Volume 17 Number 16 1989 Nucleic Acids Research

Wheat phosphoglycerate kinase: evidence for recombination between the genes for the chloroplastic and cytosolic enzymes

Marian Longstaff+, Christine A.Raines§, Eileen M.McMorrow[1], J.William Bradbeer[1] and Tristan A.Dyer*

Institute of Plant Science Research (Cambridge Laboratory), Maris Lane, Cambridge CB2 2JB and [1]Division of Biosphere Sciences, King's College London, Campden Hill Road, Kensington, London W8 7AH, UK

Received May 26, 1989; Revised and Accepted July 28, 1989 EMBL accession nos X15232, X15233

ABSTRACT

We have isolated and sequenced cDNA clones containing the entire coding region of both the chloroplast and cytosolic versions of phosphoglycerate kinase from wheat. Comparison of these sequences reveals a higher than expected level of similarity between the nucleic acids and encoded proteins. Analysis of this data in relation to that for phosphoglycerate kinase sequences of mammals, prokaryotes and yeasts suggests that the wheat genes have recombined. This has resulted in the chloroplast and cytosolic kinases being more similar to each other than would be expected if the chloroplast enzyme had evolved directly from that of a prokaryotic progenitor.

INTRODUCTION

There is now a great deal of evidence from a variety of sources to support the hypothesis that chloroplasts have evolved from prokaryotic endosymbionts (1,2). Some of this evidence is derived from a comparison of the isoenzymes of the chloroplast. For example, the cytosolic version of glyceraldehyde phosphate dehydrogenase (GAPDH) has features in common with the equivalent enzyme in mammals and yeast whilst the chloroplast enzyme is prokaryotic in nature (3,4). This is so despite the fact that both enzymes are now nuclear-encoded due presumably to the transfer of genetic material from the ancestral endosymbiont into the nucleus.

We describe here a similar comparison between the chloroplastic and cytosolic version of the enzyme phosphoglycerate kinase (PGK, EC 2.7.2.3). It is an enzyme which occurs in all cellular organisms. The chloroplastic version in plants is involved in the photosynthetic carbon reduction cycle while that of the cytosol is required, as in other organisms, for ATP generation in glycolysis. It is also involved in gluconeogenesis and in a series of reactions which shuttle reducing and phosphorylation potential across the chloroplast envelope (5) via the phosphate translocator (6).

The chloroplast and cytosolic PGKs are distinct from one another both with respect to their isoelectrofocussing properties (7) and also in their responses to substrate and Mg^{2+} levels (8). However, they strongly cross react with each other's antiserum suggesting that they have epitopes in common (E.M.M. and J.W.B., manuscript in preparation).

Although a large number of PGK sequences have been published (9; see Figure 4), no plant sequences have been reported previously. We present here the full coding sequences for both isoenzymes of wheat. Comparisons of these with each other and of the derived amino acid sequences with those of PGKs of other organisms reveals that the sequence of this enzyme is highly conserved. The most interesting observation, however, is that the sequences of the plant enzymes resemble each other more closely than they do those

