amino acid compositions of xylose isomerases from a large number of bacterial species have been determined (3, 4). The *Ampullariella* sp. strain 3876 gene for xylose isomerase codes for a protein 393 amino acids in length with a molecular weight of 43,210. This is similar to the monomer molecular weights of xylose isomerases from other bacteria (3, 4). Of the xylose isomerases for which at least partial amino acid sequence is available, we compare the amino acid composition of *Ampullariella* sp. strain 3876 xylose isomerase with the enzymes from *S. violaceus-ruber* (3) and *E. coli* (14, 24) in Table 5. The amino acid compositions show some similarities. For instance, these enzymes are low in cysteine con-

TABLE 5. Comparison of amino acid compositions of xylose isomerase from *Ampullariella* sp. strain 3876, *S. violaceus-ruber* (3), and *E. coli* (14, 24)

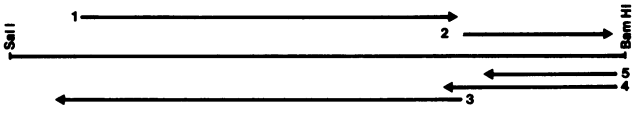
Amino acid	No. of amino acid residues (%) in:		
	Ampullariella sp. strain 3876	*S. violaceus-ruber*	*E. coli*
Alanine	44 (11.1)	54 (14.0)	44 (10.0)
Cysteine	0 (0.0)	1 (0.25)	4 (0.9)
Aspartic acid	38 (9.6)	37 (9.5)	24 (5.4)
Glutamic acid	24 (6.1)	24 (6.2)	33 (7.5)
Phenylalanine	22 (5.6)	22 (5.7)	25 (5.7)
Glycine	38 (9.6)	34 (8.7)	38 (8.6)
Histidine	11 (2.8)	10 (2.5)	19 (4.3)
Isoleucine	13 (3.3)	11 (2.8)	13 (2.9)
Lysine	18 (4.6)	12 (3.1)	24 (5.4)
Leucine	44 (11.1)	40 (10.3)	42 (9.5)
Methionine	4 (1.0)	8 (2.1)	12 (2.7)
Asparagine	13 (3.3)	9 (2.3)	20 (4.5)
Proline	21 (5.3)	19 (4.9)	14 (3.2)
Glutamine	13 (3.3)	12 (3.1)	25 (5.7)
Arginine	21 (5.3)	32 (8.2)	19 (4.3)
Serine	13 (3.3)	10 (2.6)	16 (3.6)
Threonine	18 (4.6)	17 (4.4)	19 (4.3)
Valine	21 (5.3)	18 (4.6)	23 (5.2)
Tryptophan	5 (1.3)	6 (1.5)	8 (1.8)
Tyrosine	12 (3.0)	11 (2.8)	18 (4.1)

FIG. 3. Comparison of amino acid sequences of the xylose isomerases of *Ampullariella* sp. strain 3876 , *E. coli* (14, 24), *B. subtilis* (30), and *S. violaceus-ruber* (3). Boxed residues are areas of homology. A gap is indicated by a –. Numbers to the right indicate the positions of the residues within the protein.

tent, and they share similar amounts of phenylanine, glycine, isoleucine, leucine, and threonine.

A comparison of the amino acid sequence of the *Ampullariella* sp. strain 3876 xylose isomerase to other available bacterial xylose isomerase sequences is shown in Fig. 3. There are two areas of extensive homology between the xylose isomerases of *Ampullariella* sp. strain 3876 and *E. coli* (14, 24). These occur at residues 135 to 142 and 180 to 187 of *Ampullariella* sp. strain 3876, which are almost exactly the same as residues 186 to 193 and 231 to 238 of *E. coli*. A comparison of the hydropathicity curve of amino acids 142 to 212 of *Ampullariella* sp. strain 3876 and amino acids 186 to 263 of *E. coli* is very similar, although the amino acid sequence is different (data not shown). The relationship of the sequence of the *Ampullariella* sp. strain 3876 xylose isomerase to residues 39 to 119 of the *B. subtilis* enzyme appears to be similar to the relationship of the same 64 amino acids to residues 32 to 112 of *E. coli* xylose isomerase. Although not enough data are available for an overall comparison, 6 out of the first 12 amino acids (counting from the amino-terminal serine) of *Ampullariella* sp. strain 3876 xylose isomerase are identical to 6 amino acids in the region 1 to 13 of the enzyme from *S. violaceus-ruber*, and three of the differences in this region are conservative replacements.

