FEBS Letters 356 (1994) 345–350

FEBS 14917

Primary structure of a multimeric protein, homologous to the PEP-utilizing enzyme family and isolated from a hyperthermophilic archaebacterium

Christiana Cicicopol, Jürgen Peters, Josef Kellermann, Wolfgang Baumeister*

Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

Received 16 November 1994; revised version received 16 November 1994

Abstract A large protein complex (approx. 2000 kDa) was found in the cytosol of the hyperthermophilic archaebacterium *Staphylothermus marinus*. The purified protein was shown to be a homomultimer of 93 kDa subunits, the primary structure of which was determined by nucleotide sequence analysis. The protein belongs to the family of phosphoenolpyruvate-utilizing enzymes and represents the first member characterized in archaebacteria. Its homomultimeric organisation differs from the typically dimeric structure of its eubacterial and eukaryotic counterparts.

Key words: PEP synthase; PEP-utilizing; Archaea; Hyperthermophilic

1. Introduction

The family of related phosphoenolpyruvate (PEP)-utilizing enzymes comprises three groups of proteins: phosphoenolpyruvate synthase (EC 2.7.9.2), pyruvate orthophosphate dikinase (PPDK) (EC 2.7.9.1) and, phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9), which is Enzyme I of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). The members of this family all contain an active-site histidine residue that mediates the phosphoryl transfer. PEP-synthase of *Escherichia coli*, first investigated by Cooper and Kornberg in 1965 [1], catalyses the following reaction:

$$\text{Pyruvate} + \text{ATP} + \text{H}_2\text{O} \rightleftharpoons \text{PEP} + \text{AMP} + \text{P}_\text{i}$$

It has an important function in gluconeogenesis when enterobacteria grow on small carbon substrates, thus providing an alternative route for the synthesis of PEP. Mutants devoid of this enzyme do not grow on pyruvate, lactate or alanine. Archaebacterial representatives of this enzyme family have not been isolated and characterized so far, although the enzyme activity has already been reported for some hyperthermophilic archaebacteria such as *Methanobacterium thermoautotrophicum* [2] and *Pyrococcus furiosus* [3]. The protein we isolated from the archaebacterium *Staphylothermus marinus* is homologous to PEP-synthase, but in contrast to the other members of this protein family, which are dimeric or tetrameric, this enzyme forms a multimeric high molecular weight complex.

2. Materials and methods

2.1. Bacterial strain and cultivation

Cells of *Staphylothermus marinus*, strain F1, DSM3639, were grown as described in [4] and kindly provided by K.O. Stetter.

*Corresponding author. Fax: (49) (89) 8578 2641.

Abbreviations: PEP, phosphoenolpyruvate; PPDK, pyruvate phosphate dikinase; PTS, phosphoenolpyruvate-dependant sugar phosphotransferase system; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

2.2. Protein purification procedures

2.2.1. Preparation of the total cytosolic fraction. Frozen cell paste (6 g) was thawed and mixed with 12 g of glass beads (0.1 mm diameter) and 2.5 ml of 50 mM Tris-HCl, pH 7.7, 1 μM phenylmethane sulfonylfluoride, and 50 μg/ml of DNAse II. Cells were broken in a Vibrogen cell mill for 3 × 15 s, and the mixture incubated at room temperature for 1 h. The glass beads were then separated by centrifugation at 500 × *g*, washed with 10 ml of buffer containing only 10 mM Tris, and the combined supernatants centrifuged at 28,000 × *g* for 20 min.

2.2.2. Chromatographic procedures. Tentacle anion exchange chromatography on TMAE-Fractogel 650(S) (Merck) was performed using the following gradient system: buffer A, 20 mM Tris, pH 7.7, 5% glycerol, 3 mM sodium azide; buffer B, 500 mM Tris, pH 7.7 , 700 mM NaCl, and 5% glycerol; gradient profile, 10–100% B, 120 min; flow rate, 0.5 ml/min. Fractions of 1 ml were collected. In this type of gel matrix the exchange groups are located on polymeric chains which may flexibly adapt to the surface of large proteins.

