Sequencing, Cloning, and High-Level Expression of the pfp Gene, Encoding a PPi-Dependent Phosphofructokinase from the Extremely Thermophilic Eubacterium Dictyoglomus thermophilum

YAN-HUAI R. DING,* RON S. RONIMUS, AND HUGH W. MORGAN
Thermophile Research Unit, Department of Biological Sciences, The University of Waikato, Hamilton, New Zealand

Received 21 March 2000/Accepted 26 May 2000

The sequencing, cloning, and expression of the pfp gene from Dictyoglomus thermophilum, which consists of 1,041 bp and encodes a pyrophosphate-dependent phosphofructokinase, are described. A phylogenetic analysis indicates that the enzyme is closely related to the pyrophosphate-dependent enzyme from Thermoproteus tenax. The recombinant and native enzymes share a high degree of similarity for most properties examined.

Phosphofructokinase (PFK) is one of the key enzymes in glycolysis. The textbook version of ATP-dependent PFK (ATP-PFK) is present throughout the domains of Eucarya, Bacteria, and Archaea, although it is not yet known if the single archaean ATP-PFK recently described is related by sequence (6, 11, 15). Since the first pyrophosphate-dependent PFK (PPi-PFK) was reported by Reeves et al., who identified it in Entamoeba histolytica (21), a large number of PPi-PFKs have been found in higher plants, primitive eukaryotes, and bacteria, and a single example has been found in archaea (14, 17, 22, 25, 28). It has been suggested previously that PPi-PFKs may represent a single example has been found in archaea (14, 17, 22, 25, 28). Furthermore, because PPi-PFKs can function more readily than ATP-PFKs as the near-universal energy carrier (4, 12, 16). Furthermore, because PPi-PFKs can function more readily than ATP-PFKs in the gluconeogenic direction, they could also have acted as primitive fructose-1,6-bisphosphatases. Dictyoglomus thermophilum Rt46 B.1 is an extremely thermophilic bacterium isolated from a New Zealand hot spring and is deeply rooted near the base of the order Thermotogales (18). The Dictyoglomus PPi-PFK has recently been purified and characterized (8). In this paper, we describe the cloning, phylogeny, and overexpression of the first PPi-PFK from an extremely thermophilic bacterium. D. thermophilum Rt46 B.1 was obtained from the Thermophile Research Unit Culture Collection, Hamilton, New Zealand, and grown in Dictyoglomus medium (18). The Escherichia coli strain used for cloning and expression experiments was JM109 (Promega Life Sciences), and it was grown at 37°C with vigorous aeration in Luria-Bertani broth supplemented with ampicillin (100 μg/ml). The expression plasmid pKK223-3 was obtained from Pharmacia Biotech. Genomic DNA from D. thermophilum was prepared as described by Sambrook et al. (24). Large-scale plasmid DNA was purified from E. coli by using the alkaline lysis method combined with double cesium chloride gradient purification. Restriction digests, electrophoresis, and Southern blotting were carried out according to standard methods (24). In order to clone the full-length pfp gene from D. thermophilum, a 350-bp fragment was initially amplified, using AmpliTaq Gold polymerase (Perkin-Elmer Cetus), by PCR. The sense and antisense degenerate primers were designed according to the N-terminal sequence of the purified native protein and the internal fructose-6-phosphate (F-6-P) binding conserved sequence, respectively. The PCR product was used to probe genomic DNA of Dictyoglomus strain Rt46 B.1 digested with 16 restriction enzymes. Analysis of the Southern data identified several bands of appropriate sizes within the Sau3AI, RasI, and Sau96I digests. The inverse-PCR technique was employed, and with each enzyme, a single fragment was amplified for sequencing (fragments were 360, 700, and 2,200 bp, respectively) (7).

Identification of ORF encoding the pfp gene. The ORF representing the full-length sequence of the D. thermophilum PPi-PFK gene was amplified using a forward primer corresponding to the N terminus (the first seven codons) and containing an upstream EcoRI site (in bold) and a 5′-end spacer (5′-GGAGAACAT TGT AAA ATG GGT G-3′), and a reverse primer corresponding to the C terminus (the last seven codons) and containing a flanking HindIII site (in bold) and a 5′-end spacer (5′-GGCGAGAAG CTT TAC TTA TTA AAG AAA GTT TTT ACG-3′). The complete sequence of 1,233 nucleotides obtained from overlapping inverse-PCR contigs contains an ORF of 1,041 bp with 346 codons, beginning with an ATG and ending with a TAA stop codon. One hairpin sequence downstream of the stop codon which could act as a transcription terminator was found (20). Potential promoter sites at the 5′ end of the coding region were also identified. For example, a Pribnow-like box sequence (TATAAT) is located 41 nucleotides upstream from the ATG start codon and is similar to the −10 (TATAAT) promoter sequence. In addition, the TTAGCA sequence located 17 bases upstream from the TAAAAT sequence is similar to the −35 (TTGACA) promoter sequence (20). Finally, a potential ribosome binding site (AGGAGG) was also identified and is located 4 nucleotides upstream of the start codon. The codon usage for Dictyoglomus PFK, as expected, reflected the G+C content of the genomic DNA, which is 29.3 mol% (18). For example, among the 43 glycine codons, only 2 were terminated with a C and 3 were terminated with a G. In addition, all codons for phenylalanine, proline, and threonine were terminated with either an A or a U (not shown).

