The Form II Fructose 1,6-Bisphosphatase and Phosphoribulokinase Genes Form Part of a Large Operon in *Rhodobacter sphaeroides*: Primary Structure and Insertional Mutagenesis Analysis[†,‡]

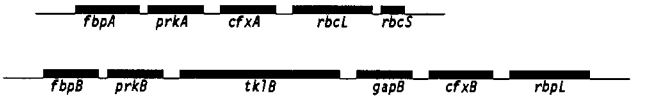
Janet L. Gibson,[§] Jiann-Hwa Chen,[∥] Paula A. Tower,[∥,⊥] and F. Robert Tabita*[,§]

Department of Microbiology and The Biotechnology Center, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210, and Department of Microbiology and Center for Applied Microbiology, The University of Texas at Austin, Austin, Texas 78712

Received March 21, 1990; Revised Manuscript Received May 29, 1990

ABSTRACT: Fructose 1,6-bisphosphatase (FBPase) and phosphoribulokinase (PRK) are two key enzymes of the reductive pentose phosphate pathway or Calvin cycle of photosynthetic carbon dioxide assimilation. Early studies had indicated that the properties of enzymes isolated from photosynthetic bacteria were clearly distinct from those of enzymes obtained from the chloroplasts of higher plants [for a review, see Tabita (1988)]. The eucaryotic enzymes, which are light activated by the thioredoxin/ferredoxin system (Buchanan, 1980), were each shown to contain a putative regulatory amino acid sequence (Marcus et al., 1988; Porter et al., 1988). The enzymes from photosynthetic bacteria are not controlled by the thioredoxin/ferredoxin system but exhibit complex kinetic properties and, in the case of PRK, there is an absolute requirement of NADH for activity. In the photosynthetic bacterium *Rhodobacter sphaeroides*, the structural genes of the Calvin cycle, including the genes that encode FBPase (*fbp*) and PRK (*prk*), are found in two distinct clusters, and the *fbp* and *prk* genes are closely associated in each cluster. In the present investigation, we have determined the nucleotide sequence of the *fbpB* and *prkB* genes of the form II cluster and have compared the deduced amino acid sequences to previously determined sequences of light-activated enzymes from higher plants and from other eucaryotic and procaryotic sources. In the case of FBPase, there are several regions that are conserved in the *R. sphaeroides* enzymes, including a protease-sensitive area located in a region equivalent to residues 51–71 of mammalian FBPase. There are also two conserved histidine residues that are separated by 56–59 amino acids in all sources of FBPase heretofore examined, suggesting an important structural role. The sequence required for light activation of the chloroplast enzyme is not present in the *R. sphaeroides* FBPase nor is the AMP-sensitive region of the mammalian enzyme. For PRK, the lack of overall homology to the plant enzyme is striking, yet there are three regions that exhibit significant similarities, including a putative ATP-binding site. Evidence is also presented that strongly suggests that the *fbpB* and *prkB* genes are cotranscribed and part of an operon with several other CO_2 assimilatory structural genes.

The Calvin reductive pentose phosphate pathway is the major route by which carbon dioxide is assimilated and reduced by photosynthetic and chemolithoautotrophic organisms. Reactions unique to this pathway are catalyzed by ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO)[1] and phosphoribulokinase (PRK). Additional steps of this biosynthetic reaction sequence are catalyzed by common enzymes of intermediary metabolism, where their function is primarily to reduce the products of CO_2 fixation and regenerate the CO_2 acceptor ribulose 1,5-bisphosphate. In the purple non-sulfur photosynthetic bacterium *Rhodobacter sphaeroides*, several of the structural genes that encode key enzymes of the Calvin cycle were found in two distinct clusters (Gibson & Tabita, 1988), the form I or A cluster and the form II or B cluster:



Both clusters contain genes that encode the enzymes fructose 1,6-bisphosphatase (*fbp* gene) (Gibson & Tabita, 1988), PRK (*prk* gene) (Gibson & Tabita, 1987; Hallenbeck & Kaplan, 1987), and RubisCO (*rbcL*, *rbcS*, or *rbpL* genes) (Gibson & Tabita, 1986; Muller et al., 1985; Quivey & Tabita, 1984) and an unknown gene product (*cfx* gene) (Hallenbeck & Kaplan, 1987; Gibson & Tabita, 1988). In addition, the form II cluster contains an approximate 3-kb insertion not found in the form I cluster, encoding the enzymes glyceraldehyde 3-phosphate dehydrogenase (*gap* gene) (Gibson & Tabita, 1988), and the recently discovered *tkl* gene, encoding the enzyme trans-

[†] This work was supported by grants from the U.S. Department of Agriculture (89-37262-4566 and 87-CRCR-1-2591) and the National Institutes of Health (GM-24497).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02922.

* Correspondence should be addressed to this author at the Department of Microbiology, The Ohio State University.

[§] The Ohio State University.

[∥] The University of Texas.

[⊥] Present address: Department of Medicine, Oregon Health Sciences University, Portland, OR 97201.

[1] Abbreviations: FBPase, fructose 1,6-bisphosphatase; PRK, phosphoribulokinase; RubisCO, ribulose 1,5-bisphosphate carboxylase/oxygenase; LB, Luria broth; IPTG, isopropyl β-D-thiogalactoside; EDTA, ethylenediaminetetraacetic acid; TEM, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 10 mM β-mercaptoethanol; TE, 25 mM Tris-HCl, pH 7.5, 1 mM EDTA; SDS, sodium dodecyl sulfate; Tp, trimethoprim.

0006-2960/90/0429-8085$02.50/0

ketolase (Tabita et al., 1990).

As a first step to understand more about the function, structure, and regulation of the form I and form II Calvin cycle enzymes, we have begun to sequence the DNA fragments that encode the genes of each cluster. The deduced sequence of two important light-activated chloroplast enzymes, fructose 1,6-bisphosphatase (FBPase) (Raines et al., 1988) and PRK (Milanez & Mural, 1988; Roesler & Ogren, 1988), has recently been reported and specific sequences required for light activation have been identified (Marcus et al., 1988; Porter et al. 1988). In phototrophic and chemolithoautotrophic bacteria, FBPase and PRK are not light-activated and the kinetic and regulatory properties of each enzyme are distinct from the light-activated eucaryotic enzymes (Springgate & Stachow, 1972; Siebert et al., 1981; Tabita, 1980; Tabita, 1988). In the present work, we describe the primary structure of FBPase and PRK from a photosynthetic CO_2-fixing bacterium. We also present evidence that *fbpB* and *prkB*, and indeed all the genes of the form II cluster, are part of a single large operon.

Experimental Procedures

Plasmid Propagation and DNA Manipulation. Different regions of the 6-kb *Bam*H1–*Eco*R1 fragment of plasmid pJG106 (Gibson & Tabita, 1988), containing the *fbpB* and *prkB* genes, were cloned into pUC8 vectors as previously reported (Gibson & Tabita, 1987, 1988). The plasmids were transferred into *E. coli* JM107 (Yanisch-Perron et al., 1985); the resultant transformants were incubated at 37 °C and the plasmids extracted by the alkaline SDS method (Ish-Horowitz & Burke, 1981). In some cases the extracted plasmid DNA was further purified through cesium chloride–ethidium bromide gradients and dialyzed extensively against TE buffer (25 mM Tris-HCl, pH 7.5, 1 mM EDTA). All *E. coli* clones were cultured in Luria broth (LB) (Davis et al., 1980) in the presence of 50 μg/mL ampicillin.