Molecular sieve chromatography on Superose 6 (Pharmacia) was performed with 10 mM Tris, pH 7.7, 200 mM NaCl, 3 mM sodium azide, and 10% glycerol as running buffer, at a flow rate of 0.5 ml/min.

2.2.3. Density gradient centrifugation. Fractions of 1 ml were loaded on a gradient of 10–30% sucrose (overlaid on 1 ml of 60% sucrose), in 50 mM Tris, pH 7.7, 200 mM NaCl, and 2 mM sodium azide using 10 ml polypropylene tubes (Beckman). After centrifugation in a SW41 type rotor (Beckman) at 38,000 rpm for 14 h, 0.5 ml fractions were collected from the bottom.

2.3. Polypeptide cleavage procedures

In order to obtain polypeptides suitable for Edman degradation, the purified protein was cleaved with 50 mg/ml of CNBr in 70% formic acid for 4 h. The lyophylized sample was size-fractionated on a TSK 2000 column (LKB) using 0.1% trifluoroacetic acid (TFA), 30% acetonitrile as running buffer. Fractions eluting between 14 and 21 min were lyophylized and separated on a C4 reverse-phase column (Vydac) or a C_{18} RPselect column (Merck) using a 0–60% acetonitrile gradient in 0.1% TFA.

Purified protein denatured with sodium dodecyl sulfate (SDS) was cleaved with endoproteinase Lys-C (Boehringer) and peptides separated by SDS-polyacrylamide gel electrophoresis (PAGE). After blotting on siliconized glass fibre, single bands stained with Coomassie were subjected to Edman degradation.

For peptide mapping of bands excised from stained polyacrylamide gels, the gel slices were incubated on a rotary shaker with 10 mg/ml of CNBr in 80% formic acid at 2°C for 60 h. The eluate was lyophylized and subjected to SDS-PAGE.

2.4. Amino acid sequence analysis

Sequencing was performed with a gas-phase sequencer (Porton LF 3 600; Beckman Fullerton CA) [5]. Phenylthiohydantoin derivatives were separated on-line by reverse-phase chromatography (System Gold; Beckman).

