Teale, F. W. J. (1960) *Biochem. J. 76*, 381.
Thanabal, V., de Ropp, J. S., & La Mar, G. N. (1987) *J. Am. Chem. Soc. 109*, 7516.
Ugarova, N. N., Rozhkova, G. D., & Berezin, I. V. (1979) *Biochim. Biophys. Acta 570*, 31.
Weber, R. E., & Tanford, C. (1959) *J. Am. Chem. Soc. 81*, 3255.
Welinder, K. G. (1979) *Eur. J. Biochem. 96*, 483.
Wiberg, K. B. (1960) *Laboratory Technique in Organic Chemistry*, pp 240–251, McGraw-Hill, New York.
Yeoman, L. C., & Hager, L. P. (1980) *Biochem. Biophys. Res. Commun. 97*, 1233.
Zaks, A., & Klibanov, A. M. (1986) *J. Am. Chem. Soc. 108*, 2767.

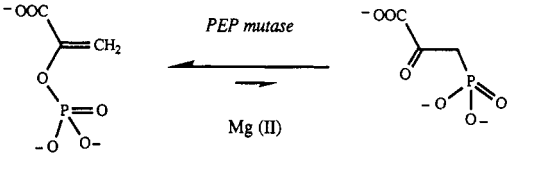
Phosphonate Biosynthesis: Molecular Cloning of the Gene for Phosphoenolpyruvate Mutase from *Tetrahymena pyriformis* and Overexpression of the Gene Product in *Escherichia coli*[†,‡]

H. Martin Seidel, David L. Pompliano,[§] and Jeremy R. Knowles*
Departments of Chemistry and Biochemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138
Received October 9, 1991; Revised Manuscript Received December 16, 1991

ABSTRACT: The phosphoenolpyruvate mutase gene from *Tetrahymena pyriformis* has been cloned and overexpressed in *Escherichia coli*. To our knowledge, this is the first *Tetrahymena* gene to be expressed in *E. coli*, a task made more complicated by the idiosyncratic codon usage by *Tetrahymena*. The N-terminal amino acid sequence of phosphoenolpyruvate mutase purified from *T. pyriformis* has been used to generate a precise oligonucleotide probe for the gene, using in vitro amplification from total genomic DNA by the polymerase chain reaction. Use of this precise probe and oligo(T) as primers for in vitro amplification from a *T. pyriformis* cDNA library has allowed the cloning of the mutase gene. A similar amplification strategy from genomic DNA yielded the genomic sequence, which contains three introns. The sequence of the DNA that encodes 10 amino acids upstream of the N-terminal sequence of the isolated protein was found by oligonucleotide hybridization to a subgenomic library. These 10 N-terminal amino acids are cleanly removed in *Tetrahymena* in vivo. The full mutase gene sequence codes for a protein of 300 amino acids, and it includes two amber (TAG) codons in the open reading frame. In *Tetrahymena*, TAG codes for glutamine. When the two amber codons are each changed to a glutamine codon (CAG) that is recognized by *E. coli* and the gene is placed behind a promoter driven by the T7 RNA polymerase, expression in *E. coli* is observed. The mutase gene also contains a large number of arginine AGA codons, a codon that is very rarely used by *E. coli*. Cotransformation with a plasmid carrying the *dnaY* gene [which encodes $tRNA^{Arg}(AGA)$] results in more than 4-fold higher expression. The mutase then comprises about 25% of the total soluble cell protein in *E. coli* transformants. The mutase gene bears significant similarity to one other gene in the available data bases, that of carboxyphosphonoenolpyruvate mutase from *Streptomyces hygroscopicus*, an enzyme that catalyzes a closely related transformation. Due to the large evolutionary distance between *Tetrahymena* and *Streptomyces*, this similarity can be interpreted as the first persuasive evidence that the biosynthesis of phosphonates is an ancient metabolic process.

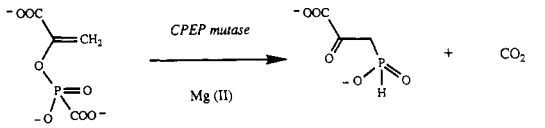
The enzyme phosphoenolpyruvate mutase (EC 5.4.2.9) catalyzes the first phosphorus–carbon bond-forming step in the biosynthesis of naturally-occurring phosphonates (Scheme I). For example, in the biosynthetic pathway that leads in *Streptomyces* to the herbicide bialaphos, the conversion of the phosphate ester in phosphoenolpyruvate (PEP) to the phosphonate in phosphonopyruvate is catalyzed by PEP mutase (Bowman et al., 1988; Seidel et al., 1988; Hidaka et al., 1989), while the reduction of the phosphonate to phosphinate (Scheme II) is catalyzed by a related enzyme, carboxyphosphonoenolpyruvate mutase (Hidaka & Seto, 1989; Hidaka et al., 1990). All compounds in nature containing a reduced form of phosphorus (that is, phosphonates or phosphinates) appear to owe their existence to PEP mutase or its congener.

Scheme I: PEP Mutase-Catalyzed Reaction



phosphoenolpyruvate (PEP) phosphonopyruvate

Scheme II: CPEP Mutase-Catalyzed Reaction



carboxyphosphonoenolpyruvate (CPEP) phosphinopyruvate

[†] Supported by the National Institutes of Health and the National Science Foundation and Eli Lilly (through predoctoral fellowships to H.M.S.).
[‡] The nucleotide sequence in this paper has been submitted to GenBank under accession number M85236.
* To whom correspondence should be addressed.
[§] Present address: Merck, Sharp & Dohme Research Laboratories, Department of Cancer Research, West Point, PA 19486.

0006-2960/92/0431-2598$03.00/0

Phosphonates (and phosphinates) are present in a variety of lower organisms ranging from *Streptomyces* to the freshwater snail schistosome vector *Biomphalaria* (Hilderbrand, 1983; Hori et al., 1984; Mastalerz, 1984). In the ciliated protozoan *Tetrahymena*, for example, aminoethylphosphonate (ciliatine), a key intermediate in the phosphonate biosynthetic pathway, is incorporated into the cell membrane as phosphonolipid. Other ciliates such as *Entodinium caudatam* also synthesize phosphonolipids. The presence of a phosphorus–carbon bond in the lipid bilayer confers a higher hydrolytic stability to the cell membrane, which is thought to provide an adaptive advantage for *Entodinium* in the gut of many ruminants (Hilderbrand, 1983).

Since the isolation of PEP mutase (Bowman et al., 1988; Seidel et al., 1988; Hidaka et al., 1989), considerable effort has been applied to the resolution of the mechanistic details of the transformation that it catalyzes, and although some progress has been made (Freeman et al., 1989; Seidel et al., 1990), the mechanistic pathway of the enzyme-catalyzed reaction remains a mystery. Physical studies on the mutase have been hampered by the relatively small amounts of the enzyme obtainable from *Tetrahymena*. The availability of large amounts of the mutase became a priority, and the cloning of the gene that encodes PEP mutase in *Tetrahymena* and the overexpression of the gene product in *Escherichia coli* was a natural solution to this problem.

To date, there have been no published reports of the expression of any *Tetrahymena* gene in other species. Indeed, the direct expression of *Tetrahymena* genes in *E. coli* will often be prevented by the existence of amber (TAG) and ochre (TAA) codons in the open reading frames of most *Tetrahymena* genes (Martindale, 1989). In *Tetrahymena* [as in several other ciliates, e.g., Gray et al. (1991); as well as the alga *Acetabularia*, Schneider et al. (1989)], amber and ochre codons code for glutamine, and these glutamine codons are used almost exclusively at the expense of the "normal" glutamine codons CAG and CAA. Expression of *Tetrahymena* genes in *E. coli* could also be hampered by the almost exclusive use by *Tetrahymena* of AGA to encode arginine, for this codon occurs with extreme rarity in *E. coli* genes (Gouy & Gautier, 1982). Indeed, genes that contain many AGA codons are known to be expressed at low levels (if at all) in *E. coli* (Brinkmann et al., 1989).

We report here the cloning of the *Tetrahymena* gene encoding PEP mutase and the resolution of all these expression problems, as evidenced by the high overexpression of the *Tetrahymena* PEP mutase gene product in *E. coli*.

Experimental Procedures

Strains and Media. The amicronucleate strain GL of *Tetrahymena pyriformis* (ATCC 30331) was used both as the source of purified phosphoenolpyruvate mutase and as the donor strain for mRNA and genomic DNA. *E. coli* strain DH5αMCR (Gibco BRL, Gaithersburg, MD) was used as the host strain for library construction and screening. *E. coli* strain DH5α was used as the host strain for the large-scale preparation of plasmid DNA (for sequencing and other manipulations). Luria broth (Gibco BRL) was used for all growths of *E. coli* (except where otherwise noted).

Phosphonopyruvate. Triethyl phosphonopyruvate was prepared according to Coutrot et al. (1978) and was deprotected with trimethylsilyl bromide (Lancaster Synthesis Ltd., Windham, NH) and purified according to Seidel et al. (1990). The phosphonopyruvate thus obtained was quantitated by three methods: (a) derivatization to the semicarbazone according to Anderson et al. (1984) ($\Delta\epsilon_{253nm}$ = 10000); (b) total phosphate assay after treatment with 70% (w/v) perchloric acid (Chen et al., 1956); and (c) enzyme assay with PEP mutase–pyruvate kinase/ADP–lactate dehydrogenase/NADH, following the decrease in NADH absorbance for which $\Delta\epsilon_{340nm}$ = 6220 (Horecker & Kornberg, 1948). The three methods gave consistent results.