DNA Sequencing. Defined restriction fragments were cloned into pUC vectors and sequenced by using universal and reverse pUC8 sequencing primers. Usually from 200 to 400 nucleotides were read. For regions beyond this area, or in regions where there was difficulty reading sequences at the upper part of the gel, 15-mer oligonucleotides, containing the end sequence of the last reading, were synthesized (Operon Technologies, Inc., San Pablo, CA) and used as sequencing primers. Sequencing was carried out on denatured double-stranded DNA according to instructions of the Sequenase kit (United States Biochemical Corp., Cleveland, OH). For labeling, [^{35}S]dATP was used and purchased from New England Nuclear at a specific activity of 500 Ci/mol. Both strands of the region containing the *fbpB*/*prkB* genes were sequenced, which is included within a 2.1-kb *Bam*H1–*Sma*I fragment (Figure 1). Because the nucleotide sequence of *Rhodobacter* DNA is G-C rich, at least one strand was also sequenced by using a dITP reaction mix, as suggested in the Sequenase kit. This procedure solved any G-C compression.

Growth of Cells. R. sphaeroides was grown photolithoautotrophically in an atmosphere of 1.5% CO_2/98.5% H_2 (Jouanneau & Tabita, 1986) and extracts prepared from sonically disrupted cells as previously described (Gibson & Tabita, 1987). Plasmid- and non-plasmid-containing *E. coli* JM107 were grown in LB broth and approximately 100 μL of an overnight culture was used to inoculate 100 mL of fresh medium in a 500-mL Erlenmeyer flask. After 1–2 h, isopropyl β-D-thiogalactoside (IPTG) was added, where required, to a final concentration of 1 mM. After 3–4 h, the cells were harvested, washed twice with ice-cold 20 mM Tris-HCl, pH 7.5, containing 1 mM EDTA and 10 mM β-mercaptoethanol (TEM), and stored frozen at −70 °C. At the desired time, the cells were thawed in cold TEM and sonicated.

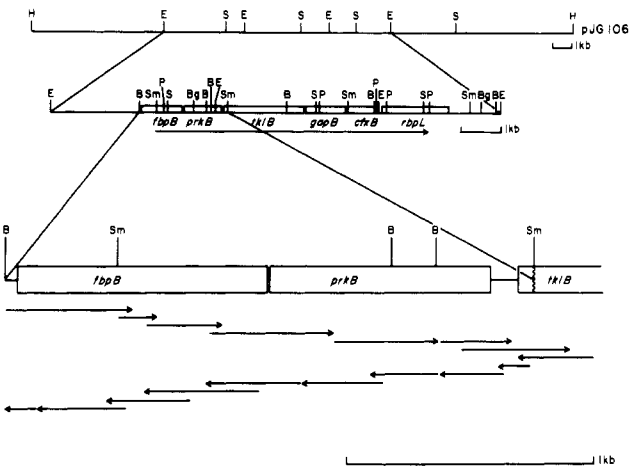


FIGURE 1: Genetic organization of form II CO_2 fixation structural genes obtained from plasmid PJG106 (Gibson & Tabita, 1988) and the strategy for sequencing the approximate 2.1-kb *Bam*H1–*Sma*I fragment containing the *fbpB*/*prkB* genes. The arrows indicate the direction and the extent of the sequencing. Restriction sites: H, *Hin*dIII; E, *Eco*R1; B, *Bam*H1; Sm, *Sma*I; P, *Pst*I; S, *Sal*I; Bg, *Bgl*II.

Immunological Procedures. Antibodies to form II RubisCO and form I PRK were prepared and used in Western immunoblot analyses as previously described (Gibson & Tabita, 1985; Gibson & Tabita, 1987).

Transposon (Tn5) Mutagenesis and Construction of R. sphaeroides Mutants. Mutants specifically deficient in each of the form II CO_2 fixation structural genes were obtained after site-directed recombination with Tn5-inactivated alleles as described in a previous report (Falcone et al., 1988). Southern blots of DNA obtained from the resultant transconjugants verified that homologous recombination led to the replacement of the wild-type gene with the Tn5-inserted allele (data not shown). In the case of *rbpL*, a trimethoprim resistance cartridge was inserted within the coding sequence (Falcone & Tabita, unpublished results).

Results

Nucleotide Sequence of the fbpB and prkB Genes. In a previous study, the *fbpB* and *prkB* genes were shown to be present within a 2.1 kb *Bam*H1–*Sma*I fragment derived from plasmid pJG106, which also contains at least four other structural genes of the Calvin cycle, including the gene that encodes form II RubisCO (Gibson & Tabita, 1988). Both the *fbpB* and *prkB* genes were expressed behind *lac* promoters in *E. coli* to yield active recombinant FBPase and PRK, respectively, and both genes were found to be homologous to similar genes found in the form I structural gene cluster. The nucleotide sequence of the *Bam*H1–*Sma*I fragment was determined by utilizing the strategy outlined in Figure 1. In Figure 2 is shown the complete nucleotide sequence and the deduced amino acid sequence of the approximate 2.1-kb *Bam*H1–*Sma*I fragment. The average G + C content of the entire fragment was 67.3%, consistent with the total genomic DNA of this organism, which shows 69% G + C (Trüper & Pfennig, 1978). It was previously determined that the 1.5-kb *Bam*H1 fragment contained the *fbpB* gene (Gibson & Tabita, 1988). In this study, we found an open-reading frame (ORF) within this fragment (from nucleotide 44 to 1039), which encodes a protein of 331 amino acids (molecular weight of 35 254). There is an excellent ribosome binding site five nu-

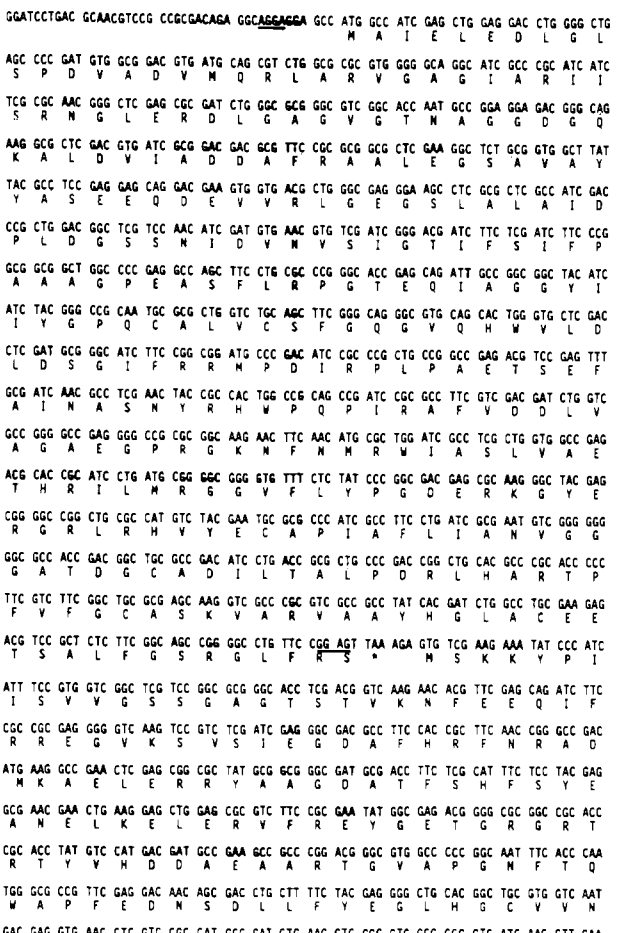


FIGURE 2: Nucleotide sequence of *fbpB* and *prkB* genes within the 2.1-kb *Bam*HI–*Sma*I fragment. The nucleotide sequence was identified for both strands. The entire deduced amino acid sequence is shown below the coding region. The underlined bases represent possible ribosome-binding sites.

cleotides upstream from the ATG initiation codon (Figure 2). Three bases from the stop codon for *fbpB*, an ORF encoding a protein of 292 amino acids (molecular weight of 33 142) was found. This *prkB* reading frame begins with GTG and ends with a TGA stop codon at nucleotide 1918. The putative ribosome binding site is seven nucleotides upstream from the GTG codon. Previous results had established that the 3.7-kb *Pst*I fragment of plasmid pJG106 (Figure 1) encoded an approximate 34 000 molecular weight polypeptide that was assembled into highly active recombinant PRK in extracts of *E. coli*. This polypeptide comigrated with the *R. sphaeroides* form II PRK polypeptide in SDS gels (Gibson & Tabita, 1987). Alignment of the deduced amino terminus of the *prkB* translated product with the previously determined amino terminus of PRK from *Alcaligenes eutrophus* and the *R. sphaeroides prkA* recombinant gene product (Hallenbeck & Kaplan, 1987) showed 63% and 87% identity, respectively.