The *Ampullariella* sp. strain 3876 xylose isomerase can retain essentially all of its original activity after heating at 75°C for 24 h (Foley et al., U.S. patent 4,308,349). *S. violaceus-ruber* xylose isomerases are stable to more than 80°C (29), and the *E. coli* enzyme also is quite heat stable to 60°C (31). The thermostability of an enzyme molecule is contributed to by various types of intramolecular bonds. A characteristic of thermostable proteins is low cysteine content (9). *Ampullariella* sp. strain 3876 xylose isomerase contains no cysteine residues, *S. violaceus-ruber* xylose isomerase contains one, and *E. coli* xylose isomerase contains four. The aliphatic index proposed by Ikai (12) is defined as the relative volume of a protein occupied by the alphatic side chains of alanine, valine, isoleucine, and leucine. The index of thermostable proteins (sample mean, 92.6) is reportedly significantly higher than that of mesostable proteins (sample mean, 78.8) (12). The indexes for the xylose isomerases are 82.6 for *Ampullariella* sp.



ORF#	INITIATOR	TERMINATOR	LENGTH
1	253	1435	1182
2	1443		485+
3	1419	177	1242
4		1382	543+
5		1492	433+

FIG. 4. Open reading frames in the 1,925-base-pair *Sal*I-*Bam*HI fragment of *Ampullariella* sp. strain 3876 DNA. Open reading frame (ORF) 1 codes for xylose isomerase. Open reading frame 2 may code for xylulose kinase. Open reading frames 2, 4, and 5 extend beyond the end of the fragment.

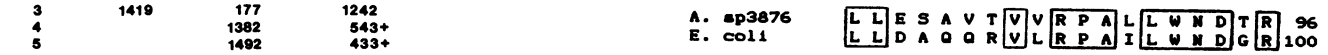


FIG. 5. Comparison of amino acids 81 to 100 of *E. coli* xylulose kinase (14) with positions 77 to 96 of the translated open reading frame downstream from the *Ampullariella* sp. strain 3876 xylose isomerase gene. Numbers to the right indicate the positions of the residues.

strain 3876, 78.4 for *S. violaceus-ruber*, and 73.4 for *E. coli*.

In addition to the open reading frame coding for xylose isomerase, there are four additional open reading frames, three of which extend beyond one end of the fragment (Fig. 4). These additional open reading frames could be due to the presence of more than one gene or may reflect the bias against the trinucleotides TAA, TGA, and TAG in a G+C-rich sequence. The large open reading frame (frame 3, Fig. 4) on the opposite strand is probably a reflection of the preference in the xylose isomerase gene for G or C in the third codon position (Table 4). In the xylose isomerase-coding region, an A is in the third codon position only twice, and frame 3 has a correspondingly low T content in the first codon positions. Whereas the significance of open reading frames 4 and 5 (Fig. 4) is unclear (and their presence may simply be a reflection of a high G+C content), it appears that open reading frame 2 may be an additional gene.

The xylose utilization genes of *E. coli* (14), *S. typhimurium* (25), and *B. subtilis* (30) are closely linked, in each case with the xylulose kinase gene downstream of the xylose isomerase genes. Several observations suggest that the xylulose kinase gene is just downstream of the xylose isomerase gene in *Ampullariella* sp. strain 3876 as well. Three nucleotides downstream of the xylose isomerase stop codon (Fig. 2), an ATG begins an open reading frame (frame 2, Fig. 4) which runs through the end of the fragment. This 485-nucleotide open reading frame is preceded by an apparent Shine-Dalgarno sequence at positions −12 to −15 from the ATG, and there is substantial homology between bases −9 to +3 of this open reading frame to the same bases in the xylose isomerase gene. The genes coding for xylose isomerase and xylulose kinase are separate transcription units in *E. coli* (14) but are thought to be part of the same transcription unit in *B. subtilis* (30). We do not observe a typical bacterial termination signal (a stem-loop structure followed by five thymidines) (21) downstream of the xylose isomerase stop codon. An absence of transcription termination signals would be expected if the gene for xylose isomerase were part of a transcription unit with the gene for xylulose kinase. Translation of the open reading frame produces an amino acid sequence with significant homology to the amino terminal sequences of the *E. coli* xylulose kinase gene (14) (Fig. 5). These considerations suggest that three bases downstream from the *Ampullariella* sp. strain 3876 gene that codes for xylose isomerase begins the gene that codes for xylulose kinase. If this is true, in all cases studied so far, the bacterial genes involved in xylose utilization are similarly arranged.

ACKNOWLEDGMENTS

We acknowledge Carol Epstein and Bill Dowd of Dow Chemical Co. and Frank Grant, Michael Tippie, Anne Bell, Teresa Gilbert, Mike Parker, Julie Holly, Lawrence Whitney, Margo Rogers, Ila McCullough, and Alan Upshall.

