J Mol Evol (1992) 34:449–464

Journal of Molecular Evolution

Comparison of Proteins of ADP-Glucose Pyrophosphorylase from Diverse Sources

Brian J. Smith-White and Jack Preiss

Department of Biochemistry and Graduate Program in Genetics, Michigan State University, East Lansing, MI 48824, USA

Summary. The primary structures of 11 proteins of ADP-glucose pyrophosphorylase are aligned and compared for relationships among them. These comparisons indicate that many domains are retained in the proteins from both the enteric bacteria and the proteins from angiosperm plants. The proteins from angiosperm plants show two main groups, with one of the main groups demonstrating two subgroups. The two main groups of angiosperm plant proteins are based upon the two subunits of the enzyme, whereas the subgroups of the large subunit group are based upon the tissue in which the particular gene had been expressed. Additionally, the small subunit group shows a slight but distinct division into a grouping based upon whether the protein is from a monocot or dicot source. Previous structure–function studies with the *Escherichia coli* enzyme have identified regions of the primary structure associated with the substrate binding site, the allosteric activator binding site, and the allosteric inhibitor binding site. There is conservation of the primary structure of the polypeptides for the substrate binding site and the allosteric activator binding site. The nucleotide sequences of the coding regions of the genes of 11 of these proteins are compared for relationships among them. This analysis indicates that the protein for the small subunit has been subject to greater selective pressure to retain a particular primary structure. Also, the coding region of the precursor gene for the small subunit diverged from the coding region of the precursor gene for the large subunits slightly prior to the divergence of the two coding regions of the genes for the two tissue-specific large subunit genes.

Offprint requests to: J. Preiss

Key words: ADP-glucose pyrophosphorylase — Protein sequence comparisons

Introduction

ADP-glucose (ADPGlc) pyrophosphorylase catalyzes the synthesis of ADPGlc from ATP and glucose-1-phosphate (Espada 1962). The enzyme is present in nonphotosynthetic and photosynthetic prokaryotes and photosynthetic eukaryotes (Preiss 1973). In bacteria and in plants the enzyme is considered to be the prime regulatory step for biosynthesis of bacterial glycogen (Preiss 1984; Preiss and Romeo 1989) and of starch (Preiss 1982, 1988). It is not known whether the higher plant enzyme evolved directly from the enzyme present in a photosynthetic prokaryote, which is related to the enzyme in a nonphotosynthethic prokaryote or whether these three groups represent one or more examples of convergent evolution.

The enzyme from all sources is found to be allosterically controlled, but the specificity toward the effector depends upon the source of the enzyme. Enzyme from enteric bacteria is activated by fructose 1,6-biphosphate (FBP) and inhibited by AMP (Preiss et al. 1966), whereas the enzyme from almost all higher plant sources is activated by 3-phosphoglyceric acid (3PGA) and inhibited by orthophosphate (Ghosh and Preiss 1966; Sanwal et al. 1968; Plaxton and Preiss 1987). The region of the primary structure of the *Escherichia coli* enzyme pertaining to the allosteric activator site has been identified (Parsons and Preiss 1978a,b). For one of the two subunits of the spinach leaf enzyme, the region of the activator binding site has been identified (Morell et al. 1988). The site in the enteric bacterial enzyme

is close to the N-terminus, and the site in the spinach leaf enzyme is at the carboxyl end of the respective primary structures.

ADPGlc pyrophosphorylase from all sources is found to be a tetrameric structure. However, the enzyme derived from enteric bacteria is homotetrameric in structure (Haugen et al. 1976), whereas the enzyme derived from angiosperm plants is comprised of two subunits, presumably $\alpha_2\beta_2$ (Morell et al. 1987; Okita et al. 1990; Preiss 1991). One report of the plant enzyme from a nonphotosynthetic tissue being a homotetramer (Sowokinos and Preiss 1982) has been shown to be incorrect (Okita et al. 1990).

Immunologic methods have been used to study the relationship between the enteric bacterial enzyme and the angiosperm enzyme, the results of which suggest that these may be unrelated (Preiss 1982, 1988). The relationship between the two subunits of the angiosperm plant enzyme has also been studied by immunologic methods. The results indicate there are some epitopes shared between the two subunits and some epitopes unique to the subunit exhibiting the greater apparent molecular weight (Morell et al. 1987). In three cereals, wheat, rice, and maize, there appears to be tissue-specific isozymes differing in apparent molecular mass with some of the polypeptides retaining some or all of the antigenic determinants (Krishnan et al. 1986). The large subunit of enzyme from nonphotosynthetic tissue is apparently less reactive to antiserum prepared against the large subunit of enzyme from photosynthetic tissue (Okita et al. 1990; Preiss et al. 1990).