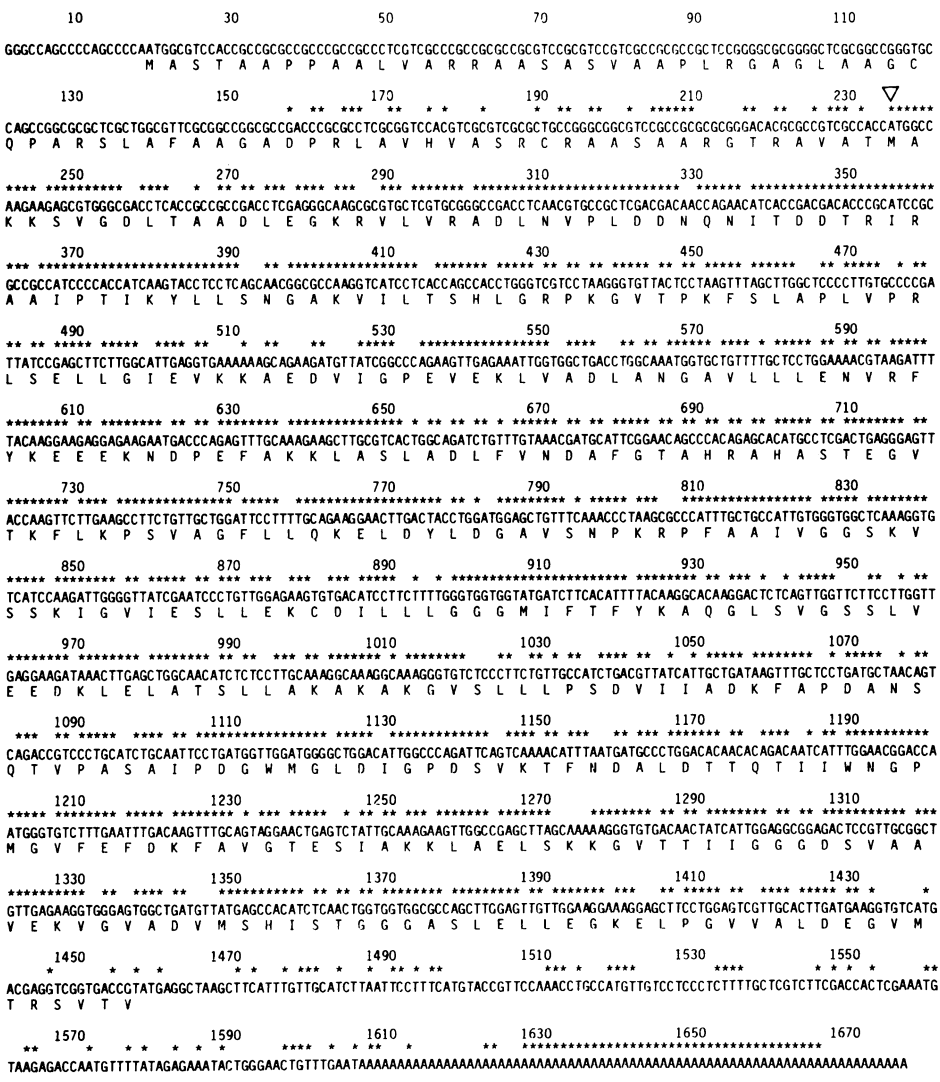

Fig. 1. Sequence of the cDNA containing the coding sequence of wheat chloroplast PGK. The stars above the sequence show where the sequence matches that of the cDNA for the cytosolic enzyme. The inverted triangle shows where the starting codon of the cytosolic enzyme lines up.

of comparable enzymes of other organisms. This suggests that their genes have evolved in an unusual and complex manner.

MATERIALS AND METHODS

cDNA library construction and screening

Using an affinity column procedure (10) highly purified PGK was isolated from barley and, using DEAE-Sephacel columns, the bulk separation of purified chloroplastic and

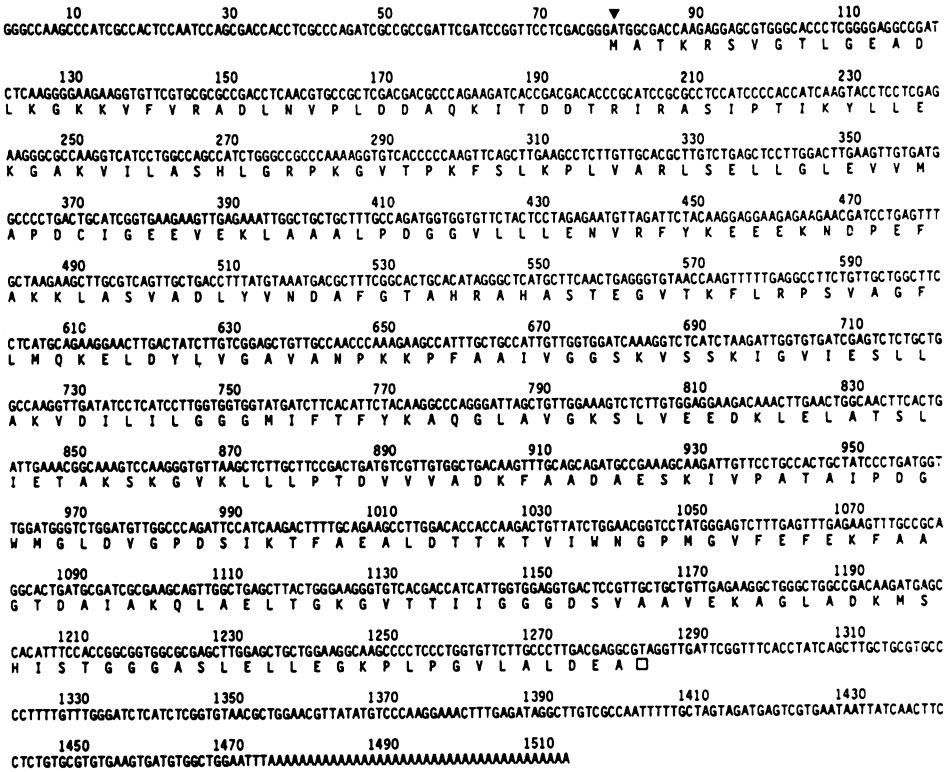

Fig. 2. Sequence of the cDNA containing the coding sequence of wheat cytosolic PGK

cytosolic enzymes was achieved (11, E.M.M. and J.W.B., manuscript in preparation). Final purification of each was on ATP-Sepharose columns and antibodies were raised in rabbits.

The methods used for the construction and screening of an expression library in λgtll and other methods were essentially the same as those described previously (12). Segments with high G+C content were sequenced using 7-deaza dGTP.