Synthetic Oligonucleotides. The oligonucleotides synthesized were (5′–3′)

01 GACTCGAGTCGACATCGATTTTTTTTTTTTTT
02 ACAGCAGGATCCTCCACCAGAAAGACYACYCAA
03 ACAGCACTGCAGCGGANARACCGTTGTGRGCTTCCAT
04 CAGCCCGAGGATCCAATGTCCTCCACCAGAAAGACTACTCAATTGAAGAACAT GATCCAATCC
05 GGAAGTTTTGGAATTCATGAG
06 GGCTTGCAAGGACACCCAAC
07 AACTACCACACCGTTCCCACC
08 TTAATTAAGCGTGTAAGAGGG
09 GATGACCATATTGACACCCCAC
10 CTTGCAAGCCTTGATCTTAGC
11 ACTCATGAATTCCAAAACTTCC
12 GGATCCTCGCGATTAATTAAGCGTGTAAGAGGG
13 CTTAAAGAACCCTCTGAAATC
14 CTTACTCATTCTTTTATTCCG
15 AGCGACGACCTGGAAATCAGC
16 ACTTCCAAAACCTGAGTGTAAGAAGC
17 GAAGAAACTGGCTTCAAGGGTATCTGGGGTTCCGGTTTGTCCATCTCTGCTGCC
18 GCGCGGATCCATTGGAGGATGATTAAATGTTGGCCAACTCTCTTAAGTCC
19 CGCTGGATCCTGCAGGGTACCACTAGTCACAGATATTTCTTGTCAGCTTG
20 GCGCGGATCCATTGGAGGATGATTAAATGTCTTCTACCAGAAAGACCACC

where Y specifies an equimolar mixture of C and T, R specifies an equimolar mixture of A and G, and N specifies an equimolar mixture of A, T, G, and C.

Plasmids. Cloning was performed in the plasmid pUC18 (Pharmacia LKB Biotechnology Inc., Piscataway, NJ) or pBS(+) (Stratagene, La Jolla, CA). Subcloning was performed in M13mp18 rf form DNA (New England Biolabs, Beverly, MA) and the plasmid T7pT5T (which was a kind gift of Dr. R. Evans, Synergen Inc., Boulder, CO). We are grateful to Dr. S. Miller (Biogen Inc., Cambridge, MA) for providing the *dnaY* gene (Garcia et al., 1986) in pSM102-2, a derivative of pACYC177 (Chang & Cohen, 1978).

Growth and Maintenance of Strains. T. pyriformis GL cultures were grown in proteose peptone (2% w/v) (Difco Laboratories, Detroit, MI) at 28–29 °C. Starter cultures (1 mL) were used to inoculate culture medium (150 mL) that was then incubated for 2 days, the resulting culture being used to inoculate fresh medium (1.5 L in 4-L baffled flasks). Cells were harvested by centrifugation (700*g*, 15 min) after 2 more days when the A_{550nm} was 1.0. Overexpressing constructs were maintained in *E. coli* strain BL21 (DE3) in ZYG medium [Bacto tryptone from Difco (10 g), Bacto yeast extract from Difco (5 g), NaCl (5 g), and glucose (4 g) in a total volume of 1 L]. Cultures were grown at 37 °C with shaking (200 rpm) to an A_{550nm} of 1.0 [the growing cultures were shaken at 30–32 °C for 1 h before induction with isopropyl thio-β-D-galactoside (IPTG, 0.5 mM; Gibco BRL)]. Cultures were grown at 30–32 °C for 4 h before being harvested by centrifugation. For routine transformations, cells were made competent by the $CaCl_2$ method (Sambrook et al., 1989). Strains carrying pBS or pUC derivatives were maintained on plates or in liquid culture, in the presence of ampicillin (200 μg/mL). Strains carrying T7pT5T derivatives were maintained in the presence of ampicillin (200 μg/mL) and tetracycline (10 μg/mL).

Strains carrying pSM102-2 were maintained in the presence of kanamycin (50 μg/mL).

Protein Determination. Protein concentrations were determined by the method of Bradford (1976), using bovine serum albumin as the calibration standard.

Enzyme Assay. Phosphoenolpyruvate mutase activity was determined by using phosphonopyruvate as substrate and measuring the rate of phosphoenolpyruvate (PEP) formation in a coupled assay (Seidel et al., 1988) with pyruvate kinase/ADP and lactate dehydrogenase/NADH, following the A_{340nm}. The assay mixture contained 100 mM triethanolamine hydrochloride buffer, pH 7.5, glycerol (10% v/v), $MgSO_4$ (10 mM), KCl (30 mM), ADP (1.5 mM), NADH (160 μM), pyruvate kinase (10 units), and lactate dehydrogenase (10 units). Alternatively, PEP formation could be observed directly by following the A_{233nm}, for which $\Delta\epsilon = 1500$ at pH 7.5. Spectrophotometric measurements were made on a Uvikon 860 (Kontron Instruments), a Hewlett Packard Model 4582A, or a Perkin-Elmer 554 UV/vis spectrophotometer.

Oligonucleotide Synthesis, DNA Manipulations, and DNA Sequencing. Oligonucleotides were synthesized on a Milli-Gen/Biosearch 7500 DNA synthesizer (Millipore Corp., Bedford, MA). Site-directed mutagenesis was performed by the method of Eckstein (Taylor et al., 1985a,b; Nakamaye & Eckstein, 1986; Sayers et al., 1988), using materials from Amersham Corp. (Arlington Heights, IL). Supercoiled template DNA was prepared by the "boiling" method (Del Sal et al., 1989). DNA sequencing was performed using the dideoxy chain termination method (Sanger et al., 1977). Standard DNA manipulations were performed according to the methods described in Sambrook et al. (1989).

Isolation of Tetrahymena mRNA and Synthesis of First-Strand cDNA. Total cellular RNA was isolated from harvested *T. pyriformis* cells by extraction with guanidinium isothiocyanate as described by Chomczynski and Sacchi (1987). Polyadenylated mRNA was purified by two successive passes through an oligo(dT)–cellulose spin column (Clontech, Palo Alto, CA). Purified *Tetrahymena* polyadenylated template mRNA (2.5 μg) and an oligo(T) adaptor/primer (1 μg, primer 01) were used to synthesize first strand cDNA using the RiboClone method (Promega, Madison, WI). To degrade the RNA from the RNA–DNA hybrid, NaOH (2 M) was added to a final concentration of 40 mM and the mixture was incubated for 30 min at 65 °C. The solution was adjusted to pH 8.0 with HCl (0.1 N) and then desalted using a Centricon-10 microconcentrator (Amicon, Beverly, MA).

Purification of Genomic DNA from Tetrahymena. Nuclei were isolated from harvested *T. pyriformis* cells as described by Gorovsky (1970). DNA was isolated from nuclei as described by Bannon et al. (1983), except that the NETS buffer used contained EDTA (0.5 M), Tris-HCl (10 mM), pH 9.5, and sodium dodecyl sulfate (1% w/v).

Polymerase Chain Reaction. The reactions were performed using *Taq* DNA polymerase (Promega), *Taq* DNA polymerase buffer (Promega), all four dNTP's (each at 1 mM), priming oligonucleotides (each at 0.5 μM), and template DNA (either *Tetrahymena* genomic DNA or *Tetrahymena* first-strand cDNA). Amplification reactions were carried out in a total volume of 100 μL overlaid with light mineral oil (50 μL). These automated reactions were carried out in a programmable thermal cycler (PTC-100; MJ Research, Cambridge, MA).

Verifying the cDNA-Derived PCR Clones. The PCR-amplified fragments were subcloned into pBS(+) and used to transform *E. coli* strain DH5α. Double-stranded template DNA was isolated from transformants and sequenced.

Genomic Sequence. A procedure identical to that described above for the cDNA PCR clones (using primers 04 and 12) was used to sequence the PCR-amplified fragment from genomic DNA, generating pHMS8a from one PCR reaction and pHMS8b from an independent PCR reaction. Two additional oligonucleotide primers (13, 14) were synthesized in order to sequence the large (527-bp) intron.

Changing the Two Amber Codons to Gln Codons Using Site-Directed Mutagenesis. One of the cDNA clones isolated as described above (pHMS1h) was found by comparison to the consensus sequence (see Discussion) to be free of PCR-induced errors. The gene was cloned into M13mp18, generating pHMS2. Single-stranded DNA was isolated after transformation of *E. coli* strain TG1, and the desired one-base changes were accomplished in two successive rounds of mutagenesis. The oligonucleotides used were 15 and 16. After the two changes were verified by single-stranded sequencing, the gene was subcloned back into pBS(+), generating pHMS3.

