Sequence analyses and evolutionary relationships among the energy-coupling proteins Enzyme I and HPr of the bacterial phosphoenolpyruvate:sugar phosphotransferase system

JONATHAN REIZER, CHRISTIAN HOISCHEN, AIALA REIZER,
TAM N. PHAM, AND MILTON H. SAIER, JR.

Department of Biology, University of California at San Diego, La Jolla, California 92093-0116

(RECEIVED September 15, 1992; REVISED MANUSCRIPT RECEIVED December 14, 1992)

Abstract

We have previously reported the overexpression, purification, and biochemical properties of the *Bacillus subtilis* Enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system (PTS) (Reizer, J., et al., 1992, *J. Biol. Chem.* 267, 9158–9169). We now report the sequencing of the *ptsI* gene of *B. subtilis* encoding Enzyme I (570 amino acids and 63,076 Da). Putative transcriptional regulatory signals are identified, and the *pts* operon is shown to be subject to carbon source-dependent regulation. Multiple alignments of the *B. subtilis* Enzyme I with (1) six other sequenced Enzymes I of the PTS from various bacterial species, (2) phosphoenolpyruvate synthase of *Escherichia coli*, and (3) bacterial and plant pyruvate:phosphate dikinases (PPDKs) revealed regions of sequence similarity as well as divergence. Statistical analyses revealed that these three types of proteins comprise a homologous family, and the phylogenetic tree of the 11 sequenced protein members of this family was constructed. This tree was compared with that of the 12 sequenced HPr proteins or protein domains. Antibodies raised against the *B. subtilis* and *E. coli* Enzymes I exhibited immunological cross-reactivity with each other as well as with PPDK of *Bacteroides symbiosus*, providing support for the evolutionary relationships of these proteins suggested from the sequence comparisons. Putative flexible linkers tethering the N-terminal and the C-terminal domains of protein members of the Enzyme I family were identified, and their potential significance with regard to Enzyme I function is discussed. The codon choice pattern of the *B. subtilis* and *E. coli* *ptsI* and *ptsH* genes was found to exhibit a bias toward optimal codons in these organisms. Quantitative analysis of this bias indicated that the *ptsH* genes have higher bias toward codons favored in highly expressed genes than the *ptsI* genes, in agreement with their relative levels of expression.

Keywords: *Bacillus subtilis*; codon usage; Enzyme I; HPr; phosphotransferase system; phylogenetic trees; Q-linkers

The phosphoenolpyruvate:sugar phosphotransferase system (PTS) consists of two general energy-coupling proteins, Enzyme I, and HPr, as well as the sugar-specific permeases, commonly referred to as the Enzyme II complex (Reizer et al., 1988b; Meadow et al., 1990). The system catalyzes the concomitant transport and phosphorylation of its sugar substrates in a process termed group translocation. The PTS permeases found in a va-

riety of bacteria may consist of one, two, three, or four distinct polypeptide chains, and, in certain instances, some of these sugar-specific proteins are fused to protein domains that serve the general energy-coupling function(s) of Enzyme I and/or HPr (Saier & Reizer, 1992). Current work in this laboratory is aimed at defining the three-dimensional structures of the proteins of the PTS of *Bacillus subtilis*. Our collaborative efforts have yielded the structures of both HPr (Kinemage 1) and Enzyme IIA^{glc} (Saier & Reizer, 1992) employing both X-ray crystallography and multidimensional NMR (Wittekind et al., 1989, 1990; Fairbrother et al., 1991, 1992a,b; Liao et al.,

Reprint requests to: Jonathan Reizer, Department of Biology, University of California at San Diego, La Jolla, California 92093-0016.

1991; Herzberg et al., 1992; Stone et al., 1992). Preliminary information regarding the complex formed by the association of these two proteins has also been forthcoming (Chen et al., 1993). We recently described the overproduction, purification, and kinetic characterization of the *B. subtilis* Enzyme I (Reizer et al., 1992). Employing the published procedure, we have prepared nearly homogeneous *B. subtilis* Enzyme I from *Escherichia coli* MZ1 cells bearing the overexpression vector pRE11. This protein is currently being used for determination of its three-dimensional structure by X-ray crystallography.

In order to facilitate the crystallographic analyses, the primary sequence of the protein is required. Determination and analysis of this sequence is the main topic of this paper. We analyze the structural and evolutionary relatedness of this protein and its homologues, i.e., other Enzymes I, phosphoenolpyruvate synthase (PPS), and pyruvate:phosphate dikinases (PPDK). We also present comparative evolutionary data on the second general energy coupling protein of the PTS, HPr. Codon usage analyses of the genes encoding these proteins allowed us to correctly predict their relative abundance in bacterial cells. The operon is shown to be subject to carbon source regulation, and potential regulatory sequences are identified. The data presented in this manuscript complete the sequence information regarding the *pts* operon in *B. subtilis*.

Results and discussion

Sequence of the *ptsI* gene of *B. subtilis*

Figure 1 shows the nucleotide sequence of the *ptsI* gene of *B. subtilis* and the deduced amino acid sequence of Enzyme I as well as the upstream and downstream flanking regions. The *ptsI* gene consists of a coding sequence of 1,713 nucleotides (571 codons) starting with an ATG (start) codon at position 360–362, terminating with a TAA (stop) codon at position 2,010–2,012, and encoding a polypeptide of 570 amino acid residues. The calculated molecular weight of the *ptsI* gene product (63,076 Da) is in good agreement with the value previously determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using the purified protein (66,000 Da; Reizer et al., 1992). It is preceded by the *ptsH* gene with which *ptsI* can be translationally coupled because the initiation codon (ATG) of *ptsI* shares one nucleotide (adenine) with the termination codon (UAA) of *ptsH* (Gonzy-Treboul et al., 1989). It should be noted, however, that the relatively poor energy value of the Shine–Dalgarno sequence for the ribosome-binding site of *ptsI* (−10.6 kcal; Tinoco et al., 1973), and our analyses of the synonymous codon usage of the *B. subtilis* *ptsI* and *ptsH* genes (see below) indicate that expression of *ptsI* is significantly weaker than that of *ptsH*.

Putative transcriptional regulatory signals controlling *pts* operon expression

Gonzy-Treboul et al. (1989) have proposed that *ptsH* and *ptsI* are both transcribed from the same promoter, which is located 50 nucleotides upstream of the *ptsH* initiation codon. The presence of an inverted repeat, which we have identified overlapping the putative −35 hexamer of the *ptsHI* promoter (see Fig. 1), suggests that this region may be recognized by a transcriptional regulatory element. No additional obvious promoter(s) for *ptsI* could be identified by “sequence gazing” within the *ptsH* or *ptsI* genes. Five nucleotides following the UAA translational stop codon of *ptsI* is a DNA sequence (nucleotides 2,018–2,053 in Fig. 1) with several attributes of a ρ -independent transcription terminator (see Fig. 2) including: (1) a potentially stable stem and loop structure, (2) a GC-rich region, and (3) a poly(T) that follows the GC-rich stem and loop configuration. The calculated free energy of stabilization of the hairpin structure is −21 kcal/mol (Tinoco et al., 1973). Two properties of the *B. subtilis* hairpin-forming sequence are noteworthy: (1) A nearly symmetric counterpart to the cluster of seven uridines, which follows the hairpin structure, is provided by a run of five consecutive adenines that precede this structure. This symmetry indicates that the *B. subtilis* hairpin can function as a transcriptional terminator in both directions. (2) The short spacing of only five nucleotides between the UAA translational termination codon of *ptsI* and the 5'-end of the factor-independent transcriptional terminator leads to the possibility that efficient coupling of transcription and translation of the *pts* operon, i.e., *ptsH* and *ptsI*, may obstruct formation of the hairpin structure. This possibility stems from the fact that a ribosome can cover approximately 10 nucleotides beyond the 3'-end of a termination codon (Platt, 1981).

Regulation of *pts* operon expression in *B. subtilis*

To determine if expression of the *pts* operon in *B. subtilis* is subject to regulation by carbon source availability, wild-type cells were grown on D,L-lactate (1%) or a sugar substrate of the PTS (0.3%). As can be seen from the results shown in Table 1, growth in the presence of methyl α -glucoside resulted in a threefold increase in the level of both HPr and Enzyme I activities. By contrast, growth in the presence of either glucose or mannitol showed lower activity, just slightly greater than that observed with lactate-grown cells. In all cases HPr and Enzyme I activities varied in parallel as expected. These results are similar to those observed for *E. coli* and *Salmonella typhimurium* where induction/repression of HPr and Enzyme I occurred in parallel, but they contrast with those observed in the enteric bacteria, as in the latter organisms, growth on any rapidly metabolized PTS sugar resulted in two- to threefold enhanced activities of both proteins (Rephaeli &

Fig. 2. Sequence of the mRNA with a potential hairpin-loop structure, which is located 3' to the *ptsI* gene of *Bacillus subtilis*. Coordinates of the DNA fragment presented in Figure 1 are shown near the corresponding bases.