RESULTS

Nucleic acid sequences

The nucleic acid sequences of the cDNA clones encoding the chloroplastic and cystolic PGK's of wheat are shown in Figures 1 and 2. The study unequivocally establishes that both enzymes are nuclear encoded and are synthesized in the cytosol. This confirms earlier deductions that were made in this regard from experiments in which selective inhibitors of protein synthesis were used (13, 14). The cDNA of the chloroplastic enzyme contains 1676 base pairs which includes an oligo dA region 70 nucleotides long and Northern blot analysis (data not shown) suggests that it is close to full length. The sequence shown for the cDNA of the cytosolic enzyme was a composite of that derived from two clones, one with the 5′ sequence and extending 104 bp downstream of the coding sequence and the other of the 3′ flanking region including the oligo dT tail and overlapping the 5′ clone

```
          10                        30                  50                  70
PGK MASTAAPPAALVAR......R..AASASVAAPLRGAGLAAGCQPAR.SLAFAAGADPRLAVHVASRCRAASAAR.G..TRAVAT
FBP MA..AA.TTTTS.R.PLLLSRQQAAASS....LQCR.LPRR..PG.SSL.FA.GQGQ......AS.............TPNVRC
PRK MA.FCSPHTTTSLRSPCTTI.....................P..NSG.FRQNQVIFFTTRS.SR.R..SNTRHGARTFQVSC
    **                                        *   *  ** *                    *  ******
```

Fig. 3. Sequence of the wheat chloroplast PGK transit peptide aligned with those of other wheat enzymes of the PCR cycle. Asterisks indicate those amino acids which occur in the three homology blocks proposed to exist in transit peptides by Karlin-Neumann and Tobin (38). The chloroplast fructose-1,6-bisphosphatase (FBP) sequence was described by Raines et al.(12) and that of phosphoribulokinase (PRK) by Raines et al. (15).

by 91 bp. As the clones were identical in the overlapping region and at the 5′ end (the first 166 bp were identical) it was assumed that they were both derived from the same gene sequence. A comparison of the chloroplastic and cytosolic coding sequences indicates that they are highly homologous (77% identity—Figure 1). However, the homology between the sequences ends immediately outside the coding region.

Deduced amino acid sequences of the chloroplast PGK

The coding sequence of the cDNA clone for the precursor peptide of the chloroplastic enzyme contains 480 codons and has two coding segments. One is for the mature enzyme and the other for the transit peptide necessary for directing the precursor of the enzyme into the chloroplast from its site of synthesis in the cytosol. A putative cleavage site for the removal of the transit peptide is between threonine-72 and methionine-73. This designation of cleavage site is based on an alignment of the deduced sequence with the N-terminal sequences of other PGK's (Figure 4). Assuming that the N-terminal amino acid of the mature protein is methionine-73, then the mature protein contains 408 residues and has a transit peptide of 72 amino acids. This would give a precursor molecular weight of 61,257 and a molecular weight of 42,992 for the mature protein. Both the chloroplast mature protein and the cytosolic isoenzyme had been estimated to have a molecular weight of between 35,000 and 38,000 from their physical properties (11, E.M.M. and J.W.B., manuscript in preparation).

The transit peptide of the chloroplast PGK resembles those of other enzymes of the photosynthetic carbon reduction cycle with respect to their general features (Figure 3—see also 15). It contains a high proportion of uncharged amino acids such as 27 alanines, 6 serines, 5 leucines and 3 threonines. The overall charge is positive as there are 9 arginine and one histidine residues balanced by the negative charge of a single glutamate. Nine of the amino acid residues are hydroxylated (serine and threonine). It is still unclear, however, as to what particular characteristics of this peptide is responsible for determining its specificity for chloroplasts and which facilitates the transport of the precursor and its processing (16).

Deduced amino acid sequence of the cytosolic PGK

The cDNA for the cytosolic enzyme contains 1512 base pairs including a segment of 39 nucleotides which corresponds to the poly A tail of the mRNA. The encoded protein contains 401 residues. This would give a molecular weight of 42,122.

Codon usage

For reasons that we do not yet understand, there is usually a strong bias in graminaceous plants towards the use of G or C in the third position of codons of nuclear-encoded chloroplast proteins. For example, in wheat the genes for chloroplastic fructose-1,6-bisphosphatase, phosphoribulokinase, Rubisco small subunit and a LHCP II

Table I. Comparison of codon usage in the coding sequences of cytosolic and chloroplastic PGK of wheat and GAPDH of maize

Codon	PGK Cyt	PGK Chl	GAPDH Cyt	GAPDH Chl	Codon	PGK Cyt	PGK Chl	GAPDH Cyt	GAPDH Chl
Ala CGG	14	2	0	39	Leu TTG	19	28	5	0
A	16	38	7	0	A	2	2	0	0
T	37	29	48	3	CTG	21	20	20	30
C	33	31	45	48	A	4	0	0	0
					T	29	28	25	3
Arg AGG	33	10	54	24	C	25	22	50	67
A	11	20	18	0					
CGG	0	10	5	25	Lys AAG	92	83	86	100
A	0	10	0	0	A	8	17	14	0
T	0	10	0	0					
C	56	40	11	52	Phe TTT	53	53	67	0
					C	47	47	33	100
Asn AAT	33	25	23	11					
C	67	75	77	89	Pro CCG	11	6	17	29
					A	28	39	17	6
Asp GAT	40	52	27	0	T	44	44	33	0
C	60	48	73	100	C	17	11	33	65
Cys TGT	0	100	0	0	Ser AGT	0	3	8	0
C	100	0	100	100	C	29	33	29	67
					TCG	0	10	17	13
Gly GGG	10	8	3	18	A	24	20	4	0
A	17	32	10	6	T	24	20	0	0
T	47	37	47	3	C	24	23	42	20
C	25	24	37	73					
					Thr ACG	9	5	5	26
Gln CAG	100	80	100	89	A	4	38	18	0
A	0	20	0	11	T	39	24	27	0
					C	48	33	50	74
Glu GAG	61	58	87	100					
A	39	42	13	0	Tyr TAT	40	0	33	0
					C	60	100	66	100
His CAT	75	75	38	0					
C	25	25	62	100	Val GTG	22	28	11	27
					A	6	10	0	0
Ile ATA	0	0	0	0	T	53	41	42	0
T	29	41	27	0	C	19	21	47	73
C	71	59	73	100					