Southern Blotting and Hybridization. Samples of *Tetrahymena* genomic DNA (10 μg) that had been digested to completion with a restriction endonuclease (*Bgl*II, *Bcl*I, *Sal*I, *Nde*I, or *Eco*RI) were subjected to electrophoresis in a 0.7% agarose gel. The DNA was transferred to a GeneScreen nylon membrane (NEN Research Products, Boston, MA) by the alkaline transfer method. The oligonucleotide probe was 5′-end-labeled with ^{32}P (>5000 Ci/mmol) by using T4 polynucleotide kinase (Gibco BRL) and [γ-^{32}P]ATP (Amersham). The labeled oligonucleotides were used following spin column purification (Boehringer Mannheim Biochemicals, Indianapolis, IN), and the filters were probed as directed by the manufacturer.

Subgenomic Library Construction. Genomic DNA from *T. pyriformis* was digested to completion with *Bgl*II and fractionated in a 1% agarose gel. DNA fragments of 1–1.4 kB were excised and purified using GeneClean (Bio 101 Inc., La Jolla, CA). These fragments were then ligated into pUC18, and supercompetent *E. coli* strain DH5αMCR (Gibco BRL) cells were transformed directly with the ligation mixtures.

Library Screening by Colony Hybridization. Transformation mixtures were plated onto 150-mm plates at a colony density of 8000 per plate. Multiple nitrocellulose filter replicas were prepared and probed as described by Woods (1984).

Subcloning. (A) Assembling the Full-Length Phosphoenolpyruvate Mutase Gene. Supercoiled pHMS6, containing the *Bgl*II fragment with the true 5′-end of the gene for phosphoenolpyruvate mutase, was digested with *Kpn*I and *Nco*I restriction endonucleases, and the *Kpn*I–*Nco*I fragment was excised from a 0.9% low-melt agarose gel. Supercoiled pHMS3, containing the 5′-truncated cDNA clone with the two amber codons changed to CAG, was cut with restriction endonucleases *Kpn*I and *Nco*I, and the large vector fragment was excised from a 0.9% low-melt agarose gel. The two complementary fragments were ligated together, generating pHMS7, which was sequenced across the entire open reading frame of the PEP mutase gene to eliminate the possibility of adventitious mutations.

(B) Subcloning of the Full-Length Phosphoenolpyruvate Mutase Gene into T7pT5T. The polymerase chain reaction was performed with *Taq* DNA polymerase as described above. Three cycles of amplification were done using pHMS7 (2.0 μg) as template. Eight reaction mixtures (each of 100 μL) were combined to allow visualization of the amplified product on an agarose gel. After digestion with *Bam*HI and *Spe*I restriction endonucleases, the fragment was excised from a

0.9% low-melt agarose gel and ligated with the equivalently-digested vector fragment of T7pT5T. The cloned insert in the resulting plasmid (pT5T-PPM1) was then sequenced.

Comparative Analysis of Phosphoenolpyruvate Mutase Gene Expression in the Presence and Absence of pSM102-2 (dnaY). Competent *E. coli* strain BL21 (DE3) cells were transformed with pT5T-PPM1 and plated on LB plates containing ampicillin (200 μg/mL) and tetracycline (10 μg/mL). *E. coli* strain BL21 (DE3) cells were separately cotransformed with pT5T-PPM1 and pSM102-2 and plated on LB agar plates containing ampicillin (200 μg/mL), tetracycline (10 μg/mL), and kanamycin (50 μg/mL). Small cultures (10 mL) of transformants from each of these experiments were then grown and induced at A_{550nm} = 1 as described above. Whole-cell extracts were prepared by pelleting the cells in a microcentrifuge for 1 min. The cells were taken up in protein sample buffer (Sambrook et al., 1989), and the resulting suspension was boiled for 10 min. Samples were then analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis by the method of Laemmli (1970).

Purification of Recombinant Full-Length PEP Mutase from E. coli BL21 (DE3)(pSM102-2)(pT5T-PPM1). Cell paste (0.3 g) of *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM1) was resuspended in ice-cold buffer A (5 mL) [100 mM triethanolamine hydrochloride buffer, pH 7.5, containing glycerol (20% v/v), dithiothreitol (1 mM) supplemented with aprotinin (1 μg/mL) (Boehringer Mannheim), and phenylmethanesulfonyl fluoride (1 mM, 50 μL of a 100 mM solution in acetone)]. Cells were lysed ultrasonically by using a Branson Model 450 sonifier (Branson Electronics, Danbury, CT) in four 1-min bursts at 37 W. Cell debris was removed by centrifugation at 200000*g* for 105 min. To the supernatant was added protamine sulfate (47 mg, Sigma, grade X) in buffer A (0.95 mL) with stirring. The mixture was stirred for 15 min at 4 °C and then centrifuged (9000*g*, 15 min). The supernatant was diluted to 15 mL with buffer A and loaded onto a column (16 mL) of hydroxylapatite (Bio-Gel HTP; Bio-Rad Laboratories, Richmond, CA) that had been equilibrated in buffer B [25 mM potassium phosphate buffer, pH 7.5, containing glycerol (10% v/v)]. The column was washed with buffer B (16 mL) and then eluted with a linear gradient (350 mL + 350 mL) from buffer C [100 mM potassium phosphate buffer, pH 7.5, containing glycerol (10% v/v) and dithiothreitol (0.5 mM)] to buffer D [500 mM potassium phosphate buffer, pH 7.5, containing glycerol (10% v/v) and dithiothreitol (0.5 mM)]. Fractions containing PEP mutase activity were pooled and concentrated by ultrafiltration with PM10 membranes (Amicon). The buffer was changed to buffer E [10 mM triethanolamine acetate, pH 8.0, containing glycerol (10% v/v)] by ultrafiltration, and the resulting solution was loaded onto an FPLC HR 5/5 column of Mono Q resin (Pharmacia LKB Biotechnology) preequilibrated with buffer E. The column was washed with buffer E (5 min, 1 mL/min) and eluted with a linear gradient (30 min, 1 mL/min) from 100% buffer E to 30% buffer E/70% buffer F [50 mM triethanolamine acetate, pH 8.0, containing glycerol (10% v/v) and sodium acetate (2 M)]. Fractions containing PEP mutase activity were pooled and concentrated by ultrafiltration. The buffer was changed to buffer A for storage.

Subcloning of 5′-Truncated Phosphoenolpyruvate Mutase. Subcloning by PCR was performed as described above for the full-length gene, but with oligonucleotide 20 as the 5′-primer. This primer introduced an ATG initiator codon just upstream of codon 10 (serine). This expression plasmid was designated pT5T-PPM2.

Purification of Recombinant 5′-Truncated PEP Mutase from E. coli BL21 (DE3)(pSM102-2)(pT5T-PPM2). Cell paste (38 g) of *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM2) was processed exactly as described for the recombinant full-length PEP mutase (buffer A, 200 mL; protamine sulfate, 2.4 g in 25 mL of buffer A). The supernatant after protamine sulfate precipitation was diluted to 300 mL with buffer A and loaded onto a column (100 mL) of hydroxylapatite (Bio-Gel HTP) that had been equilibrated in buffer B. The column was washed with buffer B (100 mL) and then eluted with a linear gradient (1.5 L + 1.5 L) from buffer C to buffer D. Fractions containing PEP mutase activity were pooled and concentrated by ultrafiltration with PM10 membranes (Amicon). The buffer was changed to buffer A for storage.

Purification of PEP Mutase from Tetrahymena pyriformis in the Presence of Protease Inhibitors. Tetrahymena cells (25 g) were taken up in buffer A (100 mL) containing aprotinin (1 μg/mL), pepstatin (1 μg/mL) (Boehringer Mannheim), and phenylmethanesulfonyl fluoride (1 mM, 1.0 mL of a 100 mM solution in acetone). Cells were lysed ultrasonically by using a Branson Model 450 sonifier in four 2-min bursts at 37 W. Cell debris was removed by centrifugation at 200000*g* for 105 min. To the supernatant was added protamine sulfate (0.75 g) in buffer A (10 mL) with stirring. The mixture was stirred for 15 min at 4 °C and then centrifuged (9000*g*, 15 min). The supernatant was diluted to 150 mL with buffer A and loaded onto a column (150 mL) of hydroxylapatite that had been equilibrated in buffer B. The column was washed with buffer B and then eluted with a linear gradient (1 L + 1 L) from buffer C to buffer D. Fractions containing PEP mutase activity were pooled and concentrated by ultrafiltration with PM10 membranes (Amicon). This partially-purified preparation of PEP mutase was then fractionated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, transferred to a ProBlott PVDF membrane (Applied Biosystems) according to the manufacturer's specifications, and stained with Ponceau S (Sigma). There was only one protein band in the 31–35-kDa range on the membrane, and this band was excised and used for N-terminal amino acid sequencing.

Amino Acid Composition and N-Terminal Amino Acid Sequencing. Amino acid analyses by acidic hydrolysis/phenylisothiocyanate derivatization and N-terminal amino acid sequencing by automated Edman degradation were performed by Dr. William S. Lane (Harvard Microchemistry Facility, Cambridge, MA).

Data Base Searches and Protein Sequence Alignments. GenBank and EMBL databases were searched using "WordSearch" and "FASTA" in the GCG software package (Devereux et al., 1984) and the EuGene software package (Lawrence & Goldman, 1988; Pearson & Lipman, 1988).