Table I: Sequence Homology at the Subtilisin-Sensitive Region of FBPase[a]

source of FBPase	sequence	ref
	□ ■□ □□ ■ ■□ □■	
wheat chloroplast	114 A P I S N L T G V Q G A T N V Q G E D Q K 134	*b*
spinach chloroplast	A G I S N L T G I Q G A V N I Q G E D Q K	*c*
pig kidney	51 A G I A H L Y G I A G S T N V T G D Q V K 71	*d*
S. cerevisiae	61 A E L V N L V G L A G A S N F T G D Q Q K 81	*e*
S. pombe	68 A E L V N L I G L S G I V N S T G D E Q K 88	*e*
E. coli	43 A G L V D I L G A S G A E N V Q G E V Q K 64	*b*
R. sphaeroides I	18 G G I E D L A G L C G - T N T D G D G Q K 37	*f*
R. sphaeroides II	35 N G L E R D L G A G V G T N A G G D G Q K 55	this work

[a] Amino acids are indicated by the single-letter code; identical residues are shown with a closed box (■) and those with conservative amino acid substitutions are shown with an open box (□). Residue numbers refer to the flanking amino acid residues of the region from each source of enzyme. The dash (–) signifies a gap introduced to optimize homology. [b] Raines et al., 1988. [c] Marcus et al., 1987. [d] Marcus et al., 1982. [e] Rogers et al., 1989. [f] Gibson and Tabita, unpublished work.

These results strongly suggest that the initiation codon for *prkB* is at nucleotide 1043. Another open-reading frame begins at nucleotide 2035, which has been identified as the start site for the *tklB* gene, encoding the Calvin cycle enzyme transketolase (Tabita et al., 1990).

Analysis of the Deduced Amino Acid Sequence. A comparison of the deduced amino acid sequence of the form I (Gibson & Tabita, unpublished results) and form II *R. sphaeroides* FBPase with the previously determined pig kidney sequence (Marcus et al., 1982) and more recently deduced sequences from wheat chloroplasts (Raines et al., 1988), *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* (Rogers et al., 1988), and *Escherichia coli* (Raines et al., 1988) is presented in Figure 3. The protein sequences of the *R. sphaeroides* FBPase isozymes are 70% identical with each other and differ significantly only at the amino terminus, where the form II enzyme contains a 17-residue extension. The molecular weights are 34 075 and 35 254 for the form I and form II isozymes, respectively. Both *R. sphaeroides* isozymes share roughly 30% identity with the yeast and pig kidney enzymes. Neither of the *R. sphaeroides* FBPase isozymes exhibit the cysteine-containing sequence thought to participate in the disulfide bond that is reduced by thioredoxin during light activation of the chloroplast enzyme (Marcus et al., 1988). *R. sphaeroides* FBPase also does not contain the cyclic AMP dependent protein kinase recognition site (Arg-Arg-X-Ser) (Rittenhouse et al., 1986) found in the *Saccharomyces* enzyme at residues 9–12 (Rogers et al., 1988). Thus, except for the amino terminus, the form II (and form I) FBPase shows homology in five of the six domains known to be conserved among eucaryotic enzymes (Rogers et al., 1988) (Figure 3). The method of Kyte and Doolittle (1982) was used to construct hydropathy plots of form I, form II, and *S. cerevisiae* FBPase (Figure 4). The profiles of the form I and form II enzymes were very similar except for the amino terminus and a region near the carboxy terminus, reflecting the high homology of these two isozymes. However, a very different profile was observed for the *S. cerevisiae* enzyme. Only at the carboxy terminus did all three enzymes show similar profiles, in a region corresponding to the F domain, which is thought to