gene, have G+C contents of 93, 63, 87 and 75% respectively. The same bias is not found in the genes for most of the cytosolic proteins of the same plants. For example, some wheat storage protein genes have a G+C content of about 45%. We have therefore examined the codon usage in the wheat chloroplastic and cytosolic PGK coding sequences to see whether this would throw any light on their evolution (Table I) and compared our results with those of for the coding sequences of chloroplastic and cytosolic GAPDH of maize (17). Clearly, the same high G+C content is not found for the wheat PGK and

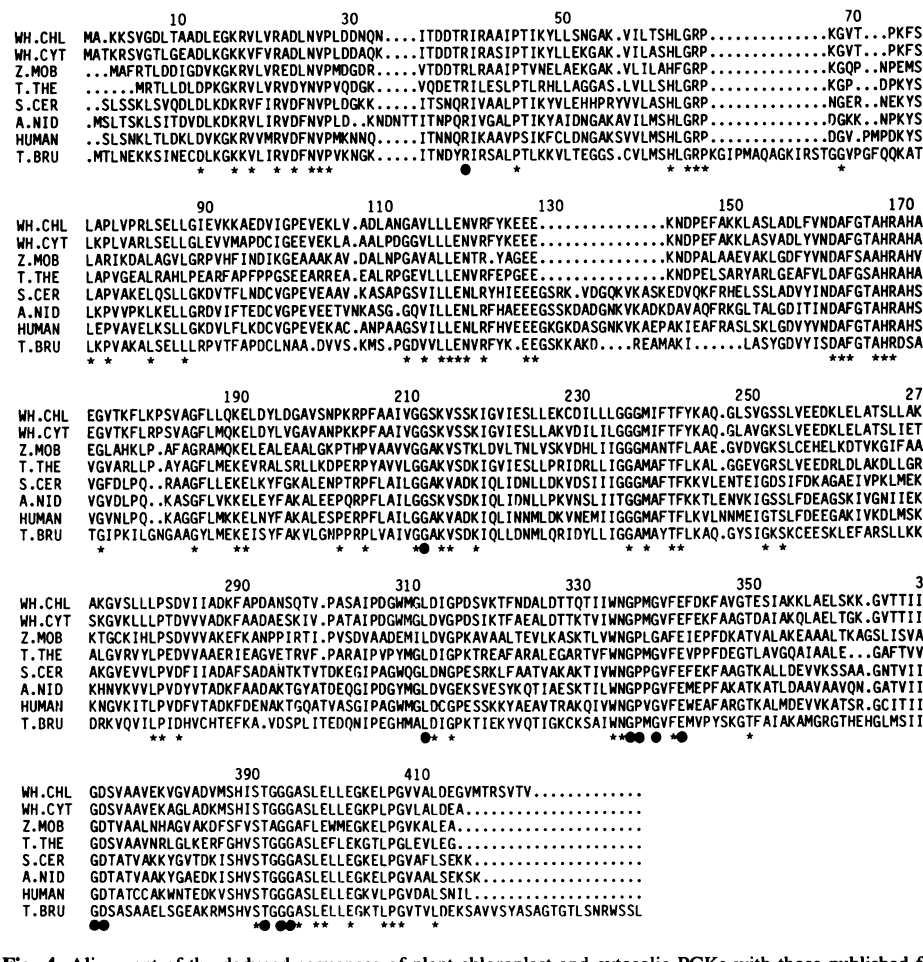

Fig. 4. Alignment of the deduced sequences of plant chloroplast and cytosolic PGKs with those published for other organisms. The closed circles show conserved residues that are at the putative ligand-binding sites and the residues shown with asterisks are also conserved in all the sequences (published and unpublished) of which we are aware. The sequence data was derived from the following sources: *Zymomonas mobilis* (Z. MOB): (39); *Thermus thermophilus* (T. THE): (40); *Saccharomyces cerevisiae* (S. CE): (20); *Aspergillus nidulans* (A. NID) (41); human: (42); *Trypanosoma brucei* (T. BRU): (43).

in fact the codon usage in the chloroplastic and cytosolic coding sequences is rather similar (53 and 58% G or C in the third position respectively as opposed to 97 and 67% in the equivalent maize GAPDH sequences). However, if the composition of codons in the chloroplastic coding sequences is examined (Figure 1) it can be seen that all but one of the codons in the transit peptide and in the amino-terminal portion of the mature protein up to codon 64 have either G or C in their third positions. A similar situation is found in the region coding for the amino-terminal portion of the cytosolic enzyme to approximately the same position. The percentage G+C in the remaining codons of both sequences is 46 and 51% respectively. It seems highly unlikely that such a skewed distribution of codon

Table II. Identities matrix, expressed as a percentage of the total number of sites, occupied by identical amino acids in pairwise comparisons of PGK (our data) and GAPDH sequences (4, 17). The values boxed with dashes are the values for cytosolic enzymes and the dots highlight comparisons of the plant enzymes with those of other organisms.