Results and Discussion

Cloning of the Phosphoenolpyruvate Mutase Gene from Tetrahymena pyriformis. Since *Tetrahymena* is a eukaryote, we anticipated that the coding sequence of the PEP mutase gene could, in genomic DNA, be interrupted by the presence of introns. Indeed, although there were no reports of the successful *expression* of a *Tetrahymena* gene in a prokaryotic host, it was known from sequence comparison studies that several *Tetrahymena* genes do contain segments of intronic DNA. It was therefore important to obtain a *Tetrahymena* cDNA library that contained full-length copies of the mutase gene. To this end, polyadenylated mRNA was isolated from log-phase *Tetrahymena* cells, and a first-strand cDNA library was synthesized using reverse transcriptase and oligonucleotide 01 (oligo(T) that carries a restriction enzyme recognition site)

as primer. To utilize the polymerase chain reaction (Oste, 1988; Mullis & Faloona, 1987) to amplify the PEP mutase gene coding sequence from the *Tetrahymena* first-strand cDNA library, a primer for the 5′-end of the gene was needed.

Two mixed oligonucleotides, containing 33 and 37 bases, respectively, were synthesized to correspond to each end (amino acids 1–8 and amino acids 22–30; additional bases on the 5′-end of each oligonucleotide carried convenient restriction sites) of the known N-terminal amino acid sequence (Seidel et al., 1988). These oligonucleotides, 02 and 03, were designed to prime DNA synthesis toward each other across a 39-bp region near the 5′-end of the PEP mutase gene. The most probable base, chosen on the basis of the highly biased *Tetrahymena* codon usage (Martindale, 1989), was incorporated at each position, though at the 3′-ends of each oligonucleotide mixtures of bases were incorporated at positions of uncertainty. PCR was performed directly on genomic DNA isolated from *T. pyriformis* GL. The major amplified product was a 109-bp DNA fragment of exactly the expected size, and this was cloned into pBS(+) as a *Pst*I–*Bam*HI fragment, generating pDLP1. The nucleotide sequence of the insert corresponded precisely with the N-terminal amino acid sequence of the protein. The 39-bp sequence between the two primers 02 and 03 was thus determined exactly.

Using oligonucleotide 04 (a primer specific to the 5′-end of the exact sequence determined above and including a best guess nucleotide sequence for the eight N-terminal amino acids known from the isolated protein) together with oligonucleotide 01 (an adaptor/primer complementary to the 3′-end of the cDNA), PCR amplification of the first-strand cDNA library yielded a DNA fragment of length 1.3 kb. For this step to be successful, it was essential first to degrade the mRNA of the mRNA–cDNA hybrid. Because primers 01 and 04 each included a unique restriction site, the amplified 1.3-kb fragment was easily cloned into pBS(+), providing pDLP2, which thus facilitated the sequencing of the gene fragment.

The two oligonucleotides 01 and 04 were used as sequencing primers in double-strand sequencing of supercoiled pDLP1. The resulting sequence information was used sequentially to define new primers (05, 06, 07, 08, 09, 10, 11), until the entire insert sequence had been determined. Because *Taq* polymerase has a relatively high error rate (Keohavong & Thilly, 1989), it was important to sequence clones derived from several independent PCR amplification reactions to find the consensus sequence. Eight independent clones (pHMS1a–h) from three independent amplifications (using two different first-strand cDNA libraries to control for reverse transcriptase errors) were sequenced. As an additional check, the entire genomic sequence was amplified by PCR using primers specific to the known 5′- and 3′-ends of the cDNA clone's coding region (04, 12). When this 1.9-kb genomic fragment was sequenced, three introns were found. The coding region was fully consistent with that obtained from the cDNA libraries described above. The nucleotide sequence and the derived amino acid sequence of the presumed open reading frame are presented in Figure 1. *Tetrahymena* uses only one of the three "universal" stop codons (opal, TGA) as a terminator, and it uses amber (TAG) and ochre (TAA) to code for glutamine (Martindale, 1989). The two in-frame amber codons in the PEP mutase gene therefore code for glutamine. On this basis, the PEP mutase gene appears to encode a protein of M_r 32 688, having 290 residues. This is in good agreement with the M_r of 33 000 that was estimated for the enzyme isolated from *Tetrahymena* previously (Seidel et al., 1988), though it is less close to the value of 38 kDa estimated by Bowman et al. (1988) for their protein. The derived sequence of the N-terminal 30 amino acid residues is in exact agreement with that reported for the purified protein from *T. pyriformis* (Seidel et al., 1988), and the codon usage is biased in the manner typical for *Tetrahymena* genes (Martindale, 1989).

Removal of the Two Amber Codons. Before expression of the PEP mutase gene in *E. coli* could be attempted, the two amber codons had to be changed to a codon that would be recognized in *E. coli* as glutamine. The simplest change was from TAG to CAG. Thus, one of the cDNA clones that had no *Taq* polymerase-induced errors (pHMS1h) was subcloned as a *Bam*HI–*Sal*I fragment into rf form M13 DNA. Single-stranded DNA was prepared from this M13 construct (pHMS2), and two rounds of site-directed mutagenesis were performed using oligonucleotides 15 and 16 to introduce the two desired changes. The *Bam*HI–*Sal*I fragment, containing the two TAG–CAG changes, was then subcloned back into pBS(+), generating pHMS3.

At this point, it was unclear whether we had cloned the entire PEP mutase gene, since the protein we had isolated previously did not have an N-terminal methionine. It was therefore necessary to determine the 5′-nucleotide sequence upstream from the cloned portion of the mutase gene. To this end, an oligonucleotide (17) complementary to the 5′-end of the gene was used to probe genomic DNA (Southern, 1975) that had been digested with various restriction endonucleases. This Southern blot analysis showed the 5′-end of the mutase gene to be on a 1.2-kb *Bgl*II fragment of *T. pyriformis* genomic DNA. A subgenomic library containing 1–1.4-kb *Bgl*II fragments of *T. pyriformis* genomic DNA was constructed in the vector pUC18, and this library was probed with oligonucleotide 17. Of the 24 000 colonies screened, three independent positive colonies were detected (pHMS4–6), each of which had an identical sequence in the region of interest. One plasmid, pHMS6, contained a single 1.2-kb insert and was sequenced entirely. This plasmid carried the 5′-end of the gene through the first intron (720 bp) as well as 480 bp upstream of the ATG start codon. From these data, the true 5′-end of the PEP mutase gene was found to have ten more codons upstream from the codon corresponding to the first amino acid defined by N-terminal sequencing (Figure 2). The sequence also verified the portion of the N-terminal sequence encompassed by the first PCR probe (02). A *Kpn*I site in the polylinker of pUC18 and the unique *Nco*I site near the 5′-end of mutase gene (Figure 1) were used to join the new upstream portion of the gene with the cDNA clone in pHMS3, generating pHMS7.

Full Genomic Sequence of the Phosphoenolpyruvate Mutase Gene. The genomic sequence contained within the open reading frame (i.e., including introns) was obtained in a similar manner to that used to obtain the cDNA sequence and is shown in Figure 2. The sequence upstream of the initiator ATG codon was obtained from pHMS6. Immediately upstream of the initiation ATG codon is a characteristic AAA triplet that has been found in virtually every *Tetrahymena* gene reported (Brunk & Sadler, 1990). However, unlike other *Tetrahymena* promoter regions, there is no "CCAT" box located upstream of the presumed translational starting point (Brunk & Sadler, 1990). The gene contains three introns, two that start *precisely* after an amber codon (TAG), and one that begins after a lysine codon (AAG). The A-T content of the 5′- and 3′-untranslated regions and the introns is 73% [consistent with the overall A-T richness (75%) of the *Tetrahymena* genome (Gorovsky, 1980)], while that of the coding sequence is 53%, again consistent with the other *Tetrahymena* genes