SSDI 0014-5793(94)01304-7

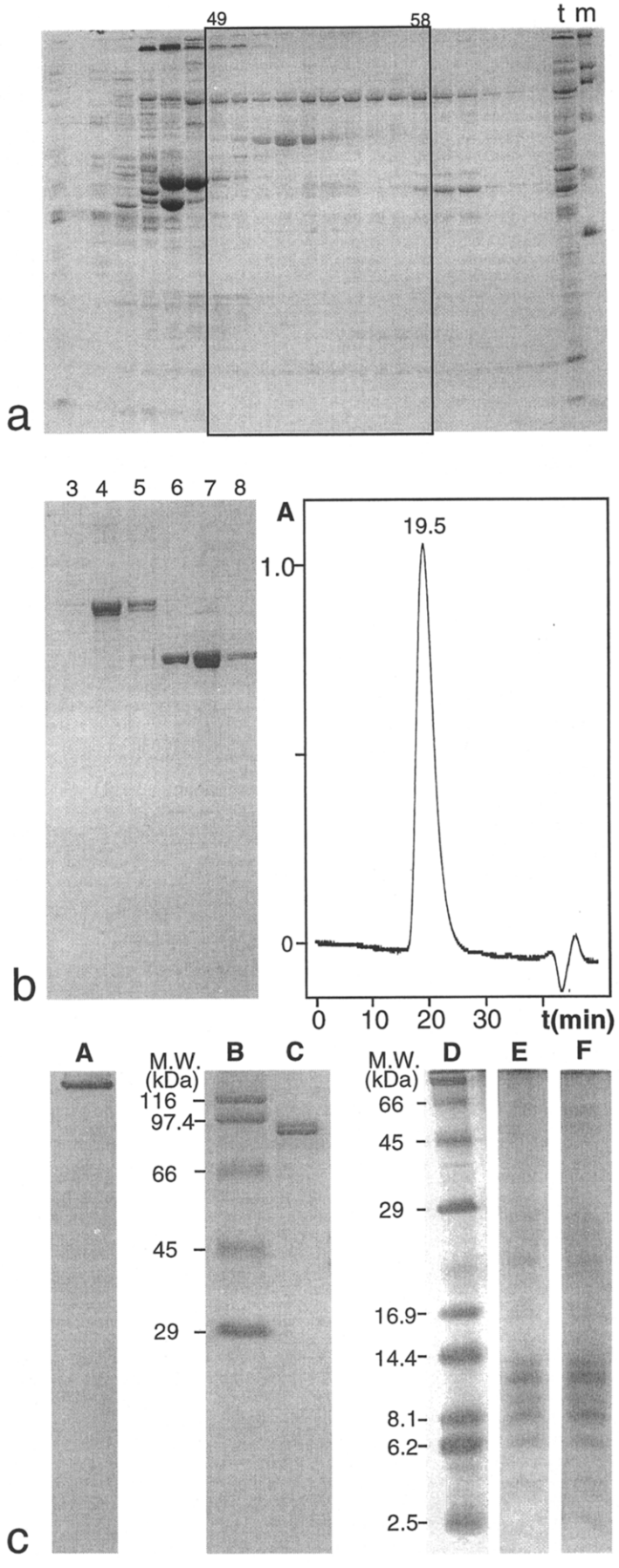

a
49
58
t m
b
3 4 5 6 7 8
A
19.5
1.0
0
0 10 20 30 t(min)
c
A
M.W. (kDa)
B
C
116
97.4
66
45
29
M.W. (kDa)
D
E
F
66
45
29
16.9
14.4
8.1
6.2
2.5

Fig. 1.

2.5. Electron microscopy

The purified protein was deposited on a glow discharged carbon film supported by copper grids and negatively stained with 2% uranyl acetate. Specimens were examined in a Philips EM420 and micrographs were recorded at a magnification of × 36,000.

2.6. Gene isolation and nucleotide sequence analysis

The nucleotide sequence was obtained using a completely PCR-based strategy. Initially PCR was carried out with 5 different amino acid-derived oligonucleotides (length 17–38 bp), employing chromosomal DNA as template, on a Perkin-Elmer Cetus Type 480 thermal cycler using *Pfu* DNA polymerase. PCR fragments were isolated from agarose gels with the Qiaex gel extraction kit (Qiagen) and subjected to cycle sequencing using the dideoxy dye terminator kit from Applied Biosystems. The degenerated oligonucleotides were employed as primers at appropriate annealing temperatures (50–58°C). Sequencing was performed on a 373A automated DNA sequencer (Applied Biosystems).

2.7. Amino acid sequence alignments

Alignments were performed with MACAW [6], an interactive program for multiple alignment construction and analysis. A search for patterns in the amino acid sequence was conducted with the program MOTIFS (Genetics Computer Group (1991), Wisconsin).

3. Results and discussion

3.1. Purification of the 93 kDa protein

The total cytosolic fraction of *S. marinus* was separated by anion-exchange chromatography. As shown in Fig. 1a, the 93 kDa protein is smeared over most of the fractions, indicating surface charge hetereogeneity. This was confirmed by iso-electric focusing studies (data not shown). The analysis of selected fractions did not reveal any significant differences with regard to particle size. The boxed fractions were pooled and separated by sucrose density gradient centrifugation. Fig. 1b shows that the 93 kDa protein appeared as a double band. The (almost homogeneous) fraction of 93 kDa multimer (fraction #4 in Fig. 1b) was subjected to gel-filtration. According to its retention time, the apparent molecular weight of the multimer is about 2000 kDa. The upper and lower bands of the 93 kDa protein varied in relative intensity. Peptide mapping of the upper and lower band yielded several bands of identical apparent molecular weight and no differences between the two fractions were found (Fig. 1c). Edman degradation indicated that the protein was blocked, but seven different N-terminal sequences were obtained after cyanogen bromide or endoproteinase Lys-C cleavage of the total protein fraction, and all of these peptides were later assigned to the 93 kDa protein. Thus, we conclude that the isolated multimer consists of the 93 kDa protein only.

3.2. Electron microscopy

Electron microscopy of the negatively stained complex (Fig. 2) showed that the 93 kDa multimer is a globular particle, approximately 20 nm in diameter, with a central pore or cavity.