	PGK								
	1	2	3	4	5	6	7	8+	9Δ
1. Chloroplast		82	53	52	50	49	47	48	58
2. Cytosol	49		52	52	51	52	48	48	59
3. Yeast	50	68		68	66	42	39	40	51
4. *Aspergillus*	-	-	-		64	43	42	40	50
5. Human	49	67	-	-		43	41	39	47
6. *Thermus*	56	45	50	-	46		45	41	50
7. *Zymomonas*	-	-	-	-	-	-		46	49
8. *E. coli**	74	64	69	-	61	49	-		46
9. *Bacillus*	64	50	53	-	51	60	-	-	
				GAPDH					

*The GAPDH of *E. coli* appears to be somewhat anomalous - see (4).
+*E. coli* PGK sequence was determined by Alefounder and Perham (44).
ΔThe *Bacillus stearothermophilus* PGK sequence was communicated to us by Dr. J. Littlechild.

usage could have arisen by chance in the corresponding parts of both sequences. Furthermore, the regions with a high G+C content are also those in which there is a very high degree of overall sequence homology (Figure 1). However, it should be kept in mind that there might be other reasons for the high G+C bias; there is a strong preference for such codons in the more highly expressed genes of *Drosophila*, for example (18).

Sequence alignments

A number of sequences have now been determined for various PGK's and we were able to compare the plant sequences with those of five prokaryotes and nine other eukaryotes. Most of those which have been published are shown in the alignment in Figure 4. The positioning of amino acids in the alignments was carried out with reference to the known conserved regions of the molecule to ensure that corresponding regions were appropriately placed. From this alignment it can be seen that in several regions there is absolute conservation of sequence. We have examined the plant chloroplastic and cytosolic sequences for amino acids in positions characteristic of either prokaryotic or eukaryotic organisms or that are unique to plants. The most outstanding feature of the plant proteins is that in both there is no 'nose' (amino acids 128 to 142—see Figure 5) which is found in all eukaryotes but in none of the prokaryotes so far studied (J.A. Littlechild, G. Davies, D. Bowen, L. Hall and H.C. Watson, manuscript in preparation). However, the cytosolic version is unusual in containing a cysteine at 97 as in other eukaryotes but which is absent from all other proteins that don't have the nose.

Both sequences also possess other prokaryotic features, such as the presence of aspartate

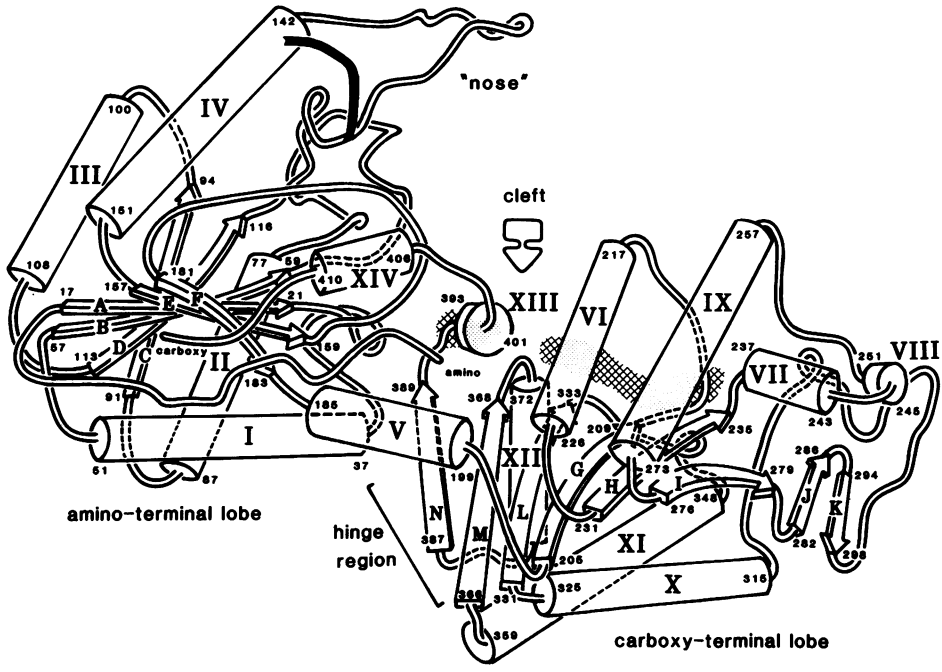

Fig. 5. Schematic representation of the open form of yeast PGK. The thick solid line represents the major deletion referred to in the text (of the 'nose') which occurs in both the chloroplast and cytosolic enzymes. Other differences between the plant and yeast enzymes are not indicated. The diagram is redrawn from that of (20), and the numbering of residues is that of the yeast enzyme. The shaded areas show putative ligand binding sites in the cleft.

at positions 35 and 36, threonine at 37, leucine at 49, lysine at 56 and 67, arginine at 90, leucine at 109, lysine at 142, aspartate at 144, leucine at 194, valine at 220 and 250, aspartate at 258, valine at 283 and alanine at 373 and 374. In other positions, both sequences contain amino acids identical to those of eukaryotes. These are lysine at position 47, glutamate at 103 and 125, phenylalanine at 133, leucine at 137, tyrosine at 181, glycine at 306 and leucine at 397. In 27 positions the wheat isoenzymes only resemble each other.