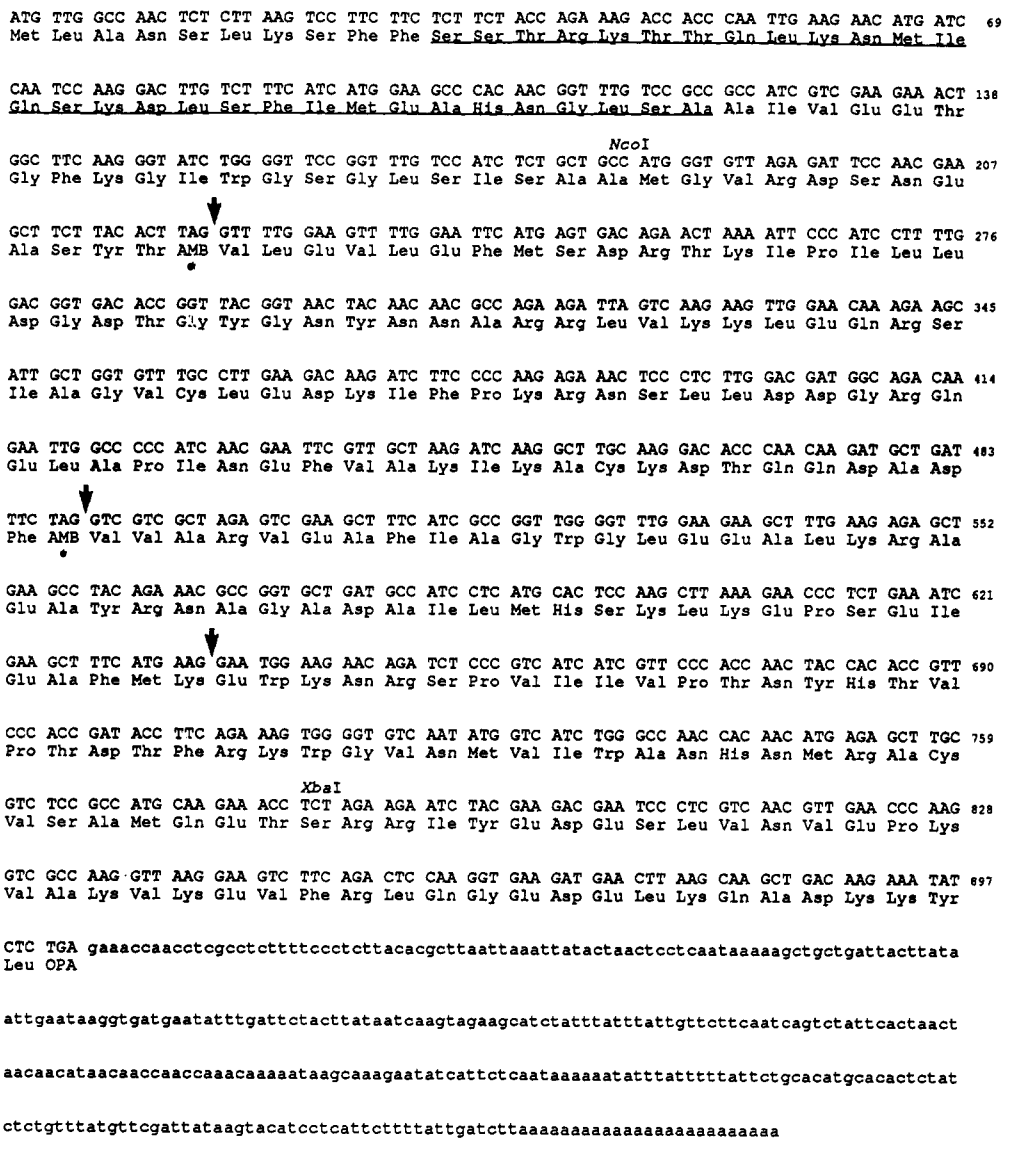


FIGURE 1: The cDNA gene sequence of PEP mutase, including the 3′-untranslated region and the 5′-terminus determined by colony hybridization (see text). Underlined amino acids correspond to the N-terminal amino acid sequence determined previously (Seidel et al., 1988). Asterisks mark the locations of the two amber codons that code for Gln in *Tetrahymena*. Arrows mark the locations of introns (see Figure 2).

reported to date [e.g., Martindale (1989)]. The base sequences around each of the three pairs of exon–intron–exon junctions conform to the consensus sequences for the exon–intron junctions (of ...AG/GTA...) and the intron–exon junctions (of ...AG/G...) from all eukaryotes (Shapiro & Senepathy, 1987). In fact, 12 out of the 17 known *Tetrahymena* introns end with the amber triplet, TAG, and of the remaining five, four end with AAG and one with ends CAG (Martindale & Taylor, 1988). These correspondences make a seductive case for the view that exons (and, indeed, introns) were "microgenes", originally terminating with an amber codon, which encoded relatively short oligopeptides that spontaneously assembled into active protein. This hypothesis is discussed more fully elsewhere (Seidel et al., submitted for publication).

Subcloning of the Phosphoenolpyruvate Mutase Gene into the Expression Vector T7pT5T. The vector chosen for the overexpression of the mutase was T7pT5T (Squires et al., 1988; Eisenberg et al., 1990), a derivative of pET3a (Rosenberg et al., 1987). The pET series of vectors uses an appropriate *E. coli* host that carries a chromosomal copy of T7-RNA polymerase under *lac* control to direct the synthesis of a target gene located behind the T7 promoter. The most common strain used for expression is *E. coli* BL21(DE3), a strain that, among its other features, lacks the *lon* protease (Studier & Moffatt, 1986). After induction, T7-RNA polymerase rapidly overwhelms the cellular translational machinery with transcripts from the plasmid, and the target gene is translated almost exclusively. The plasmid T7pT5T has been modified to couple the translation of the target gene to the synthesis of a short sequence of gene 10 of phage T7 and carries both an ampicillin and tetracycline resistance marker. The polymerase chain reaction was used to excise the open reading frame from double-stranded pHMS7 and simultaneously to introduce the necessary 5′- and 3′-sequences to enable subcloning into T7pT5T. The primers for the amplification introduced the requisite 5′-coupler sequence and *Bam*HI site (Eisenberg et al., 1990) upstream of the initiation codon (oligonucleotide 18) and a 3′-*Spe*I site downstream from the stop codon of the phosphoenolpyruvate mutase gene (oligonucleotide 19). The PCR amplification was run with *Taq* polymerase, and to minimize *Taq* polymerase-induced errors (Keohavong & Thilly, 1989), a high initial concentration of supercoiled pHMS7 was used and only three cycles of amplification were performed. The major amplified product was of the correct size (950 bp) and was digested with *Bam*HI and *Spe*I restriction endonucleases before ligating into the ap-

```
  BglII
  agatctacatgaatcggatgcatagtctttctcacaagaaataaaagatgaataaagcaatacaggaaataaaacaatcaatgatagca

cgattgagacgaaagaaaaaacaaataaattcaatttgtatctatttattattatttatttcaatttttattaattttcacccttctctc

tcaattatctctgactgattgatcagccttcattcattttgttacttttttatgtgtgacaattacgatttgcaaagagaaagagattgat

tgactggaggtgaaatcaatcaatcattatcttaatagatcaatcaatcaatatcttctaagtagtattttcaaaacttctcccaacagct

tctttaaatggaatcaaatcaaatcaaaaaaaatataattaaaaaatcaaaacaaacaaatagtttaaataaataaacaaatatctagcaa

aaacacaatcacatctaataaa  ATG TTG GCC AAC TCT CTT AAG TCC TTC TTC TCT TCT ACC AGA AAG ACC ACC
                        Met Leu Ala Asn Ser Leu Lys Ser Phe Phe Ser Ser Thr Arg Lys Thr Thr

CAA TTG AAG AAC ATG ATC CAA TCC AAG GAC TTG TCT TTC ATC ATG GAA GCC CAC AAC GGT TTG TCC GCC
Gln Leu Lys Asn Met Ile Gln Ser Lys Asp Leu Ser Phe Ile Met Glu Ala His Asn Gly Leu Ser Ala

GCC ATC GTC GAA GAA ACT GGC TTC AAG GGT ATC TGG GGT TCC GGT TTG TCC ATC TCT GCT GCC ATG GGT
Ala Ile Val Glu Glu Thr Gly Phe Lys Gly Ile Trp Gly Ser Gly Leu Ser Ile Ser Ala Ala Met Gly

GTT AGA GAT TCC AAC GAA GCT TCT TAC ACT TAG gtacttgatcttctaccattatttttcccttattctgcatctctgtc
Val Arg Asp Ser Asn Glu Ala Ser Tyr Thr AMB
                                          *

ttcttctctttcagaatcactgcttctctcaaaataactccattaatccatcatgaaaaagctgaatttgttttggatttttacattttt

gcaagagtataacacacacactaacacactcttgtgaaatactactactgctacttacttactttcaaagcaaaaaagataagaagaata

atcaagaaaagagagaaagatattcagattttataaagagagacagagagagttattgattcagtgattgagaatgacaagataaaataca

tctaaatgaattggtttaattgaactttatacttaacatacaaatag GTT TTG GAA GTT TTG GAA TTC ATG AGT GAC AGA
                                                Val Leu Glu Val Leu Glu Phe Met Ser Asp Arg

ACT AAA ATT CCC ATC CTT TTG GAC GGT GAC ACC GGT TAC GGT AAC TAC AAC AAC GCC AGA AGA TTA GTC
Thr Lys Ile Pro Ile Leu Leu Asp Gly Asp Thr Gly Tyr Gly Asn Tyr Asn Asn Ala Arg Arg Leu Val
                                                                    BglII
AAG AAG TTG GAA CAA AGA AGC ATT GCT GGT GTT TGC CTT GAA GAC AAG ATC TTC CCC AAG AGA AAC TCC
Lys Lys Leu Glu Gln Arg Ser Ile Ala Gly Val Cys Leu Glu Asp Lys Ile Phe Pro Lys Arg Asn Ser

CTC TTG GAC GAT GGC AGA CAA GAA TTG GCC CCC ATC AAC GAA TTC GTT GCT AAG ATC AAG GCT TGC AAG
Leu Leu Asp Asp Gly Arg Gln Glu Leu Ala Pro Ile Asn Glu Phe Val Ala Lys Ile Lys Ala Cys Lys

GAC ACC CAA CAA GAT GCT GAT TTC TAG  gtaagcttattccctttatttccttttatctccttccattgattggcgattcca
Asp Thr Gln Gln Asp Ala Asp Phe AMB
                                   *

gaggaaataagaactatttctcatctcttattttcaaaatacaatag GTC GTC GCT AGA GTC GAA GCT TTC ATC GCC GGT
                                                Val Val Ala Arg Val Glu Ala Phe Ile Ala Gly

TGG GGT TTG GAA GAA GCT TTG AAG AGA GCT GAA GCC TAC AGA AAC GCC GGT GCT GAT GCC ATC CTC ATG
Trp Gly Leu Glu Glu Ala Leu Lys Arg Ala Glu Ala Tyr Arg Asn Ala Gly Ala Asp Ala Ile Leu Met

CAC TCC AAG CTT AAA GAA CCC TCT GAA ATC GAA GCT TTC ATG AAG  gtaaagaaattatttcagcatattaaaaca
His Ser Lys Leu Lys Glu Pro Ser Glu Ile Glu Ala Phe Met Lys

ggatttctaaacaaaaaaacaaacaaatcgattaacatattgaatgaaagaagaagaaatagttagctataataaggatatttagagaaat

agttctatgaacaaaagaacgaactatacttaatttgtagagtttatcagattaagaagaagattagggattaaataatactttatcagca

agcaagaaagcaaacaagcaaggcaaacaaaaaagaaagaaagttattttggtcaaattaaaaaattaattgtcattcagtttttagtaga

aaccaattggcattcattctaatcaaataatcaaagaaacagtggatatcataaaagcagttttgaaaataaatcgaaataaaagaatgtg
```

```
taagatacaatatatccttacaatgataaaagatataaattaatagttcatcagtttgttgtttacaaatcaatcataatcaatcgatata

                                                    BglII
ttcttccatattaaatttcatttgttattataaag GAA TGG AAG AAC AGA TCT CCC GTC ATC ATC GTT CCC ACC AAC
                                    Glu Trp Lys Asn Arg Ser Pro Val Ile Ile Val Pro Thr Asn

TAC CAC ACC GTT CCC ACC GAT ACC TTC AGA AAG TGG GGT GTC AAT ATG GTC ATC TGG GCC AAC CAC AAC
Tyr His Thr Val Pro Thr Asp Thr Phe Arg Lys Trp Gly Val Asn Met Val Ile Trp Ala Asn His Asn

ATG AGA GCT TGC GTC TCC GCC ATG CAA GAA ACC TCT AGA AGA ATC TAC GAA GAC GAA TCC CTC GTC AAC
Met Arg Ala Cys Val Ser Ala Met Gln Glu Thr Ser Arg Arg Ile Tyr Glu Asp Glu Ser Leu Val Asn

GTT GAA CCC AAG GTC GCC AAG GTT AAG GAA GTC TTC AGA CTC CAA GGT GAA GAT GAA CTT AAG CAA GCT
Val Glu Pro Lys Val Ala Lys Val Lys Glu Val Phe Arg Leu Gln Gly Glu Asp Glu Leu Lys Gln Ala

GAC AAG AAA TAT CTC TGA gaaaccaacctcgcctcttttccctcttacacgcttaattaa
Asp Lys Lys Tyr Leu OPA
```