Saier, 1980; De Reuse & Danchin, 1991). We postulate that the lower activity observed after growth of *B. subtilis* in glucose or mannitol represents a compromise between the inductive and repressive effects of the sugar. Thus, the repressive effect could appear to be stronger in *B. subtilis* than in *E. coli*. Because *B. subtilis* lacks cyclic AMP, an entirely different mechanism of catabolite repression is probably operative.

Presence of a putative open reading frame (ORF) downstream from ptsI

Seventy-one nucleotides downstream from the *ptsI* gene is an ORF (*orfY*) starting with an ATG (start) codon (po-

sitions 2,085–2,087 in Fig. 1) and preceded by a potential Shine–Dalgarno sequence (AAGGA) nine nucleotides upstream from the ATG start codon. The free energy of binding for the interaction of this Shine–Dalgarno sequence and the 3' terminal region of the *B. subtilis* 16S rRNA is relatively poor (–10.6 kcal; Tinoco et al., 1973), indicating that the predicted low expression of *orfY* (see below) may be due, in part, to the low degree of Shine–Dalgarno complementarity and to modulation at the level of initiation of protein synthesis. The ORF (*orfY*) appears to be fused via a BamHI/Sau3A junction (positions 2,328–2,332 in Fig. 1) to the tetracycline-resistance gene of *E. coli* (Soberon et al., 1980). This BamHI/Sau3A fusion was most likely introduced during construction of plasmid pTS20 (Gonzy-Treboul & Steinmetz, 1987). The DNA and the deduced amino acid sequence of *orfY* shown in Figure 1, although incomplete, represents most of the sequence of this ORF since the termination codon of *orfY* was recently identified merely four codons downstream of the BamHI/Sau3A fusion with the tetracycline-resistance gene (W. Nicolson, pers. comm.). Thus, the complete sequence of *orfY* is 258 bp, corresponding to 86 codons (85 amino acids in the protein; molecular weight = 9,928 Da). The synonymous codon usage of *orfY* (data not shown) as well as its codon adaptation index (0.36; Sharp & Li, 1987) are well correlated with *B. subtilis* genes that are expressed at low levels.

The DNA sequence downstream of *orfY* is not presented in Figure 1, but recent sequencing data allowed us to identify an additional putative ORF (*orfZ*), which is located on the strand complementary to the *orfY* coding sequence.

Family of Enzyme I proteins

Figure 3 compares the deduced amino acid sequence of the *B. subtilis* Enzyme I with the sequences of six other Enzyme I proteins as well as the *E. coli* PPS and PPK of plants and bacteria. An overview of sequence conservation among these proteins is also provided (Fig. 3, bottom). All of these enzymes are functional oligomers (dimers or tetramers) that form a phospho-histidyl intermediate (Narindrasorasak & Bridger, 1977; Wood et al., 1977; Edwards & Nakamoto, 1985; Meadow et al., 1990). Highlighted residues (black background) indicate residues that are conserved in all 11 members of this family or in a selected subgroup. Four regions of strong conservation (A–D) are clearly observed in all protein members of this family: (A) positions 209–232; (B) positions 426–435; (C) positions 511–547; and (D) positions 599–600 (see Fig. 3). All seven Enzymes I exhibit six additional regions of strong conservation: I, positions 7–15; II, positions 49–162; III, positions 293–343; IV, positions 379–392; V, positions 455–480; and VI, positions 601–623 (see Fig. 3). Only region A has been characterized functionally. The

Table 1. Levels of HPr and Enzyme I activity after growth in the presence of various carbon sources^a

Carbon source during growth	Specific activity (nmol/mg protein/h)	
	HPr	Enzyme I
D,L-Lactate	13	12
Methyl α -glucoside	35	35
Glucose	20	16
Mannitol	16	13

^a *Bacillus subtilis* strain JH642 cells were grown and assayed for HPr and Enzyme I as described under Materials and methods.

Fig. 3. Alignment of the bacterial and plant pyruvate phosphate dikinases (PPDKs), phosphoenolpyruvate synthase of *Escherichia coli* (PPS), and the Enzyme I proteins and protein domains of the bacterial PTS. **Upper panel:** The consensus sequence (Consensus) is provided below the aligned sequences. Numbering above the aligned sequences correspond to the residue position in the multiple alignment and not to the residue number in any one of the aligned proteins. Numbers to the left of the sequences shown indicate the first amino acid in the row. Highlighted (black background) residues indicate residues that are conserved in all 11 members of this family or in a selected subgroup. An exclamation mark above the aligned sequences indicates a residue conserved in all 11 proteins or protein domains of this family. An asterisk indicates a residue conserved in all PPDKs and all Enzymes I, including the Enzyme I protein domain (designated Fru.) of the *Rhodobacter capsulatus* multiphosphoryl transfer protein (MTP), but not in PPS. Residues conserved in all Enzymes I, including the Enzyme I protein domain of *R. capsulatus*, are denoted by squares. A circle indicates residues conserved in all Enzyme I members of this family but not in the fructose-inducible Enzyme I protein domain (Enzyme I; Fru.) of *R. capsulatus*. The four regions of strong conservation in all 11 members of this protein family (A-D) and the six regions of strong conservation in all Enzymes I (I-VI) are indicated with horizontal lines above the aligned sequences (see text for details). Abbreviations used and references to the published protein sequences are as follows: PPDK, pyruvate phosphate dikinase of *Bacteroides symbiosus* (B.sy.) (Pocalky et al., 1990), *Flaveria trinervia* (F.t.) (Rosche & Westhoff, 1990), and *Zea mays* (Z.m.) (Matsuoka et al., 1988); Enzyme I of *Alcaligenes eutrophus* (A.e.) (Pries et al., 1991), *Escherichia coli* (E.c.) (Saffen et al., 1987; De Reuse & Danchin, 1988), *R. capsulatus* (R.c.) (Wu et al., 1990), *Staphylococcus carnosus* (S.c.) (Kohlbrecher et al., 1991), *Strepto-* (continued)

histidyl residue in this region (histidine₂₁₇ in the aligned proteins shown in Fig. 3) is the site of Enzyme I phosphorylation by HPr(His~P) (Meadow et al., 1990) and of PPDK by ATP or phosphoenolpyruvate (PEP) (Goss et al., 1980; Burnell, 1984; Roeske et al., 1988). The preceding threonyl residue (threonine₂₁₅ in the aligned proteins shown in Fig. 3) is the site of regulation, i.e., inactivation of the maize PPDK due to phosphorylation by an ADP-dependent bifunctional regulatory protein (Burnell, 1984; Edwards & Nakamoto, 1985; Burnell & Hatch, 1986; Roeske et al., 1988). Although the GGXTSH box in this region (positions 212–217 in Fig. 3) is conserved in all PPDK and Enzyme I members of this family, this mode of regulation is apparently not functional in *E. coli* or *B. subtilis* since no acid-stable phosphorylated derivative of the *E. coli* or *B. subtilis* Enzyme I was detected in autoradiograms of SDS-polyacrylamide gels following incubation of these proteins with extracts (50–100 µg protein) of the corresponding organisms, 0.1 mM Ap₅A (P¹, P⁵-di(adenosine-5')pentaphosphate), 0.1 mM [β -³²P]ADP (specific activity 400–800 cpm/pmol), and in the presence (5 mM) or absence of ATP or PEP (data not shown). Regions B and C are conceivably involved in PEP binding and phosphoryl transfer to the catalytic histidine residue, whereas region D is believed to contain the conserved active-site cysteinyl residue (cysteine₅₉₉ in the aligned proteins shown in Fig. 3; see also Wu et al., 1990). The regions (I–VI) that are strongly conserved in the Enzymes I, but not in PPS or the PPDKs, may conceivably function in HPr recognition or in phosphoryl transfer to HPr. Possibly, only regions I–III are involved in the recognition of HPr since reversible phosphoryl transfer occurs between HPr(His~P) and a 30-kDa N-terminal domain of the *E. coli* Enzyme I (LiCalsi et al., 1991).

Examination of the fully conserved residues in the 11 proteins aligned in Figure 3 and computer analyses of these sequences confirmed the validity of the signature sequence proposed by Bairoch (1992), i.e.,

TXXGGXTXH(STA)(STA)(LIVM)(LIVM)(STA)R,

and revealed an additional signature sequence that is present in all members of this protein family but which is absent from all other proteins in the current PIR (Version 32) and SWISSPROT (Version 22) data banks. This sequence is:

(DE)(FG)(LIVMF)S(LIVMF)G(TS)ND(LIVM)XQ

Ambiguous residues at a specific position are given in parentheses, whereas any amino acid at a position in which the residue is not specified is denoted by X. This signature sequence should assist in the identification of new protein members of this family.

A recent report has suggested that Enzyme I of *E. coli* contains a helix–turn–helix motif that might be important for binding of the protein to DNA and consequently to transcriptional regulation (LiCalsi et al., 1991). The computer program of Brennan and Matthews (1989) revealed that most of the members of this family (the Enzyme I family) contain at least one helix–turn–helix motif (data not shown), but that none of these motifs is conserved in even a majority of the proteins that comprise this family. Consequently, the physiological significance of the helix–turn–helix motif in protein members of this family is questionable.