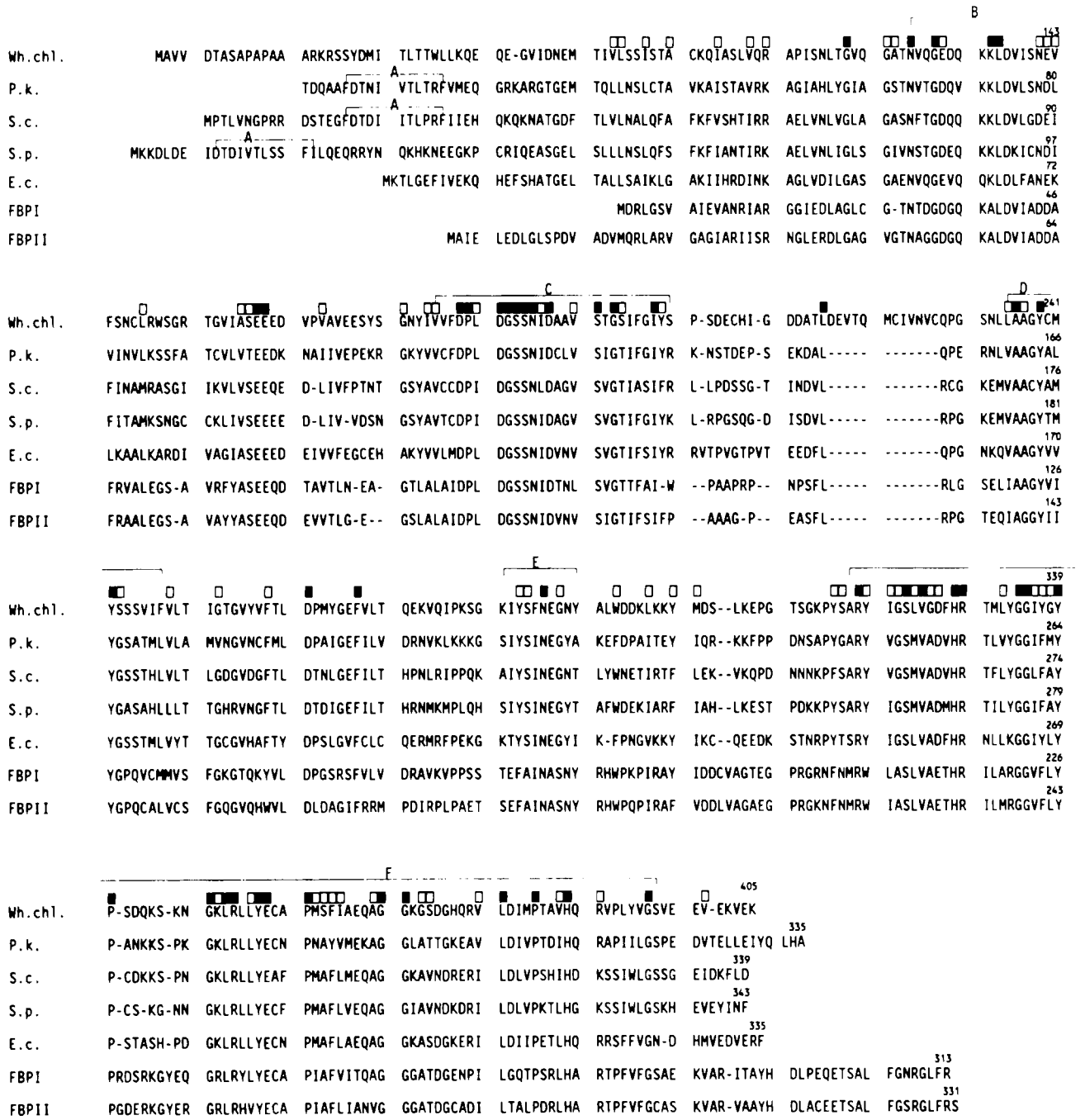


FIGURE 3: Comparison of deduced amino acid sequences of FBPase from wheat chloroplasts (Raines et al., 1988) (wh. chl.), *S. cerevisiae* and *S. pombe* (Rogers et al., 1988) (S.c. and S.p., respectively), *E. coli* (Raines et al., 1988) (E.c.), *R. sphaeroides* FBPI (Gibson and Tabita, unpublished work), and FBPII to the sequence of the pig kidney (p.k) enzyme determined by Edman degradation (Marcus et al., 1982). Amino acid sequences were aligned to give the best fit. Except for the transit sequence of the wheat chloroplast FBPase, this comparison includes the complete sequence of each protein. Amino acids showing exact identity from all sources of FBPase are indicated (■); conserved amino acid replacements are also depicted (□). When alignment of the sequence suggests an addition or deletion, the gap is represented by a dash (-). Regions previously shown to exhibit significant homology have been bracketed and labeled A–F (Rogers et al., 1988).

participate in the active site (Marcus et al., 1982). Both *R. sphaeroides* FBPase isozymes show homology to a previously determined subtilisin-sensitive region (Table I) and both *R. sphaeroides* isozymes contain homologous sequences in the environment of two invariantly conserved histidine residues near the carboxy terminus (Table II).

Comparison of known PRK sequences from spinach (Milanez & Mural, 1988; Roesler & Ogren, 1988) with those from the *R. sphaeroides* form I (Gibson & Tabita, unpublished results) and form II enzymes is shown in Figure 5, along with the recently determined sequence from the chemolithoautotrophic bacterium *Alcaligenes eutrophus* (Kossman et al., 1989). Form I and form II PRK are remarkably conserved proteins, exhibiting 89% identity at the amino acid level. The calculated molecular weights of the polypeptides encoded by *prkA* and *prkB* are 32 661 and 33 142, respectively, in agreement with the migration of the polypeptides in SDS gels (Gibson & Tabita, 1987). By contrast, the plant enzyme shares only 13% identity with the bacterial proteins. One region of similarity that was previously noted within the amino terminus of the spinach and form I PRK (Hallenbeck & Kaplan, 1987; Krieger et al., 1987; Porter et al., 1988) contains a sequence resembling an ATP-binding site. This sequence is also found at the amino terminus of form II PRK. Comparison of the known ATP-binding site at the amino terminus of spinach PRK (Krieger et al., 1987; Porter et al., 1988), and

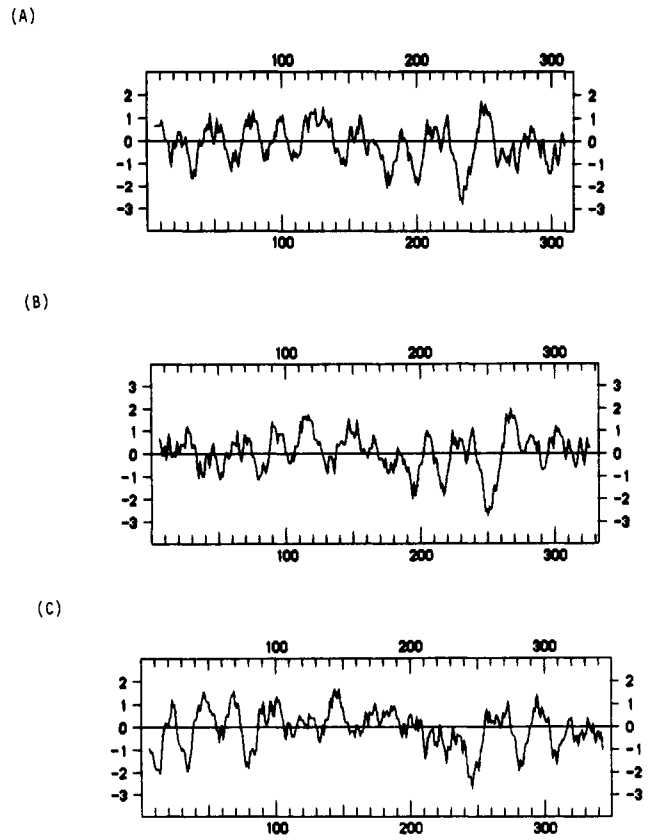


FIGURE 4: Hydropathy plots of the deduced amino acid sequences of *R. sphaeroides* FBPase I (A), *R. sphaeroides* FBPase II (B), and *S. cerevisiae* FBPase (C) (Rogers et al., 1988). The method of Kyte and Doolittle (1982) was used to generate this plot. Positive and negative numbers on the *y*-axis indicate hydrophilic and hydrophobic regions, respectively. The *x*-axis represents the amino acid residues.

Table II: Sequences around Two Conserved Histidine Residues in FBPase[a]

source of FBPase	sequence		ref
wheat chloroplast	G D F ^{328}H R T M	T A V ^{386}H Q R V	*b*
spinach chloroplast	G D F H R T L	T E I H Q R V	*c*
pig kidney	A D V ^{253}H R T L	T D I ^{311}H Q R A	*d*
S. cerevisiae	A D V ^{263}H R T F	S H I ^{321}H D K S	*e*
S. pombe	A D M ^{268}H R T I	K T L ^{326}H G K S	*e*
E. coli	A D F ^{256}H R N L	E T L ^{314}H Q R R	*b*
R. sphaeroides I	A E T ^{213}H R I L	S R L ^{271}H A R T	*f*
R. sphaeroides II	A E T ^{230}H R I L	D R L ^{288}H A R T	this work

[a] Amino acids are indicated by the single-letter code. The numbers refer to the His residue from two regions from each source of enzyme. [b] Raines et al., 1988. [c] Marcus et al., 1987. [d] Marcus et al., 1982. [e] Rogers et al., 1988. [f] Gibson and Tabita, unpublished work.

the putative ATP-binding site at the deduced amino terminus of form I and form II PRK, shows an interesting correlation to known ATP-binding sites from various proteins (Cremo et al., 1989) (Figure 6). The obvious homology of PRKI and PRKII to previously isolated glycine-rich sequences known to act on ATP or GTP is striking and has been noted to contain the general consensus sequence G-X-X-X-X-G-K-(T/S) (Möller & Amons, 1985; Higgins et al., 1986), where X is often one or more Gly residues. Also of interest is the lack of a cysteine residue in this region of bacterial PRK, which for spinach PRK is the residue at the ATP-binding site that, along with Cys-55, is reduced by the light-activated thioredoxin/ferredoxin system (Porter et al., 1988). Since the *R. sphaeroides* PRK is activated by NADH (Gibson & Tabita, 1987), and not the thioredoxin system, it is not surprising that cysteine is absent from this putative ATP-binding site. Another excellent potential nucleotide binding site in PRK I and PRK II is between residues 88 and 95 (Figure 6B); there is also a Gly-rich sequence between residues 127 and 138 that resembles known pyridine nucleotide binding sites (Nagata et al., 1988).