FIGURE 2: Genomic sequence of the PEP mutase gene and the 5′-upstream region. The sequence between the first two *Bgl*II sites was determined from a hybridization-positive clone from a subgenomic library (see text). The underlined amino acid sequence was determined previously by N-terminal amino acid sequencing. The coding sequence is in upper-case letters; the noncoding/intron sequence is in lower-case letters. Asterisks mark the locations of the two amber codons that code for Gln in *Tetrahymena*.

Table I: Purification of Full-Length PEP mutase from *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM1)

	total catalytic act. (μmol min^{-1})	unit yield (%)	sp catalytic act. (μmol min^{-1} mg^{-1})	purification factor	total protein (mg)
crude lysate	nd	nd	nd	nd	nd
protamine sulfate	270	100	1.7	1	160[a]
hydroxylapatite	190	70	17	10	11
Mono Q	180	65	113	63	1.6

[a] Includes unprecipitated protamine sulfate.

propriate fragment of T7pT5T, creating pHMS8. To minimize further the chances of polymerase-induced error, the middle portion (600 bp) of the mutase gene was excised as a *Nco*I–*Xba*I fragment and replaced with the same fragment derived from pHMS7. The insert of the resulting vector (pT5T-PPM1) was sequenced to verify that no changes had been introduced during the subcloning. The plasmid pT5T-PPM1 is stable in *E. coli* strains DH5α and BL21(DE3).

Overexpression of the Phosphoenolpyruvate Mutase Gene in E. coli BL21(DE3) and the Effect of dnaY on Expression. While the extremely biased codon usage by *Tetrahymena* (Martindale, 1989) proved to be very useful in the design of the "guessmers" 02 and 03, unusual codon usage can, of course, hinder or even prevent efficient expression in a heterologous host (Ikemura, 1981; Robinson et al., 1984; Bonekamp & Jensen, 1988). This problem is known to be acute when the target gene is rich in the arginine codons AGA and AGG, which are used very rarely by *E. coli* (Brinkmann et al., 1989). In addition, tRNA^{Arg}(AGA/AGG) is thought to be involved in chromosomal as well as plasmid replication, and a dearth of this tRNA species has been associated with poor cell viability and plasmid instability (Brinkmann et al., 1989). In fact, the phosphoenolpyruvate mutase gene contains 17 AGA codons (AGA is the only arginine codon used by *Tetrahymena*) including two pairs of *adjacent* AGA codons [which is a particularly strong translational stumbling block, see Bonekamp and Jensen (1988)], and AGA represents nearly 6% of all the codons in the mutase open reading frame. In *E. coli*, tRNA^{Arg}(AGA/AGG) is encoded by the *dnaY* gene (Garcia et al., 1986). Expression of the mutase gene in *E. coli* BL21 (DE3)(pT5T-PPM1) was therefore studied both with and without cotransformation by pSM102-2 (which carries the *dnaY* gene). In both cases, the PEP mutase gene product was visible above the background of *E. coli* cellular proteins, but the expression of the mutase gene was at least four times greater when the host was cotransformed with pSM102-2. Indeed, in the latter case, phosphoenolpyruvate mutase activity was obtained at very high levels, corresponding to about 5 mg/L of culture. Although a significant portion of the PEP mutase gene product was insoluble, it appears that, on the basis of the specific catalytic activity of the purified enzyme, transformants of *E. coli* BL21(DE3)(pSM102-2) produce *Tetrahymena* phosphoenolpyruvate mutase at about 2.5% of the *soluble* cell protein.

The purification of the mutase from *E. coli* BL21(DE3)-(pT5T-PPM1)(pSM102-2) is summarized in Table I. The recombinant protein displayed slightly different chromatographic behavior from that of the protein originally isolated from *T. pyriformis* (Seidel et al., 1988). Notably, the recombinant protein eluted earlier from hydroxylapatite and was not retained appreciably by anion-exchange resins. The reason for these differences is discussed below. The recombinant protein follows Michaelis–Menten kinetics, with a K_m of 65 μM and a k_{cat} of 60 s^{-1}. These values compare well with the K_m of 67 μM and k_{cat} of 41 s^{-1} determined for the isolated *T. pyriformis* enzyme (Seidel et al., 1988).

Subcloning and Overexpression of 5′-Truncated Phosphoenolpyruvate Mutase Gene. To determine whether the form of PEP mutase that is isolated from *T. pyriformis* (truncated by 10 residues from the N-terminus and starting with Ser-Ser-Thr-Arg-Lys-; see Figures 1 and 2) differs significantly from the full-length gene product, the truncated form of the mutase gene was subcloned into T7pT5T in a manner analogous to that described above for the full-length mutase gene. Oligonucleotide primer 20 introduced the 5′-coupler sequence, a *Bam*HI site, and an initiator methionine codon just upstream of the serine codon corresponding to the first amino acid of the isolated protein (Seidel et al., 1988). The plasmid containing the construct corresponding to the truncated form of the phosphoenolpyruvate mutase gene (PEP

Table II: Purification of Truncated PEP Mutase from *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM2)

	total catalytic act. ($\times 10^{-4}$) (μmol min^{-1})	unit yield (%)	sp catalytic act. (μmol min^{-1} mg^{-1})	purification factor	total protein (mg)
crude lysate	9.6	100	42	1	2300
protamine sulfate	9.3	97	22	0.54	4200[a]
hydroxylapatite	7.6	79	155	3.8	490

[a] Includes unprecipitated protamine sulfate.

```
PEP Mutase      MLANSLKSFFSSTRKTTQLKNMIQSKDLSFIMEAHNGLSAAIVEETGFKGI  51
                          |•  ••••••• ••   • •|•••||| •••••|| ••
CPEP Mutase               MAVTKARTFRELMNAPEILVVPSAYDALSAKVIQQAGFPAV  41

PEP Mutase      WGSGLSISAAM-GVRDSNEASYTQVLEVLEFMSDRTKIPILLDGDTGYGNY 101
                  •| ••|||•| |••| • •| ••     |• •   •••|•••|•|•|||||
CPEP Mutase     HMTGSGTSASMLGLPDLGFTSVSEQAINLKNIVLTVDVPVIMDADAGYGNA  92

PEP Mutase      NNARRLVKKLEQRSIAGVCLEDKIFPKRNSLLDDGRQELAPINEFVAKIKA 152
                 •••| ••••|• •|•|  |||•• ||| • |•• |  | •••|•••||•|
CPEP Mutase     MSVWRATREFERVGIVGYHLEDQVNPKRCGHLEGKR--LISTEEMTGKIEA 141

PEP Mutase      CKDTQQDADFQVVARVEAFIAGWGLEEALKRAEAYRNAGADAILMHSKL--- 202
                  ••••|•|| ••|||••|  •••||•||••|• •| •|||| |•••••|
CPEP Mutase     AVEAREDEDFTIIARTDA-RESFGLDEAIRRSREYVAAGADCIFLEAML--- 190
```

FIGURE 3: Sequence comparison of PEP mutase (from *Tetrahymena*) and CPEP mutase (from *Streptomyces*) using the FASTA program. Identity is signified by the | symbol; neutral or conservative substitutions are signified by the · symbol. Gaps are indicated by dashes (-).

mutase2) was designated pT5T-PPM2.