Putative flexible linkers in Enzyme I family proteins

In previous reports, we concluded from computer-aided sequence analyses that an interdomain flexible Q-linker joins the *B. subtilis* N-terminal region of the hydrophilic IIA^{glc} domain to the C-terminal end of the IICB^{glc} moiety of the glucose permease (Sutrina et al., 1990; Saier & Reizer, 1992). Two-dimensional NMR studies have established the validity of this conclusion, showing that this N-terminal region of the IIA^{glc} domain is highly mobile, exhibiting essentially unrestricted intramolecular motion (Stone et al., 1992). Interdomain Q-linkers and alanine/proline-rich hinges that occur in the boundaries of functionally and structurally distinct but otherwise interacting domains of other PTS permeases have also been identified (Erni, 1989; Sutrina et al., 1990; Wu et al., 1990). The sequences depicted in Figure 4 reveal that all protein members of the Enzyme I family contain putative Q-linkers at the ends of regions III shown in Figure 3. These linkers are rich in glutamine, arginine, proline, glutamate, and serine. Thus, 40–57% of these 18–22-residue-long hydrophilic regions are comprised of amino acids that usually occur in Q-linkers. Four aspects of these putative linkers are noteworthy. First, although Q-linkers are usually not strongly conserved in otherwise homologous proteins, the putative linkers in all Enzymes I are similar, and two of the residues, i.e., arginine and glutamate, are conserved in all Enzyme I and PPDK members of this family. Possibly these residues provide functional attributes in addi-

(continued from facing page)

coccus salivarius (S.s.) (Gagnon et al., 1992b), and *Salmonella typhimurium* (S.t.) (Byrne et al., 1988; LiCalsi et al., 1991); PPS, phosphoenolpyruvate synthase of *E. coli* (E.c.) (Niersbach et al., 1992). The statistical analyses of the alignments and the phylogenetic tree of the family are shown in Table 1 and Figure 5, respectively. **Lower panel:** Average similarity profile among members of the Enzyme I family. The average similarity at a particular position was calculated using a sliding window of 20 residues. The average similarity across the entire alignment is shown as a dotted line. The four regions (A–D) that are conserved in all 11 members of this protein family and the six regions (I–VI) of strong conservation in all 7 Enzymes I are indicated at the top of the figure and discussed in the text.

Fig. 4. Putative Q-linkers in protein members of the Enzyme I family. Amino acids that most frequently occur in Q-linkers are underlined. Each of the numbers to the left of the sequences shown denotes the position of the first residue of the linker region in the indicated protein. Numbers in parentheses denote the positions of the aligned residues in Figure 3 that correspond to the first residues of the linkers shown. The abbreviations used are as in the legend to Figure 3.

sequenced PPDKs cluster together, and the PPS of *E. coli* comprises the third branch. The PPS and PPDK clusters are closer to each other than to the Enzyme I cluster as expected from known structural and catalytic similarities of these enzymes (Saier, 1987). Within the Enzyme I cluster, it is worth noting that the gram-positive Enzymes I form a cohesive cluster, whereas the gram-negative Enzymes I form a more diffuse cluster. Interestingly, the gram-positive Enzymes I are closer to the *E. coli*/*S. typhimurium* proteins than to the *Alcaligenes eutrophus* and *Rhodobacter capsulatus* counterparts. This observation is not surprising because the *A. eutrophus* protein may function only in a regulatory capacity (Pries et al., 1991), whereas the *R. capsulatus* protein is unique in being a domain within a large fructose-inducible multiphosphoryl transfer protein (MTP; Wu et al., 1990).

Polyclonal antibodies were prepared against the purified Enzymes I of *B. subtilis* and *E. coli*. These were then used

[illegible]

Fig. 5. Phylogenetic tree of the 11 homologous proteins that comprise the family that includes PPDs, PPS, and the Enzyme I proteins and protein domains of the PTS. Relative evolutionary distances are given adjacent to the branches. Construction of the phylogenetic tree was as described by Reizer and Reizer (1992) using the progressive alignment method of Feng and Doolittle (1990). The abbreviations used are as indicated in the legend to Figure 3.

The phylogenetic tree for members of the Enzyme I family is shown in Figure 5. Three clusters that correlate with the physiological and catalytic functions of the various protein members of this family are discernible. All of the Enzymes I of the PTS cluster together, the three

Table 2. Binary comparisons of the amino acid sequences of the bacterial and plant pyruvate phosphate dikinases (PPDKs), phosphoenolpyruvate synthase (PPS), and Enzymes I of the bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS)^a

	PPDK (B.sy.) (840)	PPDK (F.t.) (953)	PPS (E.c.) (792)	Enzyme I (S.c.) (574)	Enzyme I (S.t.) (575)	Enzyme I (E.c.) (575)	Enzyme I (A.e.) (592)	Enzyme I (R.c.) (Fru.) (827)	Enzyme I (S.s.) (577)	Enzyme I (B.su.) (570)
PPDK (Z.m.) (947)	54 (596) [68]	79 (632) [102]	30 (526) [10]	29 (520) [9]	29 (522) [10]	29 (522) [10]	35 (504) [11]	33 (499) [10]	31 (498) [10]	30 (500) [11]
PPDK (B.sy.) (840)		55 (596) [72]	36 (487) [10]	30 (496) [9]	31 (487) [7]	28 (491) [6]	30 (488) [7]	30 (485) [10]	31 (478) [9]	31 (475) [8]
PPDK (F.t.) (953)			33 (515) [10]	29 (517) [11]	28 (519) [10]	27 (517) [9]	32 (512) [10]	35 (494) [12]	28 (522) [12]	30 (517) [13]
PPS (E.c.) (792)				30 (486) [10]	27 (500) [13]	27 (500) [13]	31 (488) [11]	30 (488) [10]	27 (503) [12]	31 (498) [12]
Enzyme I (S.c.) (574)					50 (566) [58]	50 (566) [59]	37 (569) [41]	38 (551) [35]	62 (573) [69]	66 (566) [74]
Enzyme I (S.t.) (575)						97 (575) [121]	39 (574) [42]	42 (545) [34]	50 (567) [58]	50 (563) [55]
Enzyme I (E.c.) (575)							39 (573) [44]	44 (548) [27]	50 (567) [56]	51 (558) [56]
Enzyme I (A.e.) (592)								39 (541) [26]	40 (571) [36]	38 (570) [39]
Enzyme I (R.c.) (Fru.) (827)									49 (544) [39]	39 (553) [37]
Enzyme I (S.s.) (577)										62 (570) [77]

^a Values in parentheses below the designation of the protein refer to the number of residues in the intact protein or protein domain. Values presented in the table that are not in brackets or parentheses represent percent identity for segments having the number of compared residues indicated in parentheses. Comparison scores in standard deviations were determined using 100 randomizations of the compared segments and are given in brackets below the values for percent identity. The ALIGN program (Dayhoff et al., 1983) was used to assess similarities of the intact Enzyme I proteins and the C-terminal end of PPS and the various PPDKs. The abbreviations used are the same as those indicated in Figure 3.

to demonstrate immuno-cross-reactivity between selected members of the Enzyme I family. As shown in Figure 6A, antibodies prepared against the *E. coli* Enzyme I reacted most strongly with this enzyme, less well with the *B. subtilis* Enzyme I, and to a still lesser degree with PPDK of *Bacteroides symbiosus*. By contrast, the antibodies prepared against the *B. subtilis* Enzyme I were most active with this protein, whereas the *E. coli* Enzyme I and PPDK were successively less reactive (Fig. 6B). The relative immunoreactivities of the two Enzymes I are in accordance with expectation based on the sequence analysis. The immunoreactivity of PPDK establishes that this protein shares epitopes with the Enzymes I.

Family of HPr proteins

Corresponding evolutionary data for the second energy-coupling protein of the PTS, HPr, are presented for comparative purposes in Figure 7, Table 3, and Figure 8. Three regions of strong sequence conservation, A, B, and C, can be observed in the 12 aligned HPrs (Fig. 7). Region A includes the phosphorylatable histidyl residue (his-

tidine₁₅) and the interactive arginyl residue (arginine₁₇), which are completely conserved in all HPrs. Region B includes seryl residue 46, which is completely conserved in all HPrs. In gram-positive bacteria, this residue is phosphorylated by an ATP-dependent kinase (Reizer et al., 1984, 1989a,b; Reizer, 1989). This region is well conserved in all gram-positive HPrs but not in the proteins from gram-negative bacteria. Region C forms a turn that connects the C-terminal α -helix and β -strand motifs of this protein (Kinemage 1; Wittekind et al., 1990; Herzberg et al., 1992) and includes the highly conserved, adjacent acidic residues (aspartate at position 69 and [usually] glutamate at position 70), which appear to be of functional significance based on site-specific mutagenesis analyses (Koch et al., unpubl. data). However, the precise function of this region is not known. Also indicated in Figure 7, for comparative purposes, are residues that are conserved in (1) HPrs of gram-negative bacteria, (2) HPrs of gram-positive bacteria, or (3) the protein domains of the fructose-inducible diphosphoryl transfer protein of *S. typhimurium* and *E. coli* (DTP; Geerse et al., 1989; Orchard & Kornberg, 1990) as well as of the *R. capsulatus* multiphosphoryl transfer protein (MTP; Wu et al., 1990).