The sequences between residues 127 and 135 and 165 and 189 (Figure 6C,D) are two other regions of *R. sphaeroides* PRK that show significant homology to plant PRK. These homologous regions, from disparate PRK enzymes that show little total homology, suggest that the sequences (Ile/Leu)-Leu-X-X-Glu-Gly-Leu-His and Lys-Ile-X-Arg-Asp-X-X-X-Arg-Gly-X-X-X-Glu might play an important role in catalysis or binding of the substrate ribulose 5-phosphate.

Cotranscription of the Structural Genes of the Form II Cluster. All of the coding regions within the form II gene cluster are transcribed in the same direction (Gibson & Tabita, 1988). In some instances, stop codons and ribosome binding sites of adjacent genes actually overlap (e.g., between *fbpB* and *prkB*) (Figure 2) and all six genes of the form II cluster are tightly spaced, suggesting that the genes of the form II cluster form part of a large operon. To investigate this possibility, transposon mutagenesis of cloned DNA fragments was employed to inactivate all the genes within the form II cluster. Plasmids were chosen for study in which Tn5 had inserted at sites within *fbpB*, *prkB*, *tklB*, or *gapB*, or the trimethoprim resistance gene was inserted within *rbpL* (Figure 7A). In all cases, the disrupted genes were introduced into *R. sphaeroides* via conjugation and colonies were selected in which double-crossover events had occurred, replacing wild-type genes with inactivated genes within the chromosome (Falcone et al., 1988). The resultant strains were then analyzed for the presence of PRK II (*prkB* gene product) or form II RubisCO (*rbpL* gene product) by Western immunoblot analysis (Gibson & Tabita, 1986). The results of the experiments with the *fbpB* and *prkB* knockout strains (Figure 7B) indicated that inactivation of the *fbpB* gene had a polar effect on the expression of *prkB*, which encodes form II PRK, i.e., the upper band of the *R. sphaeroides* PRK doublet (Gibson & Tabita, 1987). These data strongly suggest that *fbpB* and *prkB* are cotranscribed, similar to *fbpA* and *prkA* (Gibson & Tabita, 1988). Most interesting was the finding that mutations in all genes upstream from *rbpL* directly affected the expression of form II RubisCO (Figure 7C). These results indicate that all of the genes within the form II cluster are cotranscribed and that they are part of a single operon.

Discussion

In *R. sphaeroides*, the form II CO_2 fixation structural genes are clustered and are well separated from the form I structural genes, which are localized on a distinctly different chromosomal fragment (Gibson & Tabita, 1988). Indeed, it was recently proposed that *R. sphaeroides* contains two distinct chromosomes, one large and one small, in which many genes are duplicated (Suwanto & Kaplan, 1989). It was further found by these authors that the large chromosome contains the form I genes and the small chromosome contains the form II genes. The polar effect of inactivating genes upstream from *rbpL* would appear to indicate that all of the genes of the form

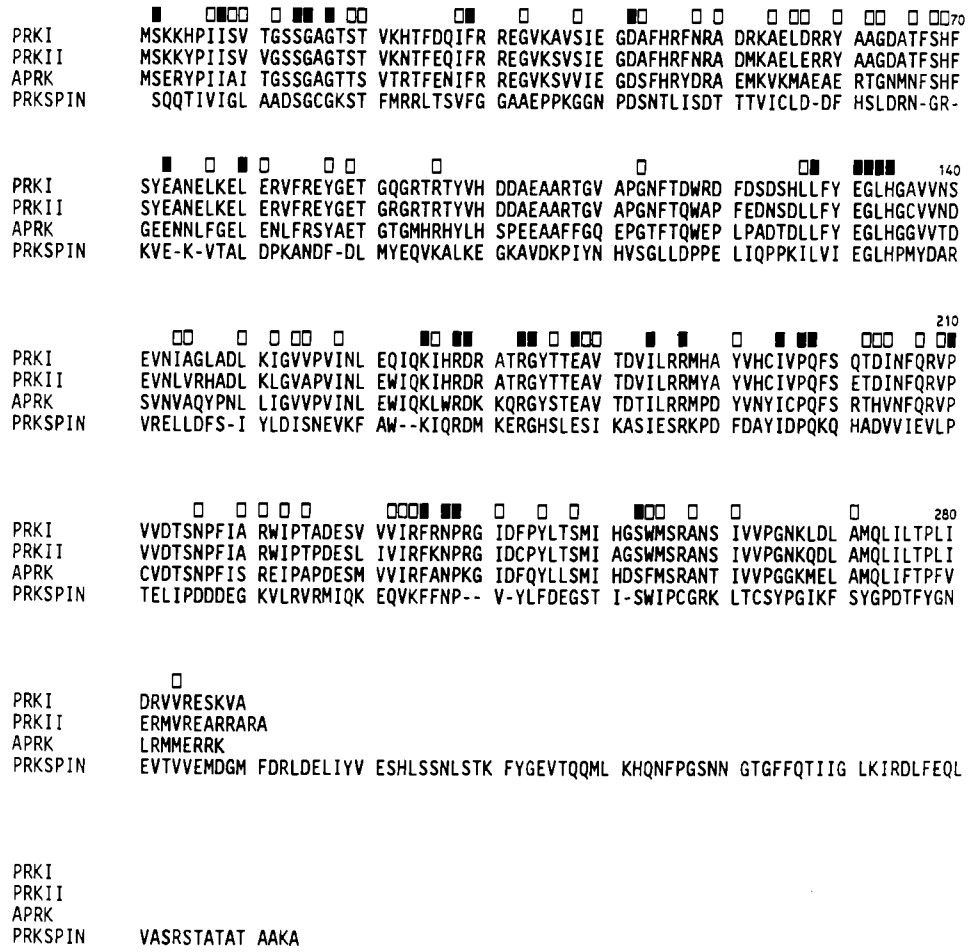


FIGURE 5: Comparison of deduced amino acid sequence of PRKI (Gibson and Tabita, unpublished work) and PRKII from *R. sphaeroides* to the sequence of PRK from spinach chloroplasts (Milanez & Mural, 1988; Roesler & Ogren, 1988) (PRKS) and *A. eutrophus* (Kossmann et al., 1989) (APRK). Amino acids were aligned to give the best fit for the complete sequence of each protein. Amino acids showing exact homology from all sources are indicated (■) along with conserved amino acid replacements (□). When alignment of the sequence suggests an addition or deletion, the gap is represented by a dash (–).

II gene cluster are cotranscribed. Yet, there is still the possibility that the genes are cotranscribed through *cfxB* and form II RubisCO is not produced because *cfxB* is required for transcription of *rbpL*. At this time, we cannot absolutely rule out this possibility since the function of the *cfxB* gene product is not known. However, there is no homology of *cfxB* or *cfxA* to known bacterial transcription activator molecules (Henikoff et al., 1988; Ronson et al., 1987) and recent results with the analogous genes of the form I cluster, *cfxA* and *rbcL rbcS*, conclusively show that these genes are cotranscribed (Gibson and Tabita, unpublished results). It should also be noted that all of the strains that contain upstream insertions that affected form II RubisCO synthesis overproduce the form I RubisCO, similar to previous results with a form II RubisCO negative strain (Falcone et al., 1988). The finding, in the present investigation, that the form II genes are cotranscribed and presumably part of a single operon suggests that investigations on the factors that contribute to the regulation of gene expression will be most interesting.