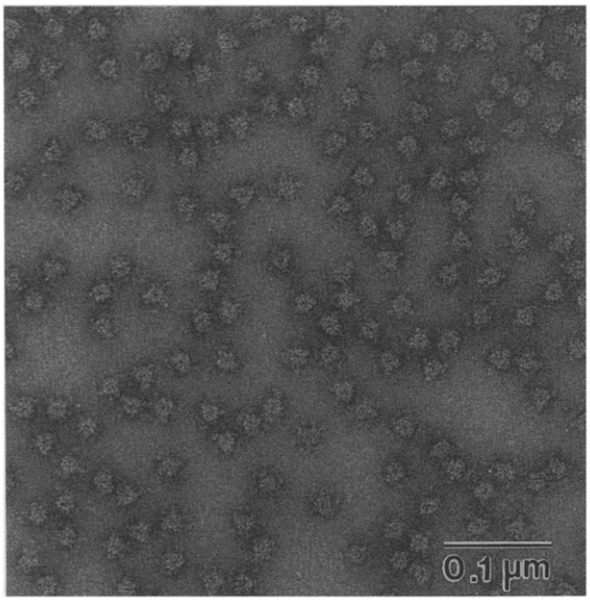


Fig. 2. Electron micrograph of purified multimeric complexes of the 93 kDa protein negatively stained with uranyl acetate.

Its surface is not always clearly defined, and we consistently observed a minority of particles which seemed to have additional material attached to their surface, giving rise to a triangular shape. For a more detailed description of the quaternary structure of the complex, more extensive electron microscopic investigations in conjunction with image analysis will be necessary. In any case, the dimensions of the complex are consistent with the molecular mass of approximately 2000 kDa as found by gel-filtration (see above), and this value has been confirmed by STEM mass measurements (data not shown).

3.3. Primary structure determination (Fig. 3)

Five degenerate oligonucleotides were synthesized according to N-terminal amino acid sequences from three different peptides and used for nested PCR. A 550 bp fragment was obtained which was sequenced using the same primers. The quality of nucleotide sequence data obtained with degenerate primers compared to sequences received with homologous primers. In the amino acid sequence derived from the single open reading frame, three experimentally obtained peptide sequences (amino acid positions 625, 657 and 682) were identified. To obtain the flanking regions of this gene fragment, the restriction endonucleases *Hin*dIII, *Asn*I, *Bgl*II, *Sau*3A, and *Msp*I were used to generate complete genomic digests, which were cyclyzed and used in inverse PCR with homologous primers derived from the

←

Fig. 1. Purification of the 93 kDa protein complex. (a) SDS-PAGE (9% acrylamide, Tricine buffer system [11]) of fractions from a TMAE anion-exchange column. Numbers on top are the fraction numbers, starting with the peak eluting first. t, total cytosolic fraction. In this lane the 93 kDa protein appears as one of the major proteins. m, marker proteins as in Fig. 1c, lane B. (b) Left panel: SDS-PAGE (conditions as in (a)) of sucrose density gradient fractions. The gradient was loaded with the boxed fractions from Fig. 1a. Numbers on top are gradient fraction numbers. The 93 kDa protein is found in fractions #4 and #5. Right panel: Superose 6 chromatogram of gradient fraction #4 subjected to gel-filtration. A, absorbance at 206 nm. (c) PAGE of Superose 6-purified 93 kDa protein (native, lane A; denatured, lane C) and of CNBr-digested bands derived from the 93 kDa protein (CNBr digest of the upper band (E) and the lower band (F), respectively in (C)). A, native PAGE (Laemmli buffer system), 3.5% acrylamide, using piperazine diacrylamide as cross-linker. B–F, SDS-PAGE, 9% (B,C) or 16% (D–F) acrylamide. B, high molecular weight standard mixture for SDS-PAGE (Sigma). D, marker protein mixture of high molecular weight standard mixture (Sigma) and peptide MW marker from horse myoglobins (Pharmacia). Lanes D–F were stained with silver. All other gels shown were stained with Coomassie.