Fig. 6. Immuno-cross-reactivity of proteins of the Enzyme I family. Polyclonal antibodies raised against the *E. coli* Enzyme I (A) or the *B. subtilis* Enzyme I (B) were used for detection of Enzyme I (EI) of *E. coli* (E.c.; open circles), Enzyme I of *B. subtilis* (B.su.; filled circles), and PPDK of *B. symbiosus* (B.sy.; denoted with X) as described in the text.

The conserved sequences around the two phosphorylation sites of the various HPrs can serve as signature patterns for all proteins and protein domains of this family. These two sequences are:

G(LIVM)H(STA)R(PA)(GSTA)(STA)

(GA)(KR)XXXX(KR)S(LIVM)(LIVMF)X(LIVM)XX(LIVM)(GA)

Ambiguous residues at a specific position are given in parentheses, whereas any amino acid at a position in which

the residue is not specified is denoted by X. Both signature sequences are less degenerate than the previously proposed two consensus patterns (Bairoch, 1992). In addition to the 12 protein members of the HPr family shown in Figure 7, 3 additional proteins in the current data banks (PIR Version 32 and SWISSPROT Version 22) were identified by the use of these signature sequences. Thus, the first signature sequence that flanks the Enzyme I-catalyzed PEP-dependent phosphorylation site, i.e., histi-

Table 3. Binary comparisons of the 12 members of the HPr family^a

	<i>K. pneumoniae</i> (85)	<i>S. typhimurium</i> (85)	<i>E. coli</i> (85)	<i>R. capsulatus</i> (85) (Fru.)	<i>E. coli</i> (92) (Fru.)	<i>S. typhimurium</i> (92) (Fru.)	<i>S. aureus</i> (87)	<i>S. carnosus</i> (88)	<i>E. faecalis</i> (89)	<i>S. salivarius</i> (87)	<i>B. subtilis</i> (88)
<i>A. eutrophus</i> (89)	36 (83) [17]	35 (83) [17]	35 (83) [17]	26 (73) [14]	28 (91) [20]	28 (91) [11]	39 (83) [18]	40 (83) [19]	40 (83) [18]	39 (77) [18]	38 (77) [18]
<i>K. pneumoniae</i> (85)		99 (85) [47]	99 (85) [47]	43 (73) [21]	33 (85) [17]	35 (85) [17]	32 (82) [15]	34 (82) [17]	36 (81) [16]	38 (82) [18]	35 (81) [17]
<i>S. typhimurium</i> (85)			100 (85) [49]	44 (73) [21]	33 (85) [17]	35 (85) [17]	33 (82) [17]	35 (82) [18]	37 (81) [15]	39 (82) [18]	36 (81) [16]
<i>E. coli</i> (85)				44 (73) [21]	33 (85) [17]	35 (85) [17]	33 (82) [14]	35 (82) [18]	37 (81) [16]	39 (82) [18]	36 (81) [16]
<i>R. capsulatus</i> (85) (Fru.)					48 (84) [24]	48 (84) [24]	36 (75) [17]	35 (74) [16]	36 (75) [16]	46 (76) [18]	41 (76) [17]
<i>E. coli</i> (92) (Fru.)						98 (92) [56]	43 (80) [18]	42 (86) [19]	44 (87) [20]	45 (83) [19]	41 (86) [19]
<i>S. typhimurium</i> (92) (Fru.)							41 (80) [18]	41 (86) [19]	41 (87) [19]	42 (83) [20]	38 (86) [17]
<i>S. aureus</i> (87)								95 (87) [43]	66 (87) [35]	66 (87) [38]	63 (86) [31]
<i>S. carnosus</i> (88)									67 (88) [35]	67 (87) [38]	63 (88) [32]
<i>E. faecalis</i> (89)										77 (87) [35]	60 (88) [33]
<i>S. salivarius</i> (87)											64 (87) [33]

^a The abbreviations used are the same as indicated in Figure 6. The HPr-like protein domains of the fructose-inducible diphosphoryl transfer proteins (DTP) of *E. coli* and *S. typhimurium* and of the multiphosphoryl transfer protein (MTP) of *R. capsulatus* are denoted (Fru.). Values in parentheses below the designation of the protein refer to the number of residues in the intact protein or protein domain. The FASTA program using the dipeptide identities mode (ktup = 2) (Pearson & Lipman, 1988) was used to assess similarities of the indicated proteins. Values presented in the table that are not in brackets or parentheses represent percent identities for segments having the numbers of compared residues indicated in parentheses. Comparison scores in standard deviations, using the RDF2 program (Pearson & Lipman, 1988) and 200 shuffles are given in brackets below the values for percent identity.

Fig. 7. Alignment of the 12 sequenced HPr proteins or HPr-like protein domains of the bacterial PTS (top panel), and profile of the average similarity among protein members of this family (bottom panel).

Upper panel: Boxed and shaded residues indicate amino acids that are conserved in all 12 members of this family. Residues conserved in all gram-negative proteins (not including the fructose-inducible protein domains (Fru.) of the *E. coli* and *S. typhimurium* diphosphoryl transfer protein (DTP), and the *R. capsulatus* MTP) are boxed and indicated by a minus (–) sign. Residues conserved in the three fructose-specific protein domains (Fru.) of the *E. coli* and *S. typhimurium* DTP and the *R. capsulatus* MTP are shaded, boxed with dashed lines, and

denoted with the (F) sign. Residues conserved in all gram-positive HPrs are boxed with dashed lines and indicated with the plus (+) sign. The consensus sequence (Consensus) is shown below the aligned sequences. Numbers indicate the residue number in each of the 12 members of this protein family. Because the complete sequence of the *E. coli* DTP is yet to be determined, the first amino acid of its HPr-like protein domain is designated X. Abbreviations and references of published sequences are as follows: *Staphylococcus carnosus* (*S. carnosus*; Eisermann et al., 1991); *Staphylococcus aureus* (*S. aureus*; Reizer et al., 1988b); *Bacillus subtilis* (*B. subtilis*; Reizer et al., 1988b, 1989b; Gonzy-Treboul et al., 1989); *Streptococcus salivarius* (*S. salivarius*; Gagnon et al., 1992a); *Enterococcus faecalis* (*E. faecalis*; Deutscher et al., 1986); *Salmonella typhimurium* (*S. typhimurium*; Powers & Roseman, 1984; Byrne et al., 1988); *Escherichia coli* (*E. coli*; De Reuse et al., 1985; Saffen et al., 1987); *Klebsiella pneumoniae* (*K. pneumoniae*; Titgemeyer et al., 1990); *Alcaligenes eutrophus* (*A. eutrophus*; Pries et al., 1991); *Rhodobacter capsulatus* (*R. capsulatus* (Fru.); Wu et al., 1990); *Salmonella typhimurium* (Fru.) (*S. typhimurium* (Fru.); Geerse et al., 1989); *Escherichia coli* (Fru.) (*E. coli* (Fru.); Orchard & Kornberg, 1990). The statistical analyses of the alignment and the phylogenetic tree of the family are shown in Table 2 and Figure 8, respectively. **Lower panel:** The average similarity at a particular position was calculated using a sliding window of 10 residues. The average similarity across the entire alignment is shown as a dotted line. The three regions (A–C) that are conserved in all 12 members of this family are indicated at the top of the figure and discussed in the text.

Fig. 8. Phylogenetic tree of the 12 sequenced members that comprise the family of HPr proteins and HPr-like protein domains. Relative evolutionary distances are given adjacent to the branches. Construction of the phylogenetic tree was as described by Reizer and Reizer (1992) using the progressive alignment method of Feng and Doolittle (1990). The HPr-like protein domains of the DTPs of *E. coli* and *S. typhimurium* and the MTP of *R. capsulatus* are denoted (Fru.).