In order to get an estimate of the relative similarities between the various primary sequences, we compared several in a pairwise fashion and scored the percentage number of positions in which there are identical residues (Table II). From this it is clear that the plant sequences resemble each other to a far greater extent than any others and that they both resemble those of prokaryotes and other eukaryotes to a similar extent. This is in marked contrast to what was found with the maize GAPDH.

Copy number, location and expression of the PGK genes

In a study to be reported in full elsewhere (19) we show that there is a single copy of each PGK gene per haploid genome and that the genes of the chloroplastic and cytosolic enzymes are on the group 1 and 6 chromosomes of wheat. Using the same mRNA preparations that were used to study the effect of light on the accumulation of mRNA for chloroplastic FBP (12) and for PRK (15), we found that the accumulation of mRNA for the chloroplast PGK is light stimulated and found only in tissues which develop the photosynthetic apparatus, whereas the mRNA for the cytosolic enzyme is found in all tissues

of wheat plants in similar quantities (results not shown). These observations clearly show that, despite the appreciable homology between their coding sequences, the chloroplastic and cytosolic genes are quite independent of one another with regard to location and expression.

DISCUSSION

The reaction which PGK catalyses is reversible

$$\text{3-phosphoglycerate} + \text{Mg ATP}^{2-} \rightleftharpoons \text{1,3-diphosphoglycerate} + \text{Mg ADP}^{-}$$

and the flux is predominantly towards diphosphoglycerate formation. Crystallographic studies have shown that the PGK of yeast (20, Figure 5) and horse (21) have a pronounced bilobal structure. The two lobes, which are of approximately equal size, are connected by a flexible hinge. The substrates bind in the cleft between the two lobes, the phosphoglycerate to one face and the adenosine phosphate to the other. All the published work to date on the structure of PGK has been performed on the open form of the enzyme. It is conceivable that there are other regulatory sites for the two substrates which have not yet been identified and this might serve to explain the non-linear kinetic data observed for both the yeast (22) and the green plant enzymes (8, and in preparation). The binding of substrates appears to trigger a movement of the lobes relative to one another so that the cleft closes bringing the substrates close together while providing the non-aqueous environment in which the phosphoryl transfer reaction takes place. The sequence comparison presented here (Figure 4) demonstrates that most of the highly conserved regions in these proteins are in loops which make up the cleft faces. There are many differences elsewhere in the sequences although a large degree of conservation does exist. There can, however, be no doubt whatsoever as to the homology and common ancestry of this enzyme in the organisms studied. Furthermore, several other types of kinase have been shown to share similar structural motifs (23). The models for the chloroplast and cytosolic isoenzymes from wheat are being constructed using the yeast model as a template and will be described in detail elsewhere (E.M.M., M.L., C.A.R., T.A.D. and J.W.B., manuscript in preparation). The differences in amino acid residues outside the highly conserved regions reflect the differences in kinetic and immunological properties already observed beween both the plant and yeast enzymes and also between the cytosolic and chloroplast isoenzymes with respect to their regulation. At this stage we cannot infer which particular amino acid differences affect these parameters.

In view of what has been found previously with GAPDH (3, 4) in which there is clearcut evidence for the divergent evolution of the chloroplastic and cytosolic enzymes it was very surprising to find that there is a much higher degree of homology between the equivalent PGK's than would be expected if they had evolved in the same sort of way.

Our results are also quite clear in relation to three further points. The first is that the primary sequence of the chloroplastic enzyme resembles the prokaryotic enzymes *less* and eukaryotic enzymes *more* than might have been expected. The second is that the primary sequence of the cytosolic enzyme resembles the prokaryotic enzymes *more* and the eukaryotic enzymes *less* than might have been expected. *Both* primary sequences resemble those of each of these groups to a *similar* extent. In immunotitrations the chloroplast antiserum showed much less cross-reaction with the yeast enzyme than did the cytosolic antiserum (E.M.M. and J.W.B., unpublished data). This suggests that the yeast enzyme shares more common epitopes with the cytosolic enzyme that it does with the chloroplast

enzyme. It is therefore important to compare the 3-dimensional structures of the isoenzymes and not just the primary sequences in order to explore those differences which will explain the different regulatory phenomena.

The explanation of the sequence data would seem to be that the genes for these enzymes, like the GAPDH, may have started evolving separately, the cytosolic one from an ancestral eukaryote and the chloroplastic one from an endosymbiotic prokaryote. However, there has subsequently been substantial recombination between the two genes so that now they are very similar to one another and code for proteins intermediate in composition between those of either of their progenitors.

Additional data also fits in with this explanation. Both enzymes have some prokaryotic features (e.g. the 'nose' is missing—residues 128−142 of yeast; J.A. Littlechild, G. Davies, D. Bowen, L. Hall and H.C. Watson, manuscript in preparation) and in some positions the amino acids found are more similar to those in the majority of prokaryotes while in other positions they are both more similar to eukaryotes. Furthermore codon usage is similar in their genes. In the amino-terminal segment of both sequences there is a very high G+C bias in the third position of codons and this feature is usually only found in the genes of chloroplastic enzymes.