Biochemical genetics (Preiss et al. 1990) allows a one-to-one correspondence between a specific subunit of the enzyme and isolated cDNA clones for only those clones derived from maize endosperm (Barton et al. 1986; Bae et al. 1990; Bhave et al. 1990). The exact identity of the others has been uncertain. The work presented in this paper shows (1) that the known primary structures of angiosperm plant ADPGlc pyrophosphorylase suggest demarcation into representatives of the two subunits with the large subunit group being comprised of two subgroups that show a correlation for the photosynthetic status of the tissue in which the respective genes are expressed, (2) that the small subunit may be distinguished by whether the host organism is a monocot or a dicot, and (3) that the coding regions of the genes of these proteins underwent at least two duplications in the distant past, one for the two subunits of the plant enzyme and one for the two subgroups of the large subunit. Additionally, we demonstrate that the regions of the protein that are involved in substrate binding and allosteric activation of the enteric bacterial protein are conserved in all known enzymes.

Table 1. Sources of nucleic acid and deduced primary structure for plant proteins

Organism/ tissue	Nomenclature	Reference(s)
Arabidopsis thaliana	*A. thaliana*	Personal communication from A. Lönneborg, Plant Molecular Biology Lab, Norway
Zea mays		
Endosperm	*bt-2*	Barton et al. (1986); Bae et al. (1990)
Endosperm	*sh-2*	Barton et al. (1986); Bhave et al. (1990)
Wheat		
Leaf	WL7	Olive et al. (1989)
Endosperm	WE3	
Endosperm	WE7	
Potato		
Tuber	Potato (T O)	Anderson et al. (1990) and the nucleotide sequences of T. Okita and J. Preiss (unpublished)
Tuber	S 25-1	Müller-Röber et al. (1990)
Tuber	B 22-1	Müller-Röber et al. (1990)
Rice		
Seed	Rice seed	Preiss et al. (1987); Anderson et al. (1989)

Materials and Methods

The source of the nucleotide sequences from which the deduced primary structures of the angiosperm plant enzymes in Fig. 1 are derived as well as the nucleic acids for Fig. 2 are presented in Table 1. The information for the small subunit of the spinach leaf enzyme, designated SL-51 kd, is from both protein sequence determination of purified small subunit (M. Morell and J. Preiss, unpublished) and the nucleotide sequences of cDNA clones of the ADPGlc pyrophosphorylase (B. Smith-White and J. Preiss, unpublished). The peptide sequences of the large subunit of spinach leaf enzyme, designated SL-54 kd, are from protein sequences determined from proteolytically derived peptides (K. Ball, J. Leykam, J. Hutny, and J. Preiss, unpublished). Only those peptides that show a relationship to previously reported primary structures are included. Those peptides whose position cannot be assigned by relation to other ADPGlc pyrophosphorylase proteins are not presented. Cloning and determination of the nucleotide sequence of the cDNAs encoding the spinach leaf enzyme subunits will be presented elsewhere. Primary structures of enzymes from bacterial sources are reported elsewhere (Baecker et al. 1983; Leung and Preiss 1987). Whether the designation in the text refers to the protein or the nucleic acid will be obvious from the context.

Sequence Alignment

Protein. Dot plot analysis of these primary structures showed between 70 and 90% identity for extensive lengths (unpublished). This suggested that these primary structures could be aligned. Final alignment is achieved manually on a computer using version 3.1 of PC Write word processing software. This is a refinement of an alignment in a recent review (Preiss 1991) that was developed from an alignment in Anderson et al. (1989). The latter alignment was produced with the alignment algorithm of the