Table 4. Relative synonymous codon usage of the *ptsH* and *ptsI* genes of *B. subtilis* and *E. coli*^a

Amino acid	Codon ^b	<i>B. subtilis</i>				<i>E. coli</i>				Standard ^c			
		<i>ptsH</i>		<i>ptsI</i>		<i>ptsH</i>		<i>ptsI</i>		<i>B. subtilis</i>		<i>E. coli</i>	
		No.	RSCU	No.	RSCU	No.	RSCU	No.	RSCU	High	Low	High	Low
Phe	UUU	1	2.00	5	0.50	2	1.00	5	0.56	0.62	1.37	0.46	1.30
	UUC↑Δ	0	0.00	15	1.50	2	1.00	13	1.44	1.38	0.63	1.54	0.70
Leu	UUA	3	2.57	10	1.18	0	0.00	5	0.52	1.55	1.20	0.11	0.88
	UUG	0	0.00	5	0.59	0	0.00	3	0.31	0.48	0.86	0.11	0.81
Ser	UCU↑Δ	4	3.43	16	2.89	1	1.00	10	2.31	3.34	1.28	2.57	0.80
	UCC↓Δ	1	0.86	0	0.00	2	2.00	4	0.92	0.08	0.82	1.91	0.85
	UCA	0	0.00	9	1.59	0	0.00	5	1.15	1.41	1.37	0.20	0.89
	UCG↓▽	0	0.00	0	0.00	0	0.00	2	0.46	0.00	0.59	0.04	0.93
Tyr	UAU	1	1.00	5	0.71	0	0.00	2	0.44	0.30	1.27	0.39	1.13
	UAC↑Δ	1	1.00	9	1.29	0	0.00	7	1.56	1.70	0.73	1.61	0.87
Cys	UGU	0	0.00	0	0.00	0	0.00	2	1.00	0.80	0.92	0.67	1.14
	UGCΔ	0	0.00	2	2.00	0	0.00	2	1.00	1.20	1.08	1.33	0.86
Trp	UGG	0	0.00	2	1.00	0	0.00	2	1.00	1.00	1.00	1.00	1.00
Leu	CUU↑	3	2.57	23	2.71	0	0.00	6	0.62	3.02	1.48	0.22	0.77
	CUC	1	0.86	4	0.47	1	0.75	3	0.31	0.11	0.66	0.20	0.49
	CUA▽	0	0.00	0	0.00	0	0.00	1	0.10	0.63	0.32	0.04	0.11
	CUGΔ	0	0.00	9	1.06	7	5.25	40	4.14	0.22	1.47	5.33	2.94
Pro	CCU↑	1	4.00	7	1.40	1	2.00	1	0.27	1.75	1.11	0.23	0.54
	CCC↓▽	0	0.00	0	0.00	0	0.00	1	0.27	0.00	0.31	0.04	0.81
	CCA↑	0	0.00	6	1.20	0	0.00	4	1.07	1.58	0.71	0.44	0.71
	CCGΔ	0	0.00	7	1.40	1	2.00	9	2.40	0.68	1.87	3.29	1.95
His	CAU	1	2.00	3	0.75	1	1.00	4	1.00	1.06	1.36	0.45	1.12
	CACΔ	0	0.00	5	1.25	1	1.00	4	1.00	0.94	0.64	1.55	0.88
Gln	CAA↑	2	2.00	8	1.00	2	0.67	4	0.38	1.70	0.97	0.22	0.80
	CAGΔ	0	0.00	8	1.00	4	1.33	17	1.62	0.30	1.03	1.78	1.20
Arg	CGU↑Δ	1	6.00	11	3.00	0	0.00	18	3.48	3.33	1.08	4.39	1.90
	CGC↑Δ	0	0.00	5	1.36	1	6.00	11	2.13	1.91	1.17	1.56	2.70
	CGA↓▽	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.59	0.02	0.54
	CGG↓▽	0	0.00	2	0.55	0	0.00	1	0.19	0.00	0.82	0.02	0.65
Ile	AUU	1	0.60	19	1.36	2	2.00	23	1.35	1.22	1.53	0.47	1.56
	AUC↑Δ	4	2.40	23	1.64	1	1.00	27	1.59	1.76	1.11	2.53	1.13
	AUA↓▽	0	0.00	0	0.00	0	0.00	1	0.06	0.02	0.36	0.01	0.31
Met	AUG	4	1.00	20	1.00	2	1.00	18	1.00	1.00	1.00	1.00	1.00
Thr	ACU↑Δ	2	1.14	9	1.06	6	2.40	5	0.69	1.84	0.59	1.80	0.48
	ACC↓Δ	0	0.00	1	0.12	4	1.60	8	1.10	0.06	0.67	1.87	1.93
	ACA	5	2.86	17	2.00	0	0.00	6	0.83	1.82	1.73	0.14	0.41
	ACG	0	0.00	7	0.82	0	0.00	10	1.38	0.28	1.01	0.18	1.17
Asn	AAU	1	0.40	4	0.44	0	0.00	6	0.71	0.51	1.12	0.10	1.13
	AAC↑Δ	4	1.60	14	1.56	2	2.00	11	1.29	1.49	0.88	1.90	0.87
Lys	AAA↑Δ	7	2.00	37	1.80	6	1.71	24	1.37	1.77	1.41	1.60	1.45
	AAG	0	0.00	4	0.20	1	0.29	11	0.63	0.23	0.59	0.40	0.55
Ser	AGU	0	0.00	3	0.53	0	0.00	1	0.23	0.30	0.54	0.22	0.76
	AGCΔ	2	1.71	6	1.06	3	3.00	4	0.92	0.87	1.40	1.05	1.78
Arg	AGA▽	0	0.00	4	1.09	0	0.00	1	0.19	0.76	1.75	0.02	0.11
	AGG↓▽	0	0.00	0	0.00	0	0.00	0	0.00	0.00	0.58	0.00	0.11
Val	GUU↑Δ	4	2.67	16	1.52	4	2.29	17	1.94	2.22	1.06	2.24	1.08
	GUC	0	0.00	9	0.86	0	0.00	4	0.46	0.20	1.12	0.15	1.18
	GUA↑	2	1.33	13	1.24	1	0.57	5	0.57	1.28	0.82	1.11	0.39
	GUG	0	0.00	4	0.38	2	1.40	9	1.03	0.30	1.00	0.50	1.35

(continued)

Table 4. Continued

Amino acid	Codon ^b	<i>B. subtilis</i>				<i>E. coli</i>				Standard ^c			
		<i>ptsH</i>		<i>ptsI</i>		<i>ptsH</i>		<i>ptsI</i>		<i>B. subtilis</i>		<i>E. coli</i>	
		No.	RSCU	No.	RSCU	No.	RSCU	No.	RSCU	High	Low	High	Low
Ala	GCU↑Δ	9	3.00	19	1.43	3	1.33	29	1.84	2.37	0.97	1.88	0.50
	GCC↓	0	0.00	2	0.15	2	0.89	13	0.83	0.05	0.79	0.23	0.68
	GCA	2	0.67	17	1.28	1	0.44	13	0.83	1.02	1.20	1.10	0.70
	GCGΔ	1	0.33	15	1.13	3	1.33	8	0.51	0.56	1.04	0.80	1.65
Asp	GAU	2	0.80	21	1.31	0	0.00	18	0.92	1.00	1.25	0.61	1.26
	GAC↑Δ	3	1.20	11	0.69	1	2.00	21	1.08	1.00	0.75	1.39	0.74
Glu	GAA↑Δ	5	1.43	42	1.31	7	1.56	44	1.47	1.52	1.35	1.59	1.30
	GAG	2	0.57	22	0.69	2	0.44	16	0.53	0.48	0.65	0.41	0.70
Gly	GGU↑Δ	2	1.00	5	0.57	2	1.33	20	2.22	1.65	0.74	2.28	1.31
	GGCΔ	3	1.50	13	1.49	4	2.67	14	1.56	0.96	1.35	1.65	1.55
	GGA▽	3	1.50	16	1.83	0	0.00	0	0.00	1.33	1.30	0.02	0.39
	GGG↓▽	0	0.00	1	0.11	0	0.00	2	0.22	0.06	0.61	0.04	0.75

^a Codon usage is presented as the number of codons (No.), and the relative synonymous codon usage (RSCU).

^b Down-pointing arrow (↓) and down-facing triangle (▽) denote nonoptimal or rare codon in *B. subtilis* and *E. coli*, respectively. These codons were defined by RSCU < 0.1 (*B. subtilis*) and RSCU < 0.05 (*E. coli*) in highly expressed genes. Optimal codons that occur in higher frequencies in highly expressed genes of *B. subtilis* or *E. coli* are denoted by an up-pointing arrow (↑) or an up-facing triangle (Δ), respectively (Sharp & Li, 1986; Sharp et al., 1990; Sharp, pers. comm.).

^c High and Low designate RSCU values obtained from a compilation of *B. subtilis* and *E. coli* highly and lowly expressed genes, respectively (Sharp & Li, 1986; Sharp et al., 1990).

dine₁₅, was used to identify the partially sequenced HPr proteins of *Streptococcus sanguis* (Jenkinson, 1989) and *Lactobacillus brevis* (Reizer et al., 1988a and unpubl. data). Interestingly, the second consensus pattern that flanks the protein kinase-catalyzed ATP-dependent phosphorylation site, i.e., serine₄₆, identified the *E. coli* glycogen synthase (Kumar et al., 1986). To the best of our knowledge, the intriguing possibility that the bacterial glycogen synthase can be regulated by post-translational modification involving ATP-dependent phosphorylation, similarly to its mammalian counterpart (Cohen, 1978), has not yet been examined.