There is abundant evidence from studies of various organsims for homologous recombination between duplicated chromosomal sequences. Proximity and length of homologous sequences are important factors in determining the frequency of recombination (24, 25). Consequently, recombination occurs most often between sequences with extensive homology and is predominantly an *intra*chromosomal event. However, *inter*chromosomal recombination does also take place. Hence, the mechanism that we are proposing to explain why the PGK genes are so similar to each other has previously been observed but in a different context.

RFLP studies have shown that the wheat genes for the chloroplast and cytosolic enzymes are on the group 1 and 6 chromosomes respectively (19). Evidence of a specific interaction of genetic material between these chromosomes comes from groups of studies of the genes for the gliadin seed storage proteins (26). In wheat, genes for this type of protein are found on the short arms of these groups of chromosomes. However, in barley (which is a close relative of wheat) there is no evidence for the presence of these genes on the group 6-equivalent chromosomes although, as in wheat, they are found on the group 1-equivalent chromosomes. This suggests that there could have been a duplicative transfer of these genes between the group 1 and 6 chromosomes of wheat quite recently. This is also so for the genes coding for high-molecular-weight ribosomal RNA (27).

Why then has recombination apparently taken place in the genes for this enzyme but not between those for others such as GAPDH? The answer may in part be that the enzyme is very highly conserved in all organisms and especially in the cleft region (9). The sequences coding for these regions would therefore remain similar which would facilitate recombination. The other feature is that there is interchangeability between the two domains of the enzyme (28) showing that chimaeric enzymes are viable.

There is precedent for a similar situation having arisen in yeast with regard to the mitochondrial and non-mitochondrial forms of citrate synthase (29). The two forms of the enzyme are much more similar to each other than they are to either *E. coli* or the pig heart enzymes. There are also several instances, however, where a single gene may code for both cytosolic and mitochondrial enzymes [e.g. yeast fumarase: (30); histidinyl-tRNA synthetase: (31); and possibly alpha-isopropylmalate synthase: (32)]. In each case

distinct mRNAs that either do or do not encode a mitochondrial targeting sequence are apparently transcribed from the same gene.

Triose phosphate isomerase is another plant enzyme which has many of the same characteristics as PGK and which has a chloroplast and cytosolic isoenzyme. Like PGK there is also cross reactivity between the respective antibodies to the chloroplastic and cytosolic enzymes (33) and the enzyme, in general, is very highly conserved (34). It seems possible therefore that the evolution of this enzyme could have been similar to that of PGK.

An alternative and more direct explanation of the similarity between the chloroplastic and cytosolic PGK sequences is that they are encoded by related genes which arose from a common progenitor by duplication. The fact that they both have a high G+C content, which is especially apparent in their 5′ regions, could be because CpG-rich islands are often associated with eukaryotic genes (35, 36). However, in the absence of sequence information about the non-coding flanking regions, we cannot comment further on this particular point here. This senario for the evolution of the PGK genes would require that either the original cytosolic or chloroplastic gene was lost subsequent to the duplication of the other. Also, it does not explain why both sequences now have some prokaryotic and eukaryotic features rather than being predominantly like one.

From all these studies it would seem that isoenzymes (and even those which have different subcellular locations) could have evolved in one of several different ways and illustrates how opportunistic gene evolution may be. Therefore, it may not be possible to make any sweeping generalizations as to how the chloroplastic and cytosolic isoenzymes have evolved.

One of the most important points yet to be resolved about PGK and other enzymes involved in carbon fixation is what influence each could have on the control of the metabolic pathways in which they function. Such control could be exerted either through changes in their amount or in the regulation of their activity. In this regard it is interesting to note that the chloroplastic PGK is nine times more abundant that the cytosolic one in barley leaves (E.M.M. and J.W.B., manuscript in preparation). Now that we have the isolated coding sequences, we are in a position to make transgenic plants with altered anounts of these isozymes. Using such plants it should be possible to determine the control coefficients of the PGKs by measuring the flux of metabolic intermediates in plants grown in a range of environmental conditions (37). This study therefore provides an important step towards achieving a greater insight into the control of photosynthetic carbon reduction.

ACKNOWLEDGEMENTS

We are very grateful to Dr. H.C. Watson and Dr. J.A. Littlechild for helping us make alignments of PGK sequences, for letting us have a copy of the *Bacillus stearothermophilus* sequence prior to its publication and for introducing us to the complexities of PGK structure. We are also grateful to Dr. R.N. Perham and to Dr. P.R. Alefounder for making their *E. coli* sequence and alignments available to us before publication. AFRC research grant no. PG 29/49 is gratefully acknowledged (J.W.B.). M.L. and C.A.R. were supported by AFRC New Initiative grants.