Tree construction and phylogenetic comparison. Ten representative amino acid sequences of PFKs from eukaryotes, bacteria, and the crenarchaeon Thermoproteus tenax were re-
FIG. 1. Multiple alignment of amino acid sequences of PFKs from *D. thermophilum* and nine other species, carried out by using Clustal W (version 1.6).

<table>
<thead>
<tr>
<th>Species</th>
<th>PFK Gene</th>
<th>Residue Involved</th>
<th>Species</th>
<th>PFK Gene</th>
<th>Residue Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. thermophilum</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP, F-6-P, PEP</td>
</tr>
<tr>
<td><em>A. manethonica</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfk</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>T. tenax</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>T. maritima</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfk</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>E. histolytica</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>S. coelicolor</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfk</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>B. brucei</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>T. pallidum</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfk</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>T. maritima</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfk</em></td>
<td>ATP</td>
</tr>
<tr>
<td><em>B. stearothermophilus</em></td>
<td>pfp</td>
<td>-</td>
<td><em>B. stearothermophilus</em></td>
<td><em>pfp</em></td>
<td>ATP</td>
</tr>
</tbody>
</table>

Asterisks, identical residues; dashes, gaps; dots, highly conserved residues.
D. thermophilum (pfk)  
A. methanonica (pfp)  
T. tenax (pfp)  
T. maritima (pfp)  
E. histolytica (pfp)  
S. coelicolor (pfk)  
T. brucei (pfk)  
T. pallidum (pfk)  
T. maritima (pfk)  
E. coli (pfk)  
B. stearothermophilum (pfk)  

D. thermophilum (pfp)  
A. methanonica (pfp)  
T. tenax (pfp)  
T. maritima (pfp)  
E. histolytica (pfp)  
S. coelicolor (pfk)  
T. brucei (pfk)  
T. pallidum (pfk)  
T. maritima (pfk)  
E. coli (pfk)  
B. stearothermophilum (pfk)  

FIG. 1—Continued.
retrieved from sequence databases and aligned (1) with the PP_i-PFK sequence from *D. thermophilum* (Fig. 1). The F-6-P binding sites were highly conserved in all PFKs examined. *Dictyoglomus* PFK has almost the same amino acid residues for F-6-P binding as those from *Amycolatopsis methanolica*, *T. tenax*, *S. coelicolor*, *B. stearothermophilus*, and *E. coli* (9, 23). The PFK from *Dictyoglomus* also possesses amino acid residues in the phosphoryl binding site identical to those for *A. methanolica*, *T. tenax* (PP_i-PFK), and *S. coelicolor* (ATP-PFK). A phylogenetic tree was generated from 22 representative amino acid sequences (Fig. 2). The *D. thermophilum* sequence is most homologous to the group III PFKs, as defined by Siebers et al. (25), including the PP_i-PFKs from *T. tenax* and *A. methanolica*, the putative PP_i-PFK from *Mycobacterium tuberculosis*, and the ATP-PFK from *S. coelicolor*. The full sequence of the *Dictyoglomus pfp* gene agreed completely with the N-terminal sequence obtained from the purified native protein, except for the lack of the first methionine residue in the native enzyme, which supports our previous suggestion, based on the sequence, that the enzyme has homology with the *T. tenax* enzyme (8). The data from the alignment of 22 sequences and the phylogenetic tree strongly support the contention that the central carbohydrate metabolism of glycolysis was established before the segregation of the three domains of life. The fact that all of the sequences presented in the phylogenetic tree can be aligned, despite their varying phosphoryl
donor specificities, demonstrates their homology and therefore, their likely evolution from a single common ancestral sequence (9, 11).

A number of X-ray crystallographic and site-directed mutagenesis studies have investigated the roles of PFK amino acid residues in substrate binding and catalysis (3, 9, 19, 23), as well as those amino acid residues related to allosteric properties (5, 27). For example, when the Glu\textsuperscript{187} of \textit{E. coli} PFK is replaced by Asp or Leu, the allosteric transition is abolished. The purified recombinant enzyme (Table 1) was obtained by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and subsequent purification steps, as described for the purification of the native enzyme (8). The final yield of the purified recombinant enzyme was 53 mg of protein from 800 ml of induced culture (approximately 5 g [wet weight]), with a yield of the purified recombinant enzyme being 100%.