The binary statistical analyses shown in Table 3 establish that all sequenced HPrs are homologous, and the phylogenetic tree depicted in Figure 8 reveals a qualitative pattern similar to that shown in the phylogenetic tree of the Enzymes I (Fig. 5). However, there are potentially significant quantitative differences. Thus, the gram-positive HPrs cluster together as do the HPrs of enteric bacteria. Two other HPr clusters include (1) the three fructose (fru.)-inducible HPr-like protein domains of the fructose phosphotransferases of *E. coli*, *S. typhimurium*, and *R. capsulatus*, and (2) the *A. eutrophus* HPr, which appears to function in a regulatory capacity (Pries et al., 1991). Surprisingly, and in contrast to the Enzyme I tree, the *A. eutrophus* HPr is closer to the gram-positive HPrs than to the other gram-negative proteins. It will be interesting to determine whether the *A. eutrophus* HPr also resembles the gram-positive HPrs in being phosphory-

lated by an ATP-dependent kinase. Careful comparison of the HPr phylogenetic tree (Fig. 8) with that of the Enzymes I (Fig. 5) reveals that the relative distance between the cluster of HPrs of gram-negative bacteria and those of gram-positive bacteria is significantly larger than the relative distance between the clusters of Enzymes I of these two groups of organisms. This surprising observation suggests that the HPrs of gram-positive bacteria are more divergent from the gram-negative HPrs than the gram-positive Enzymes I are from their counterparts in gram-negative bacteria. The explanation for this finding may have to do with the protein kinase-catalyzed ATP-dependent phosphorylation at serine₄₆ in the gram-positive HPrs, as this process is lacking in gram-negative bacteria.

Codon usage of the *ptsI* and *ptsH* genes

Previous studies have shown that synonymous codon choice differs among various genes of a specific genome and that these differences correlate with the level of gene expression. Thus, highly expressed genes exhibit a strong preference for a subset of optimal codons, i.e., those that are recognized by the more abundant tRNA species, whereas the constraints imposed by more even codon usage are well correlated with reduced levels of protein production (Gouy & Gautier, 1982; Ikemura, 1985; Sharp & Li, 1987; Sharp & Devine, 1989). The compiled codon usage data for the *ptsI* and *ptsH* genes of *B. subtilis* and *E. coli* are shown in Table 4. To enable comparison between data

sets of different sizes and different amino acids, codon usage is expressed employing relative synonymous codon usage (RSCU) values. The RSCU values are simply the observed frequency of a particular codon divided by the expected frequency if all codons for that amino acid were used randomly (Sharp & Li, 1987). Careful comparison of the calculated RSCU values of the *ptsH* and *ptsI* genes of *B. subtilis* and *E. coli* with the compiled RSCU values of highly and lowly expressed genes in these organisms clearly indicate that these *pts* genes exhibit a general pattern of bias toward optimal codons, i.e., similar to the standard highly expressed genes. This conclusion was further supported by the calculated codon adaptation index (CAI), which provides a measure of the relative level of synonymous codon usage bias and allows for comparison of the degree of bias, both within and between species (Sharp & Li, 1987). Thus, high CAI values indicate that a gene is highly expressed on the basis of strong selection of optimal codons. It should be noted that this index is relatively free of intrinsic effects of gene length and, therefore, that it provides a quantitative measure for predicting the level of expression of genes of various sizes. The CAI values for the *ptsH* gene of *E. coli* (0.63) and *B. subtilis* (0.76) indicate that both genes have high degrees of bias toward those codons favored in the highly expressed genes (see Table 5). By contrast, lower CAI values were obtained for the *ptsI* genes of *E. coli* and *B. subtilis* (0.50 and 0.57, respectively; Table 5), indicating that expression of these genes is well within the range calculated for genes expressed at moderate levels. These findings correlate with previously documented data demonstrating higher intracellular levels of HPr than of Enzyme I (Mattoo & Waygood, 1983; Thibault & Vadeboncoeur, 1985; De Reuse & Danchin, 1988). Furthermore, because the *ptsH* and *ptsI* genes of *B. subtilis* and *E. coli* appear to be transcribed from the same promoter (Gonzy-Treboul et al., 1989; De Reuse & Danchin, 1988), it is tempting to suggest that the differential expression of

these two genes can be due, in part, to a translational regulatory mechanism, which stems from the selective usage of synonymous codons. It is interesting to compare these CAI values with those for the genes encoding the *E. coli* and *B. subtilis* IIA^{glc} proteins. In *E. coli*, IIA^{glc} is expressed constitutively, mainly from a *crr*-specific promoter (De Reuse & Danchin, 1988), whereas in *B. subtilis* it is probably inducibly expressed from a promoter preceding the IICBA^{glc} permease. In accordance with these facts, the CAI value is much higher for the former protein (0.60) than for the latter protein (0.28).

In summary, we have completed the sequencing of the *B. subtilis* *pts* operon and shown that its expression is subject to regulation. The putative transcriptional regulatory elements, which may be involved both in initiation and termination, have been identified. We have further analyzed the phylogeny of the two encoded genes, those for Enzyme I and HPr of the phosphotransferase system. This work serves as an essential link in our ultimate goal to understand the phosphoryl transfer chain of the PTS in three dimensions.

Materials and methods

DNA manipulation and nucleotide sequencing

The plasmid pRE11 carrying the *ptsI* gene of *B. subtilis* was prepared as described before (Reizer et al., 1992). The nucleotide sequence of the *B. subtilis* *ptsI* gene in pRE11 and the *ptsH* gene in pRE (Reizer et al., 1989b) was determined by sequencing both DNA strands employing the dideoxy chain-terminating method (Sanger et al., 1977) with double-stranded templates, T7 DNA polymerase (Sequence version 2.0; U.S. Biochemical Corp., Cleveland, Ohio) and synthetic oligonucleotides as primers.

Immunoblotting analysis

Blotting and immunochemical detection of the *B. subtilis* and *E. coli* Enzyme I and the *B. symbiosus* PPDK was performed as previously described (Reizer et al., 1984) using a slot-blot filtration manifold (Schleicher & Schuell, Inc., Keene, New Hampshire). Quantitation of the chromogenic probe (4-chloro-1-naphthol; Bio-Rad, Hercules, California) was performed using a laser densitometer (Ultra Scan XL; LKB, Bromma, Sweden).

Preparation of [β -³²P]ADP

The reaction mixture (0.5 mL final volume) contained 2 mCi [γ -³²P]ATP (specific activity 4,000 Ci/mmol; ICN Radiochemical, Irvine, California), 2 mM ATP, 20 mM AMP, 250 units myokinase (rabbit muscle; Sigma), 50 mM KCl, 15 mM MgCl₂, 5 mM dithiothreitol, and 50 mM Tris-HCl buffer (pH 7.5). Following incubation for 2 h at 37 °C, the reaction mixture was loaded onto an anion-exchange

Table 5. Codon adaptation indices (CAI) for genes comprising the *pts* operon of *B. subtilis* and *E. coli*

	<i>B. subtilis</i>	<i>E. coli</i>
<i>ptsH</i>	0.76	0.63
<i>ptsI</i>	0.57	0.50
IIA ^{glc} gene	0.28	0.60
Standard ^a		
High	0.73	0.73
Low	0.38	0.28

^a Standard CAI values for *B. subtilis* refer to the mean values of 7 ribosomal protein genes (High) and 15 genes whose function is required for sporulation (Low). Standard CAI values for *E. coli* refer to the mean values for *ompA* and *rpsU* (High) and *lacI* and *trpR* (Low). Standard values are from Sharp and Li (1987) and Shields and Sharp (1987).

column (AG1X8, Bio-Rad), preequilibrated with 0.1 M ammonium formate, and the radiolabeled ADP was separated from [γ - 32 P]ATP by stepwise elution with 10 mL of 0.4 M, 1 M, 2.5 M, and 4 M ammonium formate. Under these conditions the purified [β - 32 P]ADP was eluted in the 1 M fraction, whereas [γ - 32 P]ATP was eluted in the 2.5 M fraction. Thin-layer chromatography on polyethylenimine-cellulose F (Merck, Darmstadt, Germany) with 1 M LiCl as the solvent confirmed the purity of the [β - 32 P]ADP formed.

Protein purification

Purification of the *B. subtilis* and *E. coli* Enzymes I was as described previously (Reizer et al., 1992).

Computer-aided analysis

All sequence analyses and database searches were performed using the GCG package from the University of Wisconsin (Devereux et al., 1984) and the DNASYS package (Smith, 1988). Comparison scores (expressed in SD) were calculated using the RDF2 program using 100 or 200 shuffles as indicated in the table legends. A value of 6 SD is suggestive of homology, whereas a value of 9 SD establishes homology. Construction of phylogenetic trees and estimation of the relative evolutionary distances among members of a protein family were as described by Reizer and Reizer (1992) using the progressive alignment method of Feng and Doolittle (1990).

Growth of cells for enzyme assays

Bacillus subtilis strain JH642 was grown in LB medium containing D,L-lactate (1%), glucose, mannitol, or methyl α -glucoside (0.3%), harvested by centrifugation, washed three times, and resuspended in M63 minimal salt medium. Cells were ruptured in a French pressure cell as previously described (Reizer et al., 1992). Cell debris was removed by centrifugation, and the crude extracts were assayed for Enzyme I and HPr by complementation employing *Staphylococcus aureus ptsI* and *ptsH* mutant crude extracts as described previously (Reizer et al., 1988a, 1992). Activity is expressed in nanomoles of glucose phosphate formed per hour per milligram of protein.