The primary structure of two important chloroplast-localized enzymes of the Calvin cycle, FBPase and PRK, was recently determined from wheat and spinach, respectively (Milanez & Mural, 1988; Raines et al., 1988; Roesler & Ogren, 1988). Both of these eucaryotic enzymes are subject to light-mediated activation in which thioredoxin reduces a disulfide bond (Buchanan, 1980), which in spinach PRK appears to be formed by Cys-16 and Cys-55 (Porter et al., 1988). In wheat FBPase, there is an insertion of 12 extra amino acids that contain Cys residues in a variable region of the enzyme not found in nonphotosynthetic eucaryotic FBPase (Raines et al., 1988). Moreover, in a recent study, the spinach chloroplast enzyme was also shown to contain an insert with two cysteines separated by four amino acid residues, similar to other enzymes with redox-active cysteines (Marcus et al., 1988). Both the wheat (Cys-Ile-Val-Asn-Val-Cys) and spinach (Cys-Val-Val-Asn-Val-Cys) chloroplast FBPase sequences are homologous, suggesting that this region may be important for light activation. In addition, although both FBPase and PRK from higher plants are found in the chloroplast, both enzymes are synthesized in the cytosol as a proenzyme containing a transit peptide that is removed or processed during transport into the chloroplast. The determination of the sequence of both FBPase and PRK from a procaryotic photosynthetic organism presents some interesting contrasts. First of all, both enzymes are encoded by genes that are closely placed on the chromosome; for *fbpB* and *prkB* there are only 3 nucleotides separating the two genes. Previous results with the related organism *Rhodospirillum rubrum* indicated that FBPase and PRK may be isolated as a two-protein complex (Joint et al., 1972); the finding that *fbpA* and *prkA* (Gibson & Tabita, 1988) and, in this investigation, *fbpB* and *prkB* are cotranscribed raises the possibility that these enzymes may be organized as part of a macromolecular complex after the proteins are synthesized in photosynthetic bacteria.

When the full FBPase sequences from pig kidney, *S. cerevisiae*, and *S. pombe* were compared to the available partial

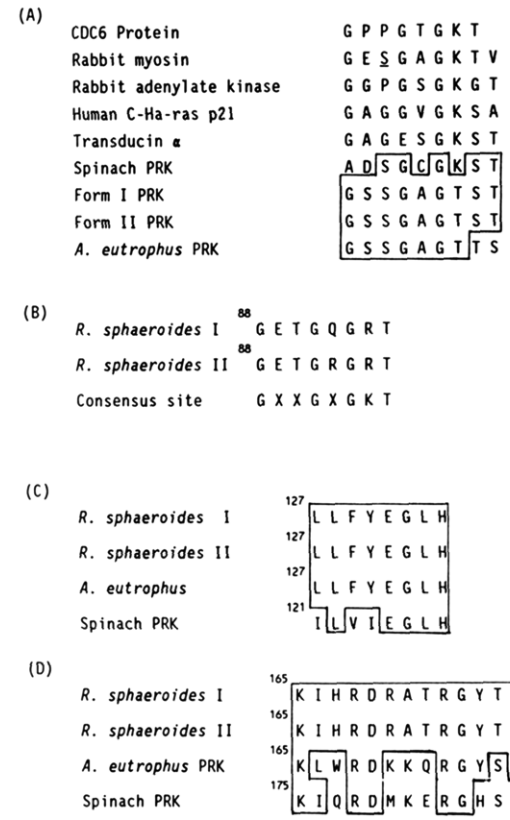


FIGURE 6: Comparison of sequences of bacterial and spinach PRK to sequences of enzymes known to act on ATP or GTP (A). These include the *S. cerevisiae* CDC6 protein (Zhou et al., 1989), rabbit myosin containing the underlined serine known to bind ATP (Cremo et al., 1989), rabbit adenylate kinase (Kuby et al., 1984), human C-Ha-*ras* p21 protein (Capon et al., 1983), transducin α (Tanabe et al., 1985), spinach PRK (Krieger et al., 1987; Porter et al., 1988), and *A. eutrophus* PRK (Kossmann et al., 1989). (B) Comparison of second putative nucleotide-binding site of *R. sphaeroides* form I and form II PRK to consensus nucleotide-binding sequence (Möller & Amons, 1985; Higgins et al., 1986). (C, D) Additional regions of bacterial and spinach PRK homology.

sequences from *E. coli* and spinach chloroplasts, six regions (A-F) of particular conservation were identified (Rogers et al., 1988). The recently deduced full *E. coli* and wheat chloroplast sequences verified this comparison (Raines et al., 1988). In the present study, we found that the form II (and form I) FBPase from *R. sphaeroides* contained regions B–F but does not contain the amino terminal A region. The mammalian enzyme is inhibited by AMP, a function that appears to be localized to the amino terminus, since deletion of the first 25 residues at this terminus results in a loss of AMP inhibition (Chatterjee et al., 1985). The lack of the A region in *R. sphaeroides* FBPase presumably might account for insensitivity to AMP inhibition, which, although not determined directly in this investigation, had previously been shown to have no effect on the enzyme from a related purple nonsulfur photosynthetic bacterium (Springgate & Stachow, 1972). The *S. cerevisiae* enzyme is also regulated by catabolite inactivation via phosphorylation at Ser-11 by a cyclic AMP dependent protein kinase (Rittenhouse et al., 1986). However, the protein kinase recognition site (Arg-Arg-X-Ser) is not found within the *R. sphaeroides* form I and form II FBPase sequence, suggesting that the two *R. sphaeroides* isozymes are insensitive to cyclic AMP dependent protein kinase phosphorylation. It is thus apparent that the *R. sphaeroides* FBPases exhibit considerable differences when compared at the regulatory amino terminus of enzymes from either photosynthetic or nonphotosynthetic eukaryotic organisms.

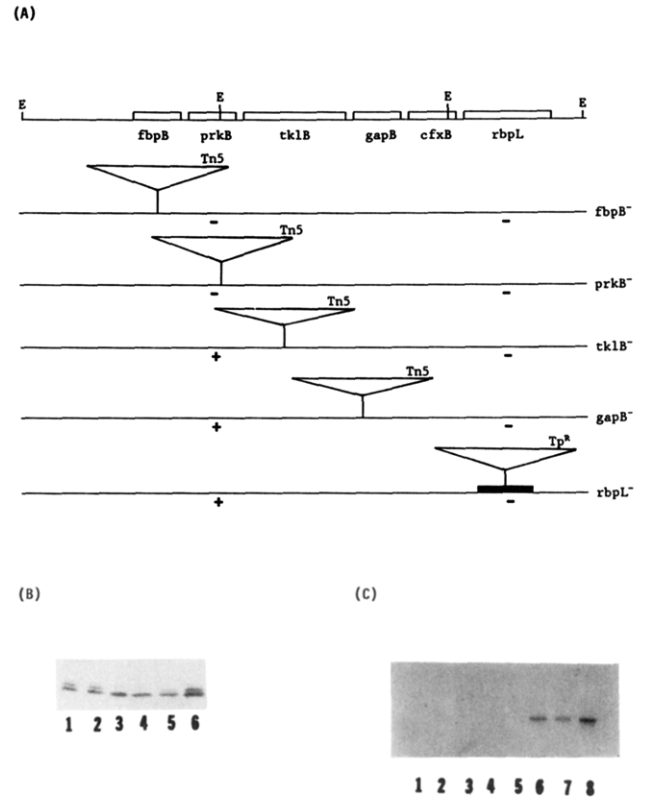


FIGURE 7: Cotranscription of the form II CO_2 fixation structural genes. (A) Position of Tn*5* and the trimethoprim resistance (TpR) cartridge within the coding sequence of the genes indicated. The presence (+) or absence (−) of form II PRK and form II RubisCO is indicated under each insertion map. After the inactivated gene replaced the wild-type allele (Falcone et al., 1988), all constructs were form II RubisCO negative. (B) Western immunoblots using antisera to *R. sphaeroides* PRKI to examine extracts of wild-type strain HR (lane 1). In addition, extracts were obtained from cells containing a Tn*5* insertion in a position outside the form II structural gene region (lane 2), a Tn*5* insertion within *fbpB* (lane 3), a Tn*5* insertion within *prkB* (lane 4), a Tn*5* insertion within *tklB* (lane 5), and a Tn*5* insertion within *gapB* (lane 6). (C) Western immunoblots using antisera to form II RubisCO to examine extracts of wild-type strain HR containing a Tp resistance cartridge within *rbpL* (lane 1) and extracts of cells containing Tn*5* insertions within *gapB* (lane 2), *tklB* (lane 3), *prkB* (lane 4), *fbpB* (lane 5), and a position outside the form II structural gene region (lane 6); crude extract from wild-type strain HR (lane 7) and purified form II RubisCO (lane 8) were also compared.