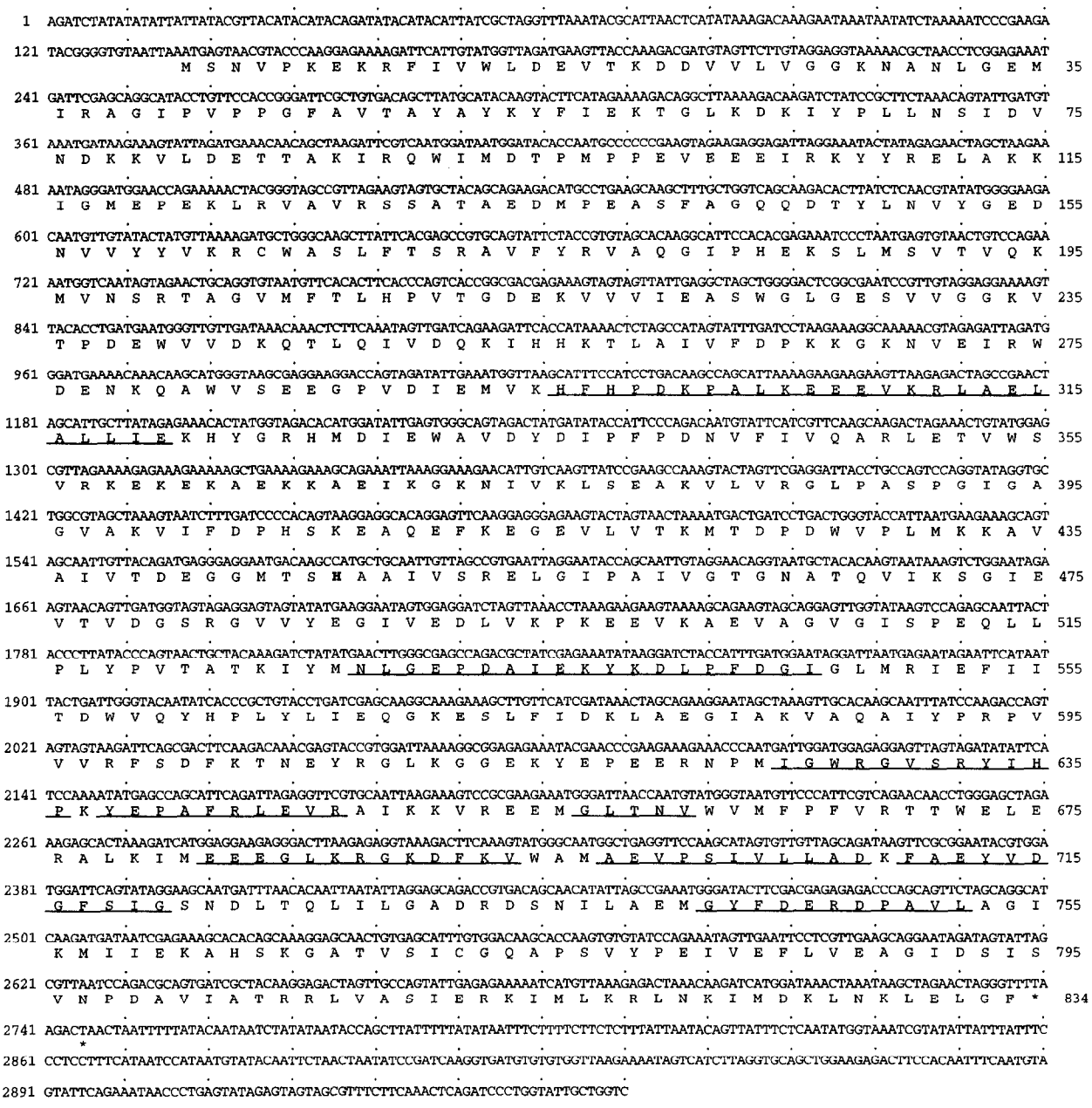
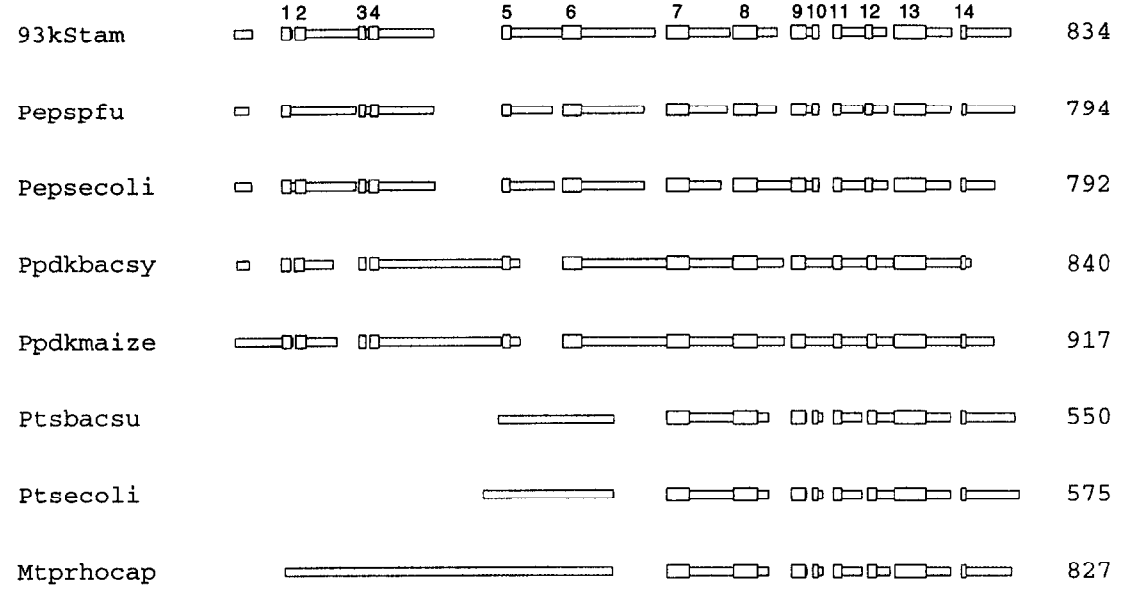