Other methods

SDS-PAGE was performed as previously described (Reizer et al., 1992). Protein was determined by the Lowry method (1951) with bovine serum albumin as a standard protein.

Acknowledgments

We thank Debra Dunaway-Mariano for providing PPK of *B. symbiosus* and Wayne L. Nicholson for sharing sequence data.

We also thank Bodo Rak, James Hoch, and Peter Geiduschek for helpful discussions. This work was supported by Public Health Service grants 5R01AI21702 and 2R01AI14176 from the National Institute of Allergy and Infectious Diseases. C.H. was supported by the Feodor-Lynen-Programm of the Alexander-v.-Humboldt-Stiftung.

References

- Bairoch, A. (1992). PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res.* 20, 2013–2018.
- Brennan, R.G. & Matthews, B.W. (1989). The helix–turn–helix DNA binding motif. *J. Bacteriol.* 264, 1903–1906.
- Burnell, J.N. (1984). Regulation of C_4 photosynthesis: Proximity of a histidine involved in catalysis and a threonine involved in regulation of pyruvate, P_i dikinase. *Biochem. Int.* 9, 686–689.
- Burnell, J.N. & Hatch, M.D. (1986). Activation and inactivation of an enzyme catalyzed by a single, bifunctional protein: A new example and why. *Arch. Biochem. Biophys.* 245, 297–304.
- Byrne, C.R., Monroe, R.S., Ward, K.A., & Kredich, N.M. (1988). DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. *J. Bacteriol.* 170, 3150–3157.
- Chen, Y., Reizer, J., Saier, M.H., Jr., Fairbrother, W.J., & Wright, P.E. (1993). Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIA^{glc}. *Biochemistry* 32, 32–37.
- Cohen, P. (1978). The role of cyclic-AMP-dependent protein kinase in the regulation of glycogen metabolism in mammalian skeletal muscle. *Curr. Top. Cell. Regul.* 14, 117–196.
- Dayhoff, M.O., Barker, W.C., & Hunt, L.T. (1983). Establishing homologies in protein sequences. *Methods Enzymol.* 91, 524–545.
- De Reuse, H. & Danchin, A. (1988). The *ptsH*, *ptsI*, and *crr* genes of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system: A complex operon with several modes of transcription. *J. Bacteriol.* 170, 3827–3837.
- De Reuse, H. & Danchin, A. (1991). Positive regulation of the PTS operon of *Escherichia coli*—Genetic evidence for a signal transduction mechanism. *J. Bacteriol.* 173, 727–733.
- De Reuse, H., Roy, A., & Danchin, A. (1985). Analysis of the *ptsH-ptsI-crr* region in *Escherichia coli* K-12: Nucleotide sequence of the *ptsH* gene. *Gene* 35, 199–207.
- Deutscher, J., Pevec, B., Beyreuther, K., Kiltz, H.-H., & Hengstenberg, W. (1986). Streptococcal phosphoenolpyruvate–sugar phosphotransferase system: Amino acid sequence and site of ATP-dependent phosphorylation of HPr. *Biochemistry* 25, 6543–6551.
- Devereux, J., Haeblerli, P., & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387–395.
- Edwards, G.E. & Nakamoto, H. (1985). Pyruvate, P_i dikinase and NADP-malate dehydrogenase in C_4 photosynthesis: Properties and mechanism of light/dark regulation. *Annu. Rev. Plant Physiol.* 36, 255–286.
- Eisermann, R., Fischer, R., Kessler, U., Neubauer, A., & Hengstenberg, W. (1991). Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system. Purification and protein sequencing of the *Staphylococcus carnosus* histidine-containing protein, and cloning and DNA sequencing of the *ptsH* gene. *Eur. J. Biochem.* 197, 9–14.
- Erni, B. (1989). Glucose transport in *Escherichia coli*. *FEMS Microbiol. Rev.* 63, 13–24.
- Fairbrother, W.J., Cavanagh, J., Dyson, H.J., Palmer, A.G., III, Sutrina, S.L., & Reizer, J. (1991). Polypeptide backbone resonance assignments and secondary structure of *Bacillus subtilis* enzyme III^{glc} determined by two-dimensional and three-dimensional heteronuclear NMR spectroscopy. *Biochemistry* 30, 6896–6907.
- Fairbrother, W.J., Gippert, G.P., Reizer, J., Saier, M.H., Jr., & Wright, P.E. (1992a). Low resolution solution structure of the *Bacillus subtilis* glucose permease IIA domain derived from heteronuclear three-dimensional NMR spectroscopy. *FEBS Lett.* 296, 148–152.
- Fairbrother, W.J., Palmer, A.G., III, Rance, M., Reizer, J., Saier, M.H., Jr., & Wright, P.E. (1992b). Assignment of the aliphatic 1H and ^{13}C resonances of the *Bacillus subtilis* glucose permease IIA domain using