*To whom correspondence should be addressed

Present addresses: [+]Sainsbury Laboratory, John Innes Institute, Colney Lane, Norwich NR4 7UH and [§]Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

REFERENCES

1. Margulis, L. (1970) Origin of Eukaryotic Cells. Yale University Press, New Haven, Connecticut.
2. Weeden, N.F. (1981) J. Mol. Evol., **17**, 133–139.
3. Shih, M-C., Lazar, G. and Goodman, H.M. (1986) Cell **47**, 73–80.
4. Martin, W. and Cerff, R. (1986) Eur. J. Biochem., **159**, 323–331.
5. Stocking, C.R. and Larson, S. (1969) Biochem. Biophys. Res. Commun., 37, 287–282.
6. Heber, U. and Heldt, H.W. (1981) Annu. Rev. Plant Phyiol., **32**, 139–168.
7. Anderson, L.E. and Advani, V.R. (1970) Plant Physiol., **45**, 583–585.
8. McMorrow, E.M. and Bradbeer, J.W. (1987) In Biggins.J. (ed), Progress in Photosynthesis Research. Martinus Nijhoff, Dordrecht, Netherlands, Vol. III, 4, pp. 333–336.
9. Mori, N., Singer-Sam, J. and Riggs, A.D. (1986) FEBS Lett., **204**, 313–317.
10. Kuntz, G.W.K., Eber, S., Kessler, W., Krietsch, H. and Krietsch, W.K.G. (1979) Eur. J. Biochem., **85**, 493–501.
11. McMorrow, E.M. (1987) Ph.D. Thesis, University of London (U.K.).
12. Raines, C.A., Lloyd, J.C., Longstaff, M., Bradley, D. and Dyer, T.A. (1988) Nucleic Acids Res., **16**, 7931–7942.
13. Ireland, H.M.M. and Bradbeer, J.W. (1971) Planta, **96**, 254–261.
14. Bradbeer, J.W. (1981) In Hatch, M.D. and Boardman, N.K. (eds), The Biochemistry of Plants. Academic Press, New York, Vol. 8, pp. 423–472.
15. Raines, C.A., Longstaff, M., Lloyd, J.C. and Dyer, T.A. (1989) Mol. Gen. Genet., in press.
16. Keegstra, K. and Baurle, C. (1988) BioEssays, **9**, 15–19.
17. Brinkmann, H., Martinez, D., Quigley, F., Martin, W. and Cerff, R. (1988) J. Mol. Evol., **26**, 320–328.
18. Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F. (1988) Mol. Biol. Evol., **5**, 704–716.
19. Chao, S., Raines, C.A., Longstaff, M., Sharp, P.J., Gale, M.D. and Dyer, T.A. (1989). Molec. Gen. Genet., in press.
20. Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.F., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J. and Kingsman, S.M. (1982) EMBO J., **1**, 1635–1640.
21. Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merrett, M. and Phillips, A.W. (1979) Nature, **279**, 773–777.
22. Larsson-Raznikiewicz, M. (1973) Arch. Biochem. Biophys., **158**, 754–762.
23. Anderson, C.M., Zucker, F.H. and Steitz, T.A. (1979) Science, **204**, 375–380.
24. Mikus, M.D. and Petes, T.D. (1982) Genetics, **101**, 369–404.
25. Liskay, R.M., Letsou, A. and Stachelek, J.L. (1987) Genetics, **115**, 161–167.
26. Law, C.N. and Payne, P.I. (1983) J. Cereal Sci., **1**, 79–93.
27. Gill, B.S. and Appels, R. (1988) Plant Syst. Evol., **160**, 77–89.
28. Mas, M.T., Chen, C.Y., Hitzeman, R.A. and Riggs, A.D. (1986) Science, 233, 788–790.
29. Rozenkrantz, M., Alam, T., Kim, K-S., Clark, B.J., Srere, P.A. and Guarente, L.P. (1986) Mol. Cell Biol., **6**, 4509–4515.
30. Wu,M. and Tzagoloff,A. (1987) J. Biol. Chem., **262**, 12275–12282.
31. Natsoulis, G., Hilger, F and Fink, G.R. (1986) Cell, **46**, 235–243.
32. Beltzer, J.P., Chang, L.-F.L., Hinkkanen, A.E. and Kohlhaw, G.B. (1986) J. Biol. Chem., **261**, 5160–5167.
33. Kurzok, H.-G. and Feierabend, J. (1984) Biochim. Biophys. Acta, **788**, 222–233.
34. Marchionni, M. and Gilbert, W. (1986) Cell, **46**, 133–141.
35. Bird, A.P. (1986) Nature **321**, 209–213.
36. Antequera, F., Bird, A.P. (1988) EMBO J. **7**, 2295–2299.
37. Kacser, H. and Porteous, J.W. (1987) TIBS, **12**, 5–14.
38. Karlin-Neumann, G.A. and Tobin, E.M. (1986) EMBO J., **5**, 9–13.
39. Conway, T., Ingram, L.O. (1988) J. Bacteriol, **170**, 1926–1933.
40. Bowen, D, Littlechild, J.A., Fothergill, J.E., Watson, H.C. and Hall, L. (1988) Biochem. J. **254**, 509–517.
41. Clements, J.M. and Roberts, C.F. (1986) Gene, **44**, 97–105.
42. Michelson, A.M., Markham, A.F. and Orkin, S.H. (1983) Proc. Natl. Acad. Sci. USA., **80**, 472–476.
43. Osinga, K.A., Swinkels, B.W., Gibson, W.C., Borst, P. Veeneman, G.H., van Boom, J.H., Michels, D.A. and Opperdoes, F.R. (1985) EMBO J., **4**, 3811–3817.
44. Alefounder, P.R. and Perham, R.N. (1989) Molec. Microbiol., **3**, 723–732.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.