Fig. 3. Determined nucleotide sequence and translation of the structural gene. Nucleotides are marked on the left, amino acids on the right. Amino acid sequences determined by Edmann degradation are underlined.

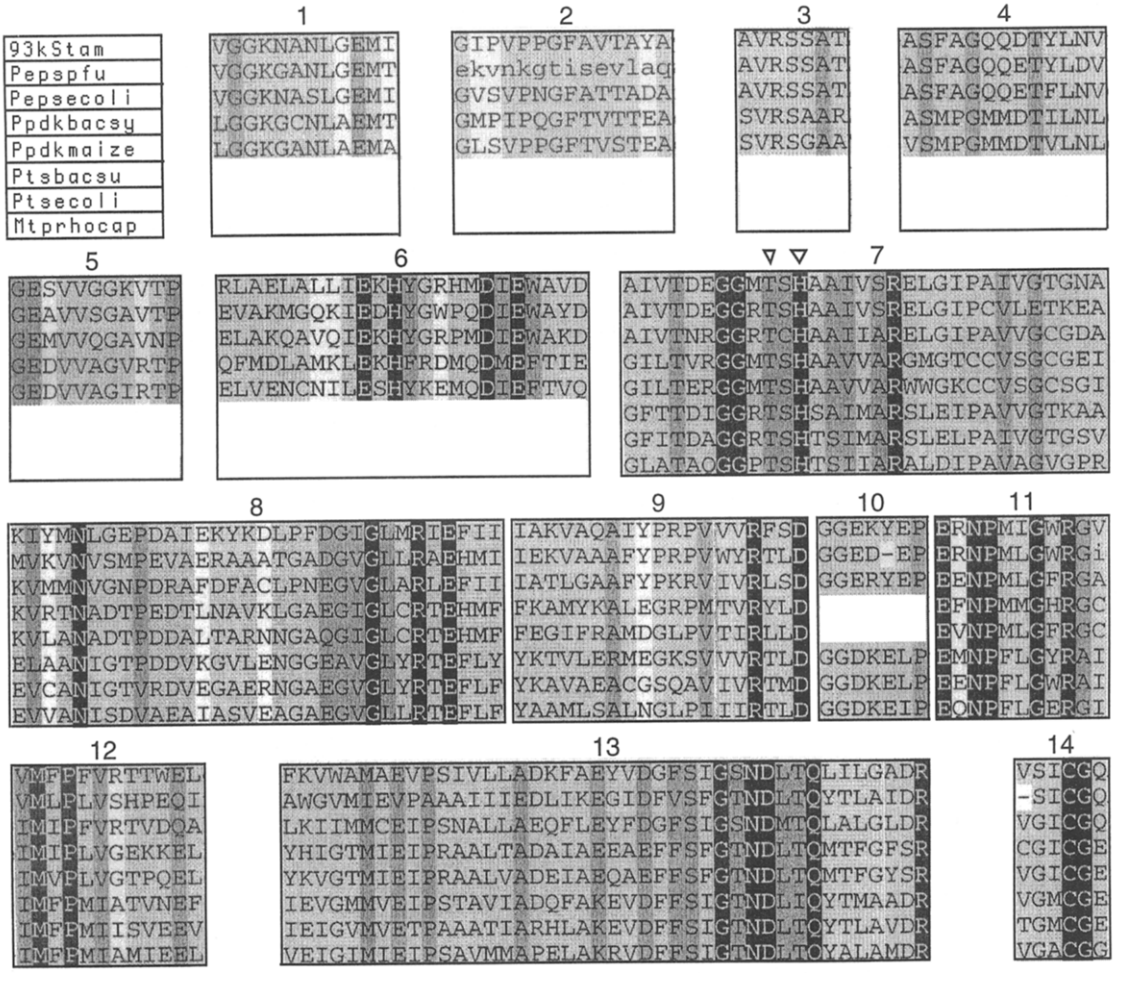
established sequence. By exploiting the overlapping sets of restriction fragments and switching between the different genomic PCR libraries, the complete nucleotide sequence, including the regions flanking the structural gene, was determined. It should be emphasized that the DNA polymerase used for PCR, *Pfu*, has an error rate of 1.6×10^{-6} (information provided by the supplier). Starting with several thousand molecules of template, any mutations would thus disappear in the back-