The conditions for purification and characterization of the recombinant \textit{D. thermophilum} PPK enzymes were essentially the same as those described for the purification of the native enzyme (8). The final yield of the purified recombinant enzyme was 53 mg of protein from 800 ml of induced culture (approximately 5 g [wet weight]) of cell pellet, indicating high-level expression of the enzyme (Table 1). A single band was obtained by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) from the purified \textit{Dictyoglomus} PPK recombinant enzyme (Table 1; Fig. 3). The recombinant and native enzymes had the same estimated molecular weights (approximately 37,000) when both were run on the same SDS-PAGE gel (data not shown). In addition, a comparison of the native and recombinant \textit{Dictyoglomus} PPK enzymes demonstrates that they possess a high degree of similarity (Table 2). Most of the biochemical and kinetic properties of the recombinant enzyme were very similar to those of the native enzyme (8), including, for example, thermostability and the extreme sensitivity to Cu\textsuperscript{2+}.

### TABLE 1. Purification of the recombinant PP\textsubscript{i}-PFK from \textit{D. thermophilum} R146 B.1

<table>
<thead>
<tr>
<th>Step</th>
<th>Total activity (U)</th>
<th>Total protein (mg)</th>
<th>Sp act (U/mg)</th>
<th>Purification (fold)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heated supernatant</td>
<td>757</td>
<td>2,300</td>
<td>0.33</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Pheny1-Sepharose</td>
<td>516</td>
<td>329</td>
<td>1.6</td>
<td>4.8</td>
<td>68</td>
</tr>
<tr>
<td>Q-Sepharose</td>
<td>363</td>
<td>79</td>
<td>4.6</td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>Red dye 120</td>
<td>328</td>
<td>53</td>
<td>6.2</td>
<td>19</td>
<td>43</td>
</tr>
</tbody>
</table>

### TABLE 2. Comparison of some properties of native and recombinant \textit{D. thermophilum} PP\textsubscript{i}-PFK

<table>
<thead>
<tr>
<th>PFK</th>
<th>Mg\textsuperscript{2+} (mM)</th>
<th>pl</th>
<th>Mol wt</th>
<th>Forward pH optimum</th>
<th>$K_a$ (mM)</th>
<th>Half-life (min) at 80°C</th>
<th>% Activity in response to 1 μM Cu\textsuperscript{2+}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant</td>
<td>1–3</td>
<td>4.3</td>
<td>37,000</td>
<td>5.8</td>
<td>0.136</td>
<td>165</td>
<td>40</td>
</tr>
<tr>
<td>Native</td>
<td>1–3</td>
<td>4.4</td>
<td>37,000</td>
<td>5.9</td>
<td>0.127</td>
<td>170</td>
<td>38</td>
</tr>
</tbody>
</table>

\(^a\) Determined by SDS-PAGE.

\(^b\) Native and recombinant enzymes (150 μg ml\textsuperscript{-1}) were incubated at 80°C in buffer containing 50 mM MOPS, 2 mM MgCl\textsubscript{2}, and 0.02% Triton X-100, pH 6.4, under a mineral oil overlay for different lengths of time. Residual activity at 50°C was then assayed and compared to that of the unheated control.

FIG. 3. SDS-PAGE gel (silver-stained) of fractions obtained during purification of the recombinant \textit{Dictyoglomus} PP\textsubscript{i}-PFK. Lane 1, molecular mass markers, as follows: phosphorylase b (molecular mass, 94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20.1 kDa), and α-lactalbumin (14.4 kDa); lane 2, cell extract following heat treatment; lanes 3 through 5, fractions obtained after purification through phenyl-Sepharose, Q-Sepharose, and red dye 120 ligand, respectively. Lanes 2, 3, 4, and 5 contained 5.0, 1.0, 0.75, and 0.75 μg of protein, respectively.
context, Mertens (13) has suggested that the role of PPi-PFK is that of a glycolytic enzyme adapted to anaerobiosis. However, other investigators have pointed out that PPi-PFKs (especially those in Archaea and extremely thermophilic bacteria) are likely to be more ancient than ATP-PFKs (2, 16). The results from the Dictyoglomus PPi-PFK strongly support the latter hypothesis. As seen with the enzyme from T. tenax, the Dictyoglomus enzyme is the first biochemically characterized extremely thermophilic bacterial PPi-dependent enzyme and offers another opportunity to gain insight into the differentiation of PFK substrate specificities and the characteristics of the phenotype of the original ancestral PFK precursor. The group III enzymes are suggested to be of a more ancient origin than group II enzymes, so it follows that the PPi-PFKs from T. tenax and extremely thermophilic bacteria may represent more ancient enzymes than the enzymes in mesophilic bacteria, primitive eukaryotes, and higher plants. In support of this latter contention, in two cases, that of S. coelicolor and Trypanosoma brucei, there is strong sequence evidence for the evolution of ATP-PFKs from PPi-PFKs (2, 15). Finally, it is hoped that the X-ray crystallographic determination of the structure of PPi-PFK from D. thermophilum RT46 B.1, which is in progress, will help clarify the phylogenetic origins of PFKs.

Nucleotide sequence accession number. The Dictyoglomus pfp gene sequence data has been submitted to GenBank under accession number AF268276.

We thank the Royal Society Marsden Science Foundation and the University of Waikato for financial support during the course of this study. We also thank Pat Gread and Cameron Evans for assistance with the mass spectrometry.

REFERENCES