At the other conserved regions (B–F), the *R. sphaeroides* FBPase isozymes show considerable homology to diverse FBPase proteins. Mammalian, chloroplast, and *E. coli* FBPase were found to contain a subtilisin-sensitive region that is highly conserved (Marcus et al., 1980, 1987; McGregor et al., 1982). In *R. sphaeroides* FBPase I and II there is a similar region from residues 25 to 38 and 42 to 55, respectively (Table I), that shows homology to previously determined subtilisin-sensitive sequences. Indeed, Gly-42, Asn-48, Gly-51, and Lys-55 of FBPase II and Gly-25, Asn-31, Gly-34, and Lys-38 of FBPase I are invariant in all FBPase enzymes examined to date. Exposed surface loops appear to be the substrate for subtilisin (McGregor et al., 1982), suggesting that this conserved region may play some significant role in either catalysis or regulation of FBPase activity in the cell.

There are several residues that may be important for FBPase activity. For example, there are two histidine residues that are conserved among all previously sequenced FBPase enzymes (His-253 and His-311 of the pig kidney enzyme) (Marcus et al., 1982). Both form I and form II *R. sphaeroides* FBPase also contain these conserved histidine residues (His-218 and

His-277 for form I; His-235 and His-295 for form II). Previous results had indicated that modification of histidine residues with diethylpyrocarbonate resulted in the inactivation of the rabbit liver enzyme (DeMaine & Benkovic, 1980), as well as the enzymes from pig kidney and spinach chloroplasts (Marcus et al., 1987). Since two histidine residues are spectrally perturbed upon addition of known effectors to the rabbit liver enzyme (Scheffler & Fromm, 1986), these facts, taken together, suggest that the conserved histidine residues may be good candidates for future structure–function studies. Indeed, the conserved histidines are separated by 56 amino acid residues (Table I) and a partial sequence of the *A. eutrophus* FBPase indicates there are 59 amino acids separating the two conserved His residues of this enzyme (Kossman et al., 1989). Lastly, a lysine residue (Lys-274) of the pig kidney enzyme was shown to be important for catalysis (Xu et al., 1982). This lysine is conserved among all previously sequenced FBPases, except in *R. sphaeroides*, where an Arg residue is found at posisitons 239 and 255 of the form I and form II enzymes, respectively. It would seem that the overall positive charge of this residue is more important that the absolute specificity of the previously identified lysine residue. It is intriguing that this lysine/arginine residue is found within the large homology domain (region F) that contains the two conservative histidine residues. Further definitive assignments of the role of various conserved amino acids obviously awaits further functional studies. In this connection, the three-dimensional structure of the pig kidney FBPase has recently been solved at 3.0-Å resolution and refined to 2.8-Å resolution (Ke et al., 1989). The binding site for Fru-2,6-P_2 was found to be shared by neighboring monomers of the enzyme, consisting of side-chain atoms of Asn-212, Tyr-244, Tyr-264, and Lys-274 and backbone atoms of Gly-246, Ser-247, Met-248, and Arg-243 from the adjacent subunit. All of these residues are identical or are conservatively replaced in the *R. sphaeroides* enzymes, suggesting that these residues may function to bind Fru-2,6-P_2 in this case as well. Residues thought to form a negative pocket near the Fru-2,6-P_2 binding site of the pig kidney enzyme (i.e., Asp-118, Asp-121, Glu-280, Glu-97, and Glu-98) are conserved in both *R. sphaeroides* isozymes. Finally, residues implicated in the binding site for AMP (Ke et al., 1989) are not conserved in the *R. sphaeroides* isozymes, consistent with the inability of AMP to inhibit the enzyme from the related organism *Rhodopseudomonas palustris* (Springgate & Stachow, 1972). Tyr-113, thought to contribute to AMP-induced UV absorbance changes (Ke et al., 1989), is replaced by a leucine in *R. sphaeroides* FBPase I and II, further suggesting insensitivity to AMP.

Much less is known about PRK, the other unique enzyme of the Calvin cycle, which along with RubisCO ensures that the Calvin cycle functions in a biosynthetic mode. The dissimilarity of the prokaryotic and eukaryotic enzymes is striking and is seemingly related to differences in the way each enzyme is regulated both in vivo and in vitro (Tabita, 1988). The only major regions of conservation, the amino terminal ATP-binding site and the two sequences (Ile/Leu)-Leu-X-X-Glu-Gly-Leu-His and Lys-Ile-X-Arg-Asp-X-X-X-Arg-Gly-X-X-X-Glu may be related to functions carried out by both prokaryotic and eukaryotic enzymes. Lys-68, recently proposed to interact with the γ-phosphoryl moiety of ATP in spinach PRK (Miziorko et al., 1990), is not conserved in the *R. sphaeroides* or *A. eutrophus* PRK primary structure.

Acknowledgments

We thank Bonggeun Lee for generating the hydropathy plots.

References

Buchanan, B. B. (1990) *Annu. Rev. Plant Physiol. 31*, 341–374.

Capon, D. J., Chen., E. Y., Levinson, A. D., Seeburg, P. H., & Goeddel, D. V. (1983) *Nature 302*, 33–37.

Chatterjee, T., Reardon, I., Heinrikson, R. L., & Marcus, F. (1985) *J. Biol. Chem. 260*, 13553–13559.

Cremo, C. R., Grammer, J. C., & Yount, R. G. (1989) *J. Biol. Chem. 264*, 6608–6611.

Davis, R. W., Botstein, D., & Roth, J. R. (1980) in *Advanced bacterial genetics: a manual for genetic engineering*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

DeMaine, M. M., & Benkovic, S. J. (1980) *Arch. Biochem. Biophys. 205*, 308–314.

Falcone, D. L., Quivey, R. G., Jr., & Tabita, F. R. (1988) *J. Bacteriol. 170*, 5–11.

Gibson, J. L., Chen, J.-H., & Tabita, F. R. (1988) *Abstr. VI Int. Symp. Photosynth. Proc.* 252.

Gibson, J. L., & Tabita, F. R. (1985) *J. Bacteriol. 164*, 1188–1193.

Gibson, J. L., & Tabita, F. R. (1986) *Gene 44*, 271–278.

Gibson, J. L., & Tabita, F. R. (1987) *J. Bacteriol. 169*, 3685–3690.

Gibson, J. L., & Tabita, F. R. (1988) *J. Bacteriol. 170*, 2153–2158.

Hallenbeck, P. L., & Kaplan, S. (1987) *J. Bacteriol. 169*, 3669–3678.

Henikoff, S., Haughn, G. W., Calvo, J. M., & Wallace, J. (1988) *Biochemistry 85*, 6602–6606.

Higgins, C. F., Hiles, I. D., Salmond, G. P. C., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W., & Hermodson, M. A. (1986) *Nature 323*, 448–450.

Ish-Horowitz, D., & Burke, J. F. (1981) *Nucleic Acids Res. 9*, 2989–2998.

Joint, I. R., Morris, I., & Fuller, R. C. (1972) *J. Biol. Chem. 247*, 4833–4838.