→

Fig. 4. Multiple alignment of some amino acid sequences belonging to the PEP-utilizing family of enzymes. 93kStam, 93 kDa protein from *Staphylothermus marinus*; Pepspfu, putative PEP synthase from *Pyrococcus furiosus* (Accession U08376); Pepsecoli, PEP synthase from *E. coli* [12], Ppdkbacsy, PPDK from *Bacteroides symbiosus* (presently *Clostridium symbiosum*) [13]; Ppdkmaize, PPDK from *Zea mays* [14]; Ptsbacsub, PTS Enzyme I from *Bacillus subtilis* [15]; Ptsecoli, PTS Enzyme I from *E. coli* [16]; Mtprhocap, multiphosphoryl transfer protein from *Rhodobacter capsulatus* [10]. (A) Schematic representation of the overall alignment generated with MACAW. The number of amino acid residues is noted on the right for each sequence. Boxes indicate regions of strong similarity present in the *Staphylothermus marinus* protein and in at least two of the three subgroups of the family. (B) Blocks of aligned sequences from A. Block #7 contains one of the signature patterns for the PEP-utilizing enzymes. Functionally important residues are marked with arrows. The intensity of the shading provides an indication of the local degree of homology.


1 2 3 4 5 6 7 8 9 10 11 12 13 14
93kStam 834
Pepspfu 794
Pepsecoli 792
Ppdkbacsy 840
Ppdkmaize 917
Ptsbacsu 550
Ptsecoli 575
Mtprhocap 827

A


93kStam
Pepspfu
Pepsecoli
Ppdkbacsy
Ppdkmaize
Ptsbacsu
Ptsecoli
Mtprhocap
1 2 3 4 5 6 7 8 9 10 11 12 13 14

B

ground, which results in more reliable data than those obtained involving cloning. In fact, sequencing of the counter strand as well as multiple sequencing of the same region using different primers strongly confirmed the fidelity of this approach.

3.4. Homology studies

The amino acid sequence obtained by the translation of the nucleotide sequence was subjected to a BLAST search [7] using the NCBI database, and the highest scoring proteins were compared with the sequence of the 93 kDa protein in MACAW (Fig. 4). The comparison revealed a high degree of homology with the family of PEP-utilizing enzymes.