The construct pT5T-PPM2 gives very high levels of PEP mutase activity (50 mg/L of culture) in *E. coli* BL21 (DE3)(pSM102-2), and the PEP mutase2 gene product partitions entirely into the soluble phase of crude *E. coli* extracts. On the basis of the specific catalytic activity of the purified PEP mutase2 gene product, transformants of *E. coli* BL21 (DE3)(pSM102-2) with pT5T-PPM2 produce phosphoenolpyruvate mutase at about 25% of soluble cell protein. The purification from *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM2) is summarized in Table II. This method represents an improvement of about 1000-fold in the yield of phosphoenolpyruvate mutase as compared to the purification from *Tetrahymena pyriformis* (Seidel et al., 1988). The truncated mutase behaves identically to the enzyme originally isolated from *T. pyriformis*, and it follows Michaelis–Menten kinetics with a K_m of 63 μM and a k_{cat} of 77 s^{-1}. These values compare well with the K_m of 67 μM and k_{cat} of 41 s^{-1} determined for the *T. pyriformis* enzyme isolated originally (Seidel et al., 1988).

Comparison of the Phosphoenolpyruvate Mutase and Phosphoenolpyruvate Mutase2 Gene Products. As mentioned above, the two recombinant mutases have somewhat different chromatographic and solubility characteristics, though the steady-state kinetic parameters for the two proteins are very similar. When the two purified recombinant gene products were subjected to N-terminal amino acid analysis, the full-length PEP mutase gene product gave a heterogeneous mixture of N-terminal sequences dominated by Ser-Leu-Lys-Ser-Phe-, while the truncated PEP mutase2 gene product yielded a clean N-terminal sequence of Ser-Ser-Thr-Arg-Lys- (Figure 1). It seems that the N-terminal sequence of 10 amino acids is cleanly removed in vivo. This sequence seems likely to be a relatively unstructured tail, which is not specifically recognized in *E. coli* and is subject to the action of intracellular proteases that produce the heterogeneous N-terminal sequence that we have observed. In *Tetrahymena*, however, this N-terminal tail is processed precisely. Whether this tail sequence has some kind of signal function in *Tetrahymena* remains to be seen, but it is clear that the full-length enzyme is cleanly processed in *Tetrahymena* even in the presence of protease inhibitors. Thus, when the preparation of the mutase from *T. pyriformis* was repeated in the presence of a battery of protease inhibitors [no protease inhibitors were included in the original purification of the mutase from which the N-terminal sequence was determined (Seidel et al., 1988)], a clean N-terminal sequence of Ser-Ser-Thr-Arg-Lys- was again found. Unfortunately, too few *Tetrahymena* genes are known for us to discern any analogous feature at the N-terminus of other *Tetrahymena* gene products.

The PEP mutase2 gene product was chosen for use in future studies because of its significantly greater solubility in induced *E. coli* BL21 (DE3)(pSM102-2)(pT5T-PPM2) cells and because it appears to be the in vivo form of PEP mutase in *Tetrahymena*.

Comparison to Other Work on Phosphoenolpyruvate Mutase. Work on phosphoenolpyruvate mutase has been reported from several other laboratories (Bowman et al., 1988, 1990; Hidaka et al., 1988). The characteristics of the *Tetrahymena* enzyme reported by Bowman et al. (1990) differ somewhat from those described here and in our previous work (Seidel et al., 1988). Thus, Bowman et al. (1990) report a K_m of 2 μM and a k_{cat} of 38 s^{-1} for the mutase, in contrast to the K_m of 63 μM and k_{cat} of 77 s^{-1} reported in this work. The difference in K_m cannot be attributed to differences in pH (in both cases the enzyme was assayed at pH 7.5), and the concentration of substrate phosphonopyruvate used in our studies was quantitated in three different ways (see Experimental Procedures). Further, in the amino acid analysis of the mutase of Bowman et al. (1990), the values for 9 of the 20 amino acids deviate by more than 50% from the composition predicted by either the full-length PEP mutase or the PEP mutase2 gene sequences. Bowman et al. (1990) also report a subunit molecular weight of 38 000 and a dimer molecular weight of 81 000. The predicted values from the gene sequence are 33 827 from 300 amino acids (for the full-length PEP mutase)

or 32 819 from 290 amino acids (for the naturally truncated enzyme), which agree well with our original estimate of 33 000 for the isolated protein. We cannot account for these discrepancies, and we have found no evidence for the existence of any PEP mutase isoenzymes in *T. pyriformis* GL.

Sequence Alignments. The derived amino acid sequence of the *T. pyriformis* PEP mutase has significant similarity (Doolittle, 1986) to only one sequence in the GenBank and EMBL databases, that of carboxyphosphonoenolpyruvate mutase from *Streptomyces hygroscopicus* (30.3% identity over a 180-residue overlap). The similarity is illustrated in Figure 3. The enzyme carboxyphosphonoenolpyruvate mutase, an enzyme in the bialaphos biosynthetic pathway, catalyzes a very similar rearrangement to that mediated by phosphoenolpyruvate mutase (Scheme II), and this similarity is therefore not surprising. Interestingly, the best alignment of the two sequences starts precisely with the N-terminal serine found in the *truncated* protein isolated from *T. pyriformis*.

On the basis of the electrical discharge experiments of Rabinowitz et al. (1969) and the belief that the ancient earth had a reducing atmosphere (Mastalerz, 1984), it has been argued that the biosynthesis of phosphonates is an evolutionarily ancient metabolic process. The existence of structural homology (>24.8% identity over a >80-residue overlap implies homology; Sander & Schneider, 1991) between the PEP mutase gene sequences of a Gram-positive bacterium (*Streptomyces*) and that of a eukaryote (*Tetrahymena*), two evolutionarily very distant relatives, provides the most persuasive evidence that the biosynthesis of phosphonates is evolutionarily ancient.

Finally, the motif of -Thr-X-His- found in some PEP-handling enzymes (Pocalyko et al., 1990) and the -Arg-His-Gly- motif found in some enzymes that involve a phosphohistidine intermediate (Bazan et al., 1989) are not present in PEP mutase.

In summary, we report here the cloning and expression in *E. coli* of the phosphoenolpyruvate mutase from *Tetrahymena*. To our knowledge, this is the first *Tetrahymena* protein that has been expressed in a foreign host. The existence of two amber TAG codons in the open reading frame of the *Tetrahymena* mutase gene necessitated appropriate site-directed mutagenesis to change these codons so that they would be recognized as encoding glutamine by the translational machinery of *E. coli*. Further, the presence of extrachromosomal *E. coli dnaY* [that encodes tRNA(AGA)] in the host enables efficient overexpression of the phosphoenolpyruvate mutase gene, despite the exclusive use of the rarely-used arginine AGA codons in the mutase gene.

With large quantities of phosphoenolpyruvate mutase now available, the way is clear for a more detailed examination of this unusual enzyme-catalyzed reaction.

ACKNOWLEDGMENTS

We are most grateful to Maya T. Kunkel and Lorraine C. Santy for excellent experimental assistance, to B. Bertolaet for synthesis of oligonucleotides, to D. Herschlag for advice, to R. Evans of Synergen Inc. for the plasmid T7pT5T, to S. Miller of Biogen Inc. for the plasmid pSM102-2, and to W. S. Lane for amino acid analyses and N-terminal amino acid sequencing.

REFERENCES

Anderson, V. E., Weiss, P. M., & Cleland, W. W. (1984) *Biochemistry 23*, 2779–2786.

Bannon, G. A., Calzone, F. J., Bowen, J. K., Allis, C. D., & Gorovsky, M. A. (1983) *Nucleic Acids Res. 11*, 3903–3917.

Bazan, J. F., Fletterick, R. J., & Pilkis, S. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9642–9646.

Bonekamp, F., & Jensen, K. F. (1988) *Nucleic Acids Res. 16*, 3013–3024.

Bowman, E., McQueney, M., Barry, R. J., & Dunaway-Mariano, D. (1988) *J. Am. Chem. Soc. 110*, 5575–5576.