Jouanneau, Y., & Tabita, F. R. (1986) *J. Bacteriol. 165*, 620–624.

Ke, H., Thorpe, C. M., Seaton, B. A., Marcus, F., & Lipscomb, W. N. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1475–1479.

Klintworth, R., Husemann, M., Salnikow, J., & Bowien, B. (1985) *J. Bacteriol. 164*, 954–956.

Kossmann, J., Klintworth, R., & Bowien, B. (1989) *Gene 85*, 246–252.

Krieger, T. J., Mede-Mueller, L., & Miziorko, H. M. (1987) *Biochim. Biophys. Acta 915*, 112–119.

Kuby, S. A., Palmieri, R. H., Frischat, A., Fischer, A., Wu, H., Maland, L., & Manship, M. (1984) *Biochemistry 23*, 2393–2399.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Marcus, F., Edelstein, I., Nishizawa, A. N., & Buchanan, B. B. (1980) *Biochem. Biophys. Res. Commun. 97*, 1304–1310.

Marcus, F., Edelstein, I., Reardon, I., & Heinrikson, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7161–7165.

Marcus, F., Harrsch, P. B., Moberly, L., Edelstein, I., & Latshaw, S. P. (1987) *Biochemistry 26*, 7029–7035.

Marcus, F., Moberley, L., & Latshaw, S. P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5379–5383.

McGregor, J. S., Hannappel, E., Xu, G.-J., Pontremoli, S., & Horecker, B. L. (1982) *Arch. Biochem. Biophys. 217*, 652–664.

Milanez, S., & Mural, R. J. (1988) *Gene 66*, 55-63.
Miziorko, H. M., Brodt, C. A., & Krieger, T. J. (1990) *J. Biol. Chem. 265*, 3642-3647.
Möller, W., & Amons, R. (1985) *FEBS Lett. 186*, 1-7.
Muller, E. D., Chory, J., & Kaplan, S. (1985) *J. Bacteriol. 161*, 469-472.
Nagata, S., Tanizawa, K., Escki, N., Sakamoto, Y., Oshima, T., Tanaka, H., & Soda, K. (1988) *Biochemistry 27*, 9056-9062.
Porter, M. A., Stringer, C. D., & Hartman, F. C. (1988) *J. Biol. Chem. 263*, 123-129.
Quivey, R. J., Jr., & Tabita, F. R. (1984) *Gene 31*, 91-101.
Raines, C. A., Lloyd, J. C., Longstaff, M., Bradley, D., & Dyer, T. (1988) *Nucleic Acids Res. 16*, 7931-7942.
Rittenhouse, J., Harrsch, P. B., Kim, J. N., & Marcus, F. (1986) *J. Biol. Chem. 261*, 3939-3953.
Roesler, K. R., & Ogren, W. L. (1988) *Nucleic Acids Res. 16*, 7192.
Rogers, D. T., Hiller, E., Mitsock, L., & Orr, E. (1988) *J. Biol. Chem. 263*, 6051-6057.
Ronson, C. W., Nixon, B. T., & Ausubel, F. M. (1988) *Cell 49*, 578-581.
Scheffler, J. E., & Fromm, H. J. (1986) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 45*, 1868.
Siebert, K., Schobert, P., Bowien, B. (1981) *Biochim. Biophys. Acta 658*, 35-94.
Springgate, C. F., & Stachow, C. S. (1972) *Arch. Biochem. Biophys. 152*, 13-19.
Suwanto, A., & Kaplan, S. (1989) *J. Bacteriol. 171*, 5850-5859.
Tabita, F. R. (1980) *J. Bacteriol. 143*, 1275-1280.
Tabita, F. R. (1988) *Microbiol. Rev. 52*, 155-189.
Tabita, F. R., Gibson, J. L., Falcone, D. L., Lee, B., & Chen, J.-H. (1990) *FEMS Microbiol. Rev.* (in press).
Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H., & Numa, S. (1985) *Nature 315*, 242-245.
Trüper, H. G., & Pfennig, N. (1978) in *The Photosynthetic Bacteria* (Clayton, R. K., & Sistrom, W. R., Eds.) pp 19-27, Plenum, New York.
Xu, G.-J., Natalini, P., Suda, H., Tsolas, O., Dzugaj, A., Sun, S. C., Pontremoli, S., & Horecker, B. L. (1982) *Arch. Biochem. Biophys. 214*, 688-694.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103-119.
Zhou, C., Hwang, S.-H., & Jong, A. Y. (1989) *J. Biol. Chem. 264*, 9022-9029.

Human IgE-Binding Protein: A Soluble Lectin Exhibiting a Highly Conserved Interspecies Sequence and Differential Recognition of IgE Glycoforms†,‡

Michael W. Robertson, Keith Albrandt, Daryle Keller, and Fu-Tong Liu*
Medical Biology Institute, Division of Molecular Biology, 11077 North Torrey Pines Road, La Jolla, California 92037
Received February 27, 1990; Revised Manuscript Received May 23, 1990

ABSTRACT: IgE-binding protein (εBP) refers to a protein originally identified in rat basophilic leukemia cells by virtue of its affinity for IgE. It is now known to be a β-galactoside-binding lectin equivalent to carbohydrate-binding protein 35 (CBP 35). More recently, its identity to Mac-2, a macrophage cell-surface protein, has been established. cDNA coding for human εBP has been cloned from a human HeLa cell cDNA library and contains an open reading frame of 750 base pairs encoding a 250 amino acid protein. Like the rat and murine counterparts, the human εBP amino acid sequence can be divided into two domains with the amino-terminal domain consisting of a highly conserved repetitive sequence (YPGXXXPGA) and the carboxyl-terminal domain containing sequences shared by other S-type lectins. The human εBP sequence exhibits extensive homology to murine and rat εBP with 84% and 82% identity, respectively. The homology is particularly striking in the carboxyl-terminal domain where 95% identity is found between human and murine sequences in a stretch of over 70 amino acids. A survey of εBP mRNA expression from several lymphocyte cell lines revealed that the level of εBP transcription may reflect a relationship between cell differentiation and εBP expression. Finally, human εBP was purified from several human cell lines and shown to possess lactose-binding characteristics and cross-species reactivity to murine IgE. Surprisingly, three different human myeloma IgE proteins did not show reactivity to human εBP. However, after neuraminidase treatment of each human IgE, pronounced binding to εBP was observed, thereby indicating that only specific IgE glycoforms can be recognized by εBP.

IgE-binding protein (εBP)[1] (Liu, 1990) refers to a M_r 31 000 protein with IgE-binding activity, originally identified in rat basophilic leukemia (RBL) cells (Liu & Orida, 1984; Liu et al., 1985). Cloning and sequencing of cDNA revealed a novel sequence with several interesting structural features (Liu et al., 1985; Albrandt et al., 1987). The protein is composed of 2 domains: the amino-terminal domain contains tandem repeats of a highly conserved sequence of 9 amino acids [Tyr-Pro-Gly-(Pro/Gln)-(Ala/Thr)-(Pro/Ala)-Pro-Gly-Ala]; the

† This work was supported by National Institutes of Health Grants AI-19747 and AI-20958 to F.-T.L.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02921.
* To whom correspondence should be addressed at the Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, 10666 N. Torrey Pines Rd., La Jolla, CA 92037.

[1] Abbreviations: εBP, IgE-binding protein; RBL, rat basophilic leukemia; CBP 35, carbohydrate-binding protein 35; RL-29, rat lung lectin 29; HL-29, human lung lectin 29; hnRNP, heterogeneous nuclear ribonucleoprotein; Mac-2, murine macrophage cell-surface protein 2.

0006-2960/90/0429-8093$02.50/0