- double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* 31, 4413–4425.
- Feng, D.-F. & Doolittle, R.F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol.* 183, 375–387.
- Gagnon, G., Vadeboncoeur, C., & Frenette, M. (1992a). Sequence and expression of the *Streptococcus salivarius ptsH* gene. *J. Dent. Res.* 72, 733 (Abstr. 1739).
- Gagnon, G., Vadeboncoeur, C., Levesque, R.C., & Frenette, M. (1992b). Cloning, sequencing and expression in *Escherichia coli* of the *ptsI* gene encoding the Enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system from *Streptococcus salivarius*. *Gene* 121, 71–88.
- Geerse, R.H., Izzo, F., & Postma, P.W. (1989). The PEP:fructose phosphotransferase system in *Salmonella typhimurium*: FPr combines Enzyme III^{Fru} and pseudo-HPr activities. *Mol. Gen. Genet.* 216, 517–525.
- Gonzy-Treboul, G. & Steinmetz, M. (1987). Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: Cloning of the region containing the *ptsH* and *ptsI* genes and evidence for a *crr*-like gene. *J. Bacteriol.* 169, 2287–2290.
- Gonzy-Treboul, G., Zagorec, M., Rain-Guion, M.-C., & Steinmetz, M. (1989). Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: Nucleotide sequence of *ptsX*, *ptsH* and the 5'-end of *ptsI* and evidence for a *ptsHI* operon. *Mol. Microbiol.* 3, 103–112.
- Goss, N.H., Evans, C.T., & Wood, H.G. (1980). Pyruvate phosphate dikinase: Sequence of the histidyl peptide, the pyrophosphoryl and phosphoryl carrier. *Biochemistry* 19, 5805–5808.
- Gouy, M. & Gautier, C. (1982). Codon usage in bacteria: Correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074.
- Herzberg, O., Reddy, P., Sutrina, S.L., Saier, M.H., Jr., Reizer, J., & Kapadia, G. (1992). Structure of the histidine-containing phosphocarrier protein HPr from *Bacillus subtilis* at 2.0 Å resolution. *Proc. Natl. Acad. Sci. USA* 89, 2499–2503.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2, 13–34.
- Jenkinson, H.F. (1989). Properties of a phosphocarrier protein (HPr) extracted from intact cells of *Streptococcus sanguis*. *J. Gen. Microbiol.* 135, 3183–3197.
- Kohlbrecher, D., Eisermann, R., & Hengstenberg, W. (1991). Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system: Molecular cloning and nucleotide sequence of the *Staphylococcus carnosus ptsI* gene and expression and complementation studies of the gene product. *J. Bacteriol.* 174, 2208–2214.
- Kumar, A., Larsen, C.E., & Preiss, J. (1986). Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose: α -1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the *glgA* gene. *J. Biol. Chem.* 261, 16256–16259.
- Liao, D.I., Kapadia, G., Reddy, P., Saier, M.H., Jr., Reizer, J., & Herzberg, O. (1991). Structure of the IIA domain of the glucose permease of *Bacillus subtilis* at 2.2-Å resolution. *Biochemistry* 30, 9583–9594.
- LiCalsi, C., Croceni, T.S., Freire, E., & Roseman, S. (1991). Sugar transport by the bacterial phosphotransferase system: Structural and thermodynamic domains of Enzyme I of *Salmonella typhimurium*. *J. Biol. Chem.* 266, 19519–19527.
- Lowry, O.H., Rosebrough, N.J., Farr, A.L., & Randall, R.J. (1951). Protein measurements with the Folin phenol reagent. *J. Biol. Chem.* 193, 265–275.
- Matsuoka, M., Ozeki, Y., Yamamoto, N., Hirano, H., Kano-Murakami, Y., & Tanaka, Y. (1988). Primary structure of maize pyruvate, orthophosphate dikinase as deduced from cDNA sequence. *J. Biol. Chem.* 263, 11080–11083.
- Mattoo, R.L. & Waygood, E.B. (1983). Determination of the levels of HPr and Enzyme I of the phosphoenolpyruvate-sugar phosphotransferase system in *Escherichia coli* and *Salmonella typhimurium*. *Can. J. Biochem.* 61, 29–37.
- Meadow, N.D., Fox, D.K., & Roseman, S. (1990). The bacterial phosphoenolpyruvate:glycose phosphotransferase system. *Annu. Rev. Biochem.* 59, 497–542.
- Narindrasorasak, S. & Bridger, W.A. (1977). Phosphoenolpyruvate synthetase of *Escherichia coli*. Molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate. *J. Biol. Chem.* 252, 3121–3127.
- Niersbach, M., Kreuzaler, F., Geerse, R.H., Postma, P.W., & Hirsch, H.J. (1992). Cloning and nucleotide sequence of the *Escherichia coli* K-12 *ppsA* gene, encoding PEP synthase. *Mol. Gen. Genet.* 231, 332–336.
- Orchard, L.M.D. & Kornberg, H.L. (1990). Sequence similarities between the gene specifying 1-phosphofructokinase (*fruK*), genes specifying other kinases in *Escherichia coli* K12, and *lacC* of *Staphylococcus aureus*. *Proc. R. Soc. Lond.* 242, 87–90.
- Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.
- Platt, T. (1981). Termination of transcription and its regulation in the tryptophan operon of *E. coli*. *Cell* 24, 10–23.
- Poculyko, D.J., Carroll, L.J., Martin, B.M., Babbitt, P.C., & Dunaway-Mariano, D. (1990). Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, Enzyme I of the bacterial phosphoenolpyruvate:sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. *Biochemistry* 29, 10757–10765.
- Powers, D.A. & Roseman, S. (1984). The primary structure of *Salmonella typhimurium* HPr, a phosphocarrier protein of the phosphoenolpyruvate:glycose phosphotransferase system. *J. Biol. Chem.* 259, 15212–15214.
- Pries, A., Priefert, H., Kruger, N., & Steinbuchel, A. (1991). Identification and characterization of two *Alcaligenes eutrophus* gene loci relevant to the poly(β -hydroxybutyric acid)-leaky phenotype which exhibits homology to *ptsH* and *ptsI* of *Escherichia coli*. *J. Bacteriol.* 173, 5843–5853.
- Reizer, A. & Reizer, J. (1992). Progressive multiple alignment of protein sequences and the construction of phylogenetic trees. In *Methods in Molecular Biology: Computer Analysis of Sequence Data* (Griffin, A.M. & Griffin, H.G., Eds.). Humana Press, Totowa, New Jersey, in press.
- Reizer, J. (1989). Regulation of sugar uptake and efflux in gram-positive bacteria. *FEMS Microbiol. Rev.* 5, 149–156.
- Reizer, J., Deutscher, J., & Saier, M.H., Jr. (1989a). Metabolite-sensitive, ATP-dependent, protein kinase-catalyzed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system in gram-positive bacteria. *Biochimie* 71, 989–991.
- Reizer, J., Novotny, M.J., Hengstenberg, W., & Saier, M.H., Jr. (1984). Properties of ATP-dependent protein kinase from *Streptococcus pyogenes* that phosphorylates a seryl residue in HPr, a phosphocarrier protein of the phosphotransferase system. *J. Bacteriol.* 160, 333–340.
- Reizer, J., Peterkofsky, A., & Romano, A.H. (1988a). Evidence of the presence of HPr and ATP-dependent HPr kinase in heterofermentative lactobacilli lacking a phosphoenolpyruvate:glycose phosphotransferase activity. *Proc. Natl. Acad. Sci. USA* 85, 2041–2045.
- Reizer, J., Saier, M.H., Jr., Deutscher, J., Grenier, F., Thompson, J., & Hengstenberg, W. (1988b). The phosphoenolpyruvate:sugar phosphotransferase system in gram-positive bacteria: Properties, mechanism, and regulation. *Crit. Rev. Microbiol.* 15, 297–338.
- Reizer, J., Sutrina, S.L., Saier, M.H., Jr., Stewart, G.C., Peterkofsky, A., & Reddy, P. (1989b). Mechanistic and physiological consequences of HPr(Ser) phosphorylation on the activities of the phosphoenolpyruvate:sugar phosphotransferase system in gram-positive bacteria: Studies with site-specific mutants of HPr. *EMBO J.* 8, 2111–2120.
- Reizer, J., Sutrina, S.L., Wu, L.-F., Deutscher, J., Reddy, P., & Saier, M.H., Jr. (1992). Functional interactions between proteins of the phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis* and *Escherichia coli*. *J. Biol. Chem.* 267, 9158–9169.
- Rephaeli, A.W. & Saier, M.H., Jr. (1980). Regulation of genes coding for enzyme constituents of the bacterial phosphotransferase system. *J. Bacteriol.* 141, 658–663.
- Roeske, C.A., Kutny, R.M., Budde, R.J.A., & Chollet, R. (1988). Sequence of the phosphothreonyl regulatory site peptide from inactive maize leaf pyruvate, orthophosphate dikinase. *J. Biol. Chem.* 263, 6683–6687.
- Rosche, E. & Westhoff, P. (1990). Primary structure of pyruvate, orthophosphate dikinase in the dicotyledonous C(4) plant *Flaveria trinervia*. *FEBS Lett.* 273, 116–121.
- Saffen, D.W., Presper, K.A., Doering, T.L., & Roseman, S. (1987). Sugar transport by the bacterial phosphotransferase system. Molecular cloning and structural analysis of the *Escherichia coli ptsH*, *ptsI*, and *crr* genes. *J. Biol. Chem.* 262, 16241–16253.
- Saier, M.H., Jr. (1987). *Enzymes in Metabolic Pathways*. Harper & Row, New York.
- Saier, M.H., Jr. & Reizer, J. (1992). Proposed uniform nomenclature

- for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* 174, 1433–1438.
- Sanger, F., Nicklen, S., & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Sharp, P.M. & Devine, K.M. (1989). Codon usage and gene expression level in *Dictyostelium discoideum*: Highly expressed genes do “prefer” optimal codons. *Nucleic Acids Res.* 17, 5029–5039.
- Sharp, P.M., Higgins, D.G., Shields, D.C., Devine, K.M., & Hoch, J.A. (1990). *Bacillus subtilis* gene sequences. In *Genetics and Biotechnology*, Vol. 3 (Zukowski, M., Ganesan, A.T., & Hoch, J.A., Eds.), pp. 89–98. Academic Press, New York.
- Sharp, P.M. & Li, W.-H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24, 28–38.
- Sharp, P.M. & Li, W.-H. (1987). The codon adaptation index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295.
- Shields, D.C. & Sharp, P.M. (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational bias. *Nucleic Acids Res.* 10, 8023–8040.
- Smith, D.W. (1988). A complete, yet flexible, system for DNA/protein sequence analysis using VAX/VMS computers. *Comput. Appl. Biosci.* 4, 212.
- Soberon, X., Covarrubias, L., & Bolivar, F. (1980). Construction and characterization of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. *Gene* 9, 287–305.
- Stone, M.J., Fairbrother, W.J., Palmer, A.G., III, Reizer, J., Saier, M.H., Jr., & Wright, P.E. (1992). The backbone dynamics of the *Bacillus subtilis* glucose permease IIA domain determined from ¹⁵N NMR relaxation measurements. *Biochemistry* 31, 4394–4406.
- Sutrina, S.L., Reddy, P., Saier, M.H., Jr., & Reizer, J. (1990). The glucose permease of *Bacillus subtilis* is a single polypeptide chain that functions to energize the sucrose permease. *J. Biol. Chem.* 265, 18581–18589.
- Thibault, L. & Vadeboncoeur, C. (1985). Phosphoenolpyruvate-sugar phosphotransferase transport system of *Streptococcus mutans*: Purification of HPr and Enzyme I and determination of their intracellular concentrations by rocket immunoelectrophoresis. *Infect. Immun.* 50, 817–825.
- Tinoco, J., Borer, P.N., Dengler, B., & Levine, M.D. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature* 246, 40–41.
- Titgemeyer, F., Eisermann, R., Hengstenberg, W., & Lengeler, J.W. (1990). The nucleotide sequence of *ptsH* gene from *Klebsiella pneumoniae*. *Nucleic Acids Res.* 18, 1898.
- Wittekind, M., Reizer, J., Deutscher, J., Saier, M.H., Jr., & Klevit, R.E. (1989). Common structural changes accompany the functional inactivation of HPr by seryl phosphorylation or by serine to aspartate substitution. *Biochemistry* 28, 9908–9912.
- Wittekind, M., Reizer, J., & Klevit, R.E. (1990). Sequence-specific ¹H NMR resonance assignments of *Bacillus subtilis* HPr: Use of spectra obtained from mutants to resolve spectral overlap. *Biochemistry* 7, 7191–7200.
- Wood, H.G., O'Brien, W.E., & Michaels, G. (1977). Properties of carboxytransphosphorylase; pyruvate, phosphate dikinase; pyrophosphate-phosphofructokinase and pyrophosphate-acetate kinase and their roles in the metabolism of inorganic pyrophosphate. *Adv. Enzymol.* 45, 85–155.
- Wu, L.-F., Tomich, J.M., & Saier, M.H., Jr. 1990. Structure and evolution of a multidomain multiphosphoryl transfer protein. Nucleotide sequence of the *fruB* (*HI*) gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. *J. Mol. Biol.* 213, 687–703.