Bowman, E., McQueney, M., Scholten, J. D., & Dunaway-Mariano, D. (1990) *Biochemistry 29*, 7059–7063.

Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.

Brinkmann, U., Mattes, R. E., & Buckel, P. (1989) *Gene 85*, 109–114.

Brunk, C. F., & Sadler, L. A. (1990) *Nucleic Acids Res. 18*, 323–329.

Chang, A. C. Y., & Cohen, S. N. (1978) *J. Bacteriol. 134*, 1141–1156.

Chen, P. S., Toribara, T. Y., & Warner, H. (1956) *Anal. Biochem. 28*, 1756.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

Coutrot, P., Savignac, P., & Mathey, F. (1978) *Synthesis*, 36–38.

Del Sal, G., Manfioletti, G., & Schneider, C. (1989) *BioTechniques 7*, 514–520.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Doolittle, R. F. (1986) *Of Urfs and Orfs: A Primer on How To Analyze Derived Amino Acid Sequences*, University Science Books, Mill Valley, CA.

Eisenberg, S. P., Evans, R. J., Arend, W. P., Verderber, E., Brewer, M. T., Hannum, C. H., & Thompson, R. C. (1990) *Nature (London) 343*, 341–346.

Freeman, S., Seidel, H. M., Schwalbe, C. H., & Knowles, J. R. (1989) *J. Am. Chem. Soc. 111*, 9233–9234.

Garcia, G. M., Mar, P. K., Mullin, D. A., Walker, J. R., & Prather, N. E. (1986) *Cell 45*, 453–459.

Gorovsky, M. A. (1970) *J. Cell Biol. 47*, 619–630.

Gorovsky, M. A. (1980) *Annu. Rev. Genet. 14*, 203–239.

Gouy, M., & Gautier, C. (1982) *Nucleic Acids Res. 10*, 7055–7074.

Gray, J. T., Celander, D. W., Price, C. M., & Cech, T. R. (1991) *Cell 67*, 807–814.

Hidaka, T., & Seto, H. (1989) *J. Am. Chem. Soc. 111*, 8012–8013.

Hidaka, T., Mori, M., Imai, S., Hara, O., Nagaoka, K., & Seto, H. (1989) *J. Antibiot. 42*, 491–494.

Hidaka, T., Imai, S., Hara, O., Anzai, H., Murakami, T., Nagaoka, K., & Seto, H. (1990) *J. Bacteriol. 172*, 3066–3072.

Hilderbrand, R. L. (1983) *The Role of Phosphonates in Living Systems*, CRC Press, Boca Raton, FL.

Horecker, B. L., & Kornberg, A. (1948) *J. Biol. Chem. 175*, 385–390.

Hori, T., Horiguchi, M., & Hayashi, A. (1984) *Biochemistry of Natural C–P Compounds*, Maruzen, Tokyo.

Ikemura, T. (1981) *J. Mol. Biol. 151*, 389–409.

Keohavong, P., & Thilly, W. G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9253–9257.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Lawrence, C. B., & Goldman, D. A. (1988) *CABIOS 4*, 25.

Martindale, D. W. (1989) *J. Protozool. 36*, 29–34.

Martindale, D. W., & Taylor, F. M. (1988) *Nucleic Acids Res. 16*, 2189–2201.

Mastalerz, P. (1984) in *Natural Products Chemistry* (Zalewski, R. I., & Skolik, J. J., Eds.) pp 171–184, Elsevier, Amsterdam.

Mullis, K. B., & Faloona, F. A. (1987) *Methods Enzymol. 155*, 335-350.
Nakamaye, K., & Eckstein, F. (1986) *Nucleic Acids Res. 14*, 9679-9698.
Oste, C. (1988) *BioTechniques 6*, 162-167.
Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 185*, 2444-2448.
Pocalyko, D. J., Carroll, L. J., Martin, B. M., Babbitt, P. C., & Dunaway-Mariano, D. (1990) *Biochemistry 29*, 10757-10765.
Rabinowitz, J., Woeller, F., Flores, J., Kresbach, R. (1969) *Nature (London) 224*, 796-798.
Robinson, M., Lilley, R., Little, S., Emtage, J. S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., & Humphreys, G. (1984) *Nucleic Acids Res. 12*, 6663-6671.
Rosenberg, A. H., Lade, B. N., Chui, D.-s., Liu, S.-W., Dunn, J. J., & Studier, F. W. (1987) *Gene 56*, 125-135.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sander, C., & Schneider, R. (1991) *Proteins: Struct., Funct., Genet. 9*, 56-68.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Sayers, J. R., Schmidt, W., & Eckstein, F. (1988) *Nucleic Acids Res. 16*, 791-802.
Schneider, S. U., Leible, M. B., & Yang, X. P. (1989) *Mol. Gen. Genet. 218*, 445-452.
Seidel, H. M., Freeman, S., Seto, H., & Knowles, J. R. (1988) *Nature (London) 335*, 457-458.
Seidel, H. M., Freeman, S., Schwalbe, C. H., & Knowles, J. R. (1990) *J. Am. Chem. Soc. 112*, 8149-8155.
Shapiro, M. B., & Senepathy, P. (1987) *Nucleic Acids Res. 15*, 7155-7174.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503-517.
Squires, C. H., Childs, J., Eisenberg, S. P., Polverini, P. J., & Sommer, A. (1988) *J. Biol. Chem. 263*, 16297-16302.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113-130.
Taylor, J. W., Schmidt, W., Cosstick, R., Okruszek, A., & Eckstein, F. (1985a) *Nucleic Acids Res. 13*, 8749-8764.
Taylor, J. W., Ott, J., & Eckstein, F. (1985b) *Nucleic Acids Res. 13*, 8764-8785.
Woods, D. (1984) *Focus 6*, 1-3.

Characterization of an Improved Reaction Center Preparation from the Photosynthetic Green Sulfur Bacterium *Chlorobium* Containing the FeS Centers F_A and F_B and a Bound Cytochrome Subunit†

Ute Feiler,*,‡,§ Wolfgang Nitschke,‖ and Hartmut Michel‡

Molekulare Membranbiologie, Max-Planck-Institut für Biophysik, Heinrich Hoffmann Strasse 7, D-6000 Frankfurt am Main 71, FRG, and Département de Biologie, Service de Biophysique, CE Saclay, F-91191 Gif sur Yvette Cedex, France

Received October 15, 1991; Revised Manuscript Received December 23, 1991

ABSTRACT: A photosynthetic reaction center complex was prepared from the green sulfur bacterium *Chlorobium* by solubilization of chlorosome-depleted membranes with lauryl maltoside, followed by anion-exchange chromatography and molecular sieve chromatography. The purified complex was characterized by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, optical spectroscopy, and EPR spectroscopy. The major bands migrated at apparent molecular masses of 50, 42, and 32 kDa (heme-staining) and additional weaker bands at 22, 15, and 12 kDa. The isolated reaction center complex contained about 40 bacteriochlorophyll *a* molecules per primary electron donor, P_{840}, assayed by photooxidation. It was competent in stable low-temperature photoreduction of the FeS centers F_A and F_B. The spectra of these acceptors and their low-temperature photochemistry in the purified complex were the same as found in intact *Chlorobium* membranes and similar to what had been described for photosystem I from plants. Membrane-bound cytochrome c_{553} copurified with the reaction center complex. A ratio of about four hemes per P_{840} was determined. This result indicates that cytochrome c_{553} that is closely associated with the reaction center is a tetraheme cytochrome, as described for some purple bacteria.

The first published data pointing toward a relatedness of the green sulfur bacterial reaction center to photosystem I (PSI)[1] date back to 1968 (Buchanan & Evans, 1968). These early results have been supported by (however somewhat controversial) results suggesting that one or more FeS centers might participate in stabilization of light-induced charge separation within the reaction center (Jennings & Evans, 1977; Swarthoff et al., 1981).

† Supported by the Leibniz-Programm (to H.M.) of the Deutsche Forschungsgemeinschaft (DFG), the Max-Planck Gasellschaft (U.F. and H.M.), and the DFG (W.N.). Supported in part by the CNRS (URA 1290).

* To whom correspondence should be addressed.
‡ Max-Planck-Institut für Biophysik.
§ Present address: Sëction de Biophysique des Protëines et des Membranes, Dëpartement de Biologie Cellulaire et Molëculaire, CE Saclay, F-91191 Gif sur Yvette Cedex, France.
‖ Service de Biophysique, CE Saclay.

[1] Abbreviations: BChl, bacteriochlorophyll; Chaps, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; EDTA, ethylenediaminetetraacetic acid; FeS center, iron–sulfur center; kDa, kilodalton; LDAO, *N,N*-dimethyldodecylamine *N*-oxide; LM, lauryl β-D-maltoside; Mega-9, decanoyl-*N*-methylmaltosylamine; MW, molecular weight; OTG, *n*-octyl β-D-thioglucoside; P_{840}/P_{700}, the primary electron donor in green sulfur bacteria/PSI; PSI, photosystem I; RC, reaction center; SB12, *N*-dodecyl-*N,N*-dimethylammonio-3-propanesulfonate; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/92/0431-2608$03.00/0