LITERATURE CITED

1. **Ahlem, C., W. Huisman, G. Neslund, and A. S. Dahms.** 1982. Purification and properties of a periplasmic D-xylose-binding protein from *Escherichia coli K-12*. J. Biol. Chem. **257**: 2926–2931.
2. **Bolivar, F., R. L. Rodriquez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. Gene **2**:95–113.
3. **Callens, M., H. Kersters-Hilderson, J. Vandekerckhore, O. Van Opstal, and C. K. De Bruyne.** 1985. Purification and some physiochemical properties of D-xylose isomerase from *Streptomyces violaceus-ruber*. Biochem. Intl. **11**:467–475.
4. **Chen, W.-P.** 1980. Glucose isomerase (a review). Process Bioc. **15**:30–41.
5. **David, J. D., and H. Wiesmeyer.** 1970. Control of xylose metabolism in *Escherichia coli*. Biochim. Biophys. Acta **201**:497–499.
6. **Dunn, R., J. McCoy, M. Simsek, A. Majumdar, S. H. Chang, U. L. Raj Bhandary, and H. G. Khorna.** 1981. The bacteriorhodopsin gene. Proc. Natl. Acad. Sci. USA **78**:6744–6748.
7. **Ensign, J. C.** 1978. Formation, properties, and germination of actinomycete spores. Annu. Rev. Microbiol. **32**:185–219.
8. **Farina, G., and S. G. Bradley.** 1970. Reassociation of deoxyribonucleic acids from *Actinoplanes* and other actinomycetes. J. Bacteriol. **102**:30–35.
9. **Freidman, S. M. (ed.).** 1978. Biochemistry of thermophily. Academic Press, Inc., New York.
10. **Ghangas, G. S., and D. B. Wilson.** 1984. Isolation and characterization of the *Salmonella typhimurium* LT2 xylose regulon. J. Bacteriol. **157**:158–164.
11. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translation initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
12. **Ikai, A.** 1980. Thermostability and aliphatic index of globular proteins. J. Biochem. **88**:1895–1898.
13. **Kagawa, Y., H. Nojime, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima.** 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. J. Biol. Chem. **259**:2956–2960.
14. **Lawliss, V. B., M. S. Dennis, E. Y. Chen, D. H. Smith, and D. J. Henner.** 1984. Cloning and sequencing of the xylose isomerase and xylulose kinase genes of *Escherichia coli*. Appl. Environ. Microbiol. **47**:15–21.
15. **Maleszka, R., P. Y. Wang, and H. Schneider.** 1982. A Col E1 hybrid plasmid containing *Escherichia coli* genes complementing D-xylose negative mutants of *Escherichia coli* and *Salmonella typhimurium*. Can. J. Biochem. **60**:144–151.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
17. **Mills, D. R., and F. R. Kramer.** 1979. Structure independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA **76**:2232–2235.
18. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26**:101–106.
19. **Parenti, F., and C. Coronelli.** 1979. Members of the genus *Actinoplanes* and their antibiotics. Annu. Rev. Microbiol. **33**:389–411.
20. **Robbins, P. W., R. B. Trimble, D. F. Wirth, C. Hering, F. Maley, G. F. Maley, R. Das, B. W. Gibson, N. Royal, and K. Biemann.** 1984. Primary structure of the *Streptomyces* enzyme endo-β-N-acetyl glucosaminidase H. J. Biol. Chem. **259**: 7577–7583.
21. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
22. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **43**:161–178.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:64–70.
24. **Schellenberg, G. D., A. Sarthy, A. E. Larson, M. P. Backer, J. W. Crabb, M. Lidstrom, B. D. Hall, and C. E. Furlong.** 1984. Xylose isomerase from *Escherichia coli*. J. Biol. Chem. **259**:6826–6832.
25. **Shamanna, D. K., and K. E. Sanderson.** 1979. Genetics and regulation of D-xylose utilization in *Salmonella typhimurium*

LT2. J. Bacteriol. **139**:71–79.
26. **Shamanna, D. K., and K. E. Sanderson.** 1979. Uptake and catabolism of D-xylose in *Salmonella typhimurium* LT2. J. Bacteriol. **139**:64–70.
27. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16s ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
28. **Smith, H. O.** 1980. Recovery of DNA from gels. Methods Enzymol. **65**:371–379.
29. **Takasaki, Y., Y. Kosugi, and A. Kanbayashi.** 1969. Studies on sugar-isomerizing enzyme. Agric. Biol. Chem. **33**:1527–1534.
30. **Wilhelm, M., and C. P. Hollenberg.** 1984. Selective cloning of *Bacillus subtilis* xylose isomerase and xylulose kinase genes in *Escherichia coli* by IS5-mediated expression. EMBO J. **3**:2555–2560.
31. **Wovcha, M. G., D. L. Steuerwald, and K. E. Brooks.** 1983. Amplification of D-xylose and D-glucose isomerase activities in *Escherichia coli* by gene cloning. Appl. Environ. Microbiol. **45**:1402–1404.
32. **Yamanaka, K.** 1966. D-xylose isomerase. Methods Enzymol. **9**:588–593.