The C-terminal regions from representatives of all three protein classes exhibit a particularly high degree of homology with many stretches of strong conservation (Fig 4A). The alignment of these significant stretches is shown in Fig. 4B. Segment 7 contains the highly conserved region around the histidine that is considered to be the phosphorylation site [8,9]. The preceding threonyl residue was identified as a regulatory site of the maize PPDK [9].

The 93 kDa protein of *S. marinus* contains the two signatures shared by all the members of this PEP-utilizing family and by no other proteins in the current protein data banks. The first pattern is $TX_2GGXTXH(S,T,A)_2(L,I,V,M)_2(S,T,A)R$ (see segment 7, Fig. 4B); and the second one (D,E)(F,G)(L,I,V,M,F)S(L,I,V,M,F)G(S,T)ND(L,I,V,M)XQ(L,I,V,M,F,Y)(S,T, A) (L,I,V,M,F) $(G,A)X_2R$ (see segment 13, Fig. 4B). The biological significance of the latter is not yet known, but the conserved cysteinyl residue from segment 14 (Fig. 4B) is believed to belong to the active site [10].

In comparison with the other members of the PEP-utilizing family of enzymes, the structure of the 93 kDa protein is particularly unusual: most of these proteins occur in dimeric assemblies (PPDK from maize is a tetramer). This difference raises questions as to the functional significance of the multimeric arrangement in the archaebacterium *Staphylothermus marinus* in relation to the eubacterial and eukaryotic counterparts. The sequence most homologous to the 93 kDa protein is the translation of a gene isolated from *Pyrococcus furiosus* which was recently deposited in the databank (Patchx U08376). As this is the only other archaebacterial representative of this class of enzymes, it may be instructive to characterise and compare this protein complex with the quaternary structure of the 93 kDa multimer from *Staphylothermus marinus*.

Acknowledgements: We wish to thank Andrei Lupas for help with the sequence alignments, Marius Boicu for automated DNA sequencing, Ute Santarius for technical assistance with electron microscopy, and Bob Glaeser for critically reading the manuscript.

References

[1] Cooper, R.A. and Kornberg, H.L. (1965) Biochim. Biophys. Acta 104, 616–618.
[2] Eyzaguirre, J., Jansen, K. and Fuchs G. (1982) Arch. Microbiol. 132, 67–74.
[3] Schäfer, T. and Schönheit, P. (1993) Arch. Microbiol. 159, 354–363.
[4] Fiala, G., Stetter, K.O., Jannasch, H.W., Langworthy, T.A. and Madon, J. (1986) System. Appl. Microbiol. 8, 106–113.
[5] Lottspeich, F., Kellermann, J., Henschen, A., Rauth, G and Müller-Esterl, W. (1984) Eur. J. Biochem. 142, 227–232.
[6] Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) Proteins 9, 180–190.
[7] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[8] Meadow, N.D., Fox, D.K. and Roseman, S. (1990) Annu. Rev. Biochem. 59, 497–542.
[9] Roeske, C.A., Kutny, R.M., Budde, R.J.A. and Chollet, R. (1988) J. Biol. Chem. 263, 6683–6687.
[10] Wu, L.-F., Tomich, J.M. and Saier Jr., M.H. (1990) J.Mol.Biol. 213, 687–703.
[11] Schägger, H. and Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[12] Niersbach, M., Kreuzaler, F., Geerse, R.H., Postma, P.W. and Hirsch, H.J. (1992) Mol. Gen. Genet. 231, 332–336.
[13] Pokalyko, D.J., Carrol, L.J., Martin, B.M., Babbit, P.C. and Dunaway-Mariano D. (1990) Biochemistry 29, 10757–10765.
[14] Matsuoka, M., Ozeki, Y., Yamamoto, N., Hirano, H., Kano-Murakami, Y. and Tanaka, Y. (1988) J. Biol. Chem. 263, 11080–11083.
[15] Reizer, J., Hoischen, C. Reizer, A., Pham, T.N. and Saier Jr., M.H. (1993) Protein Sci. 2, 506–521.
[16] Saffen, D.W., Presper, K.A., Doering, T.L. and Roseman S. (1987) J. Biol. Chem. 262, 16241–16253.