ALIGNMENT OF AMINO ACID SEQUENCES OF KNOWN ADP-GLUCOSE PYROPHOSPHORYLASES

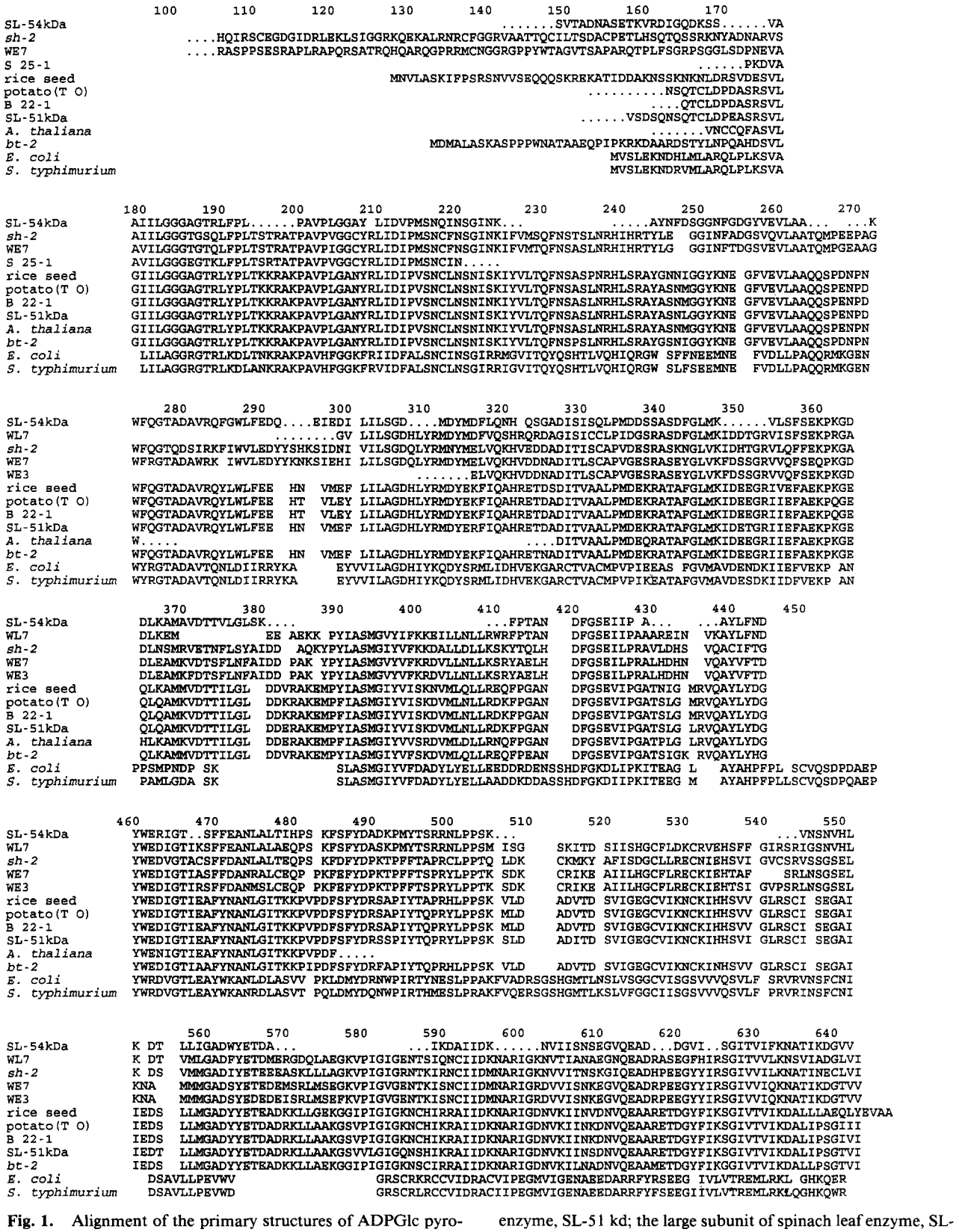

Fig. 1. Alignment of the primary structures of ADPGlc pyrophosphorylase from rice seed, *Zea mays sh-2* and *bt-2* loci, small [potato (T O) and B 22-1] and large (S 25-1) subunit from potato tuber, wheat endosperm (WE3 and WE7), wheat leaf (WL7), spinach leaf 51-kd and 54-kd subunits, *Arabidopsis thaliana, Escherichia coli,* and *Salmonella typhimurium.* Nomenclature is presented in Table 1 except for the small subunit of spinach leaf enzyme, SL-51 kd; the large subunit of spinach leaf enzyme, SL-54 kd; and those from the enteric bacteria *E. coli* and *S. typhimurium,* designated as, respectively, *E. coli* and *S. typhimurium.* The order is chosen to maximize the visual impact of relationships among members of apparent groups of proteins. The first digit of the numbers is atop the residue corresponding to that value.

Table 2. Similarity matrix for plant ADPGlc pyrophosphorylase proteins

	SL-54 kDa	WL7	WE3	WE7	*sh-2*
Number of residues in primary structure		301	298	497	542
SL-54 kd	1.0	0.72/0.80	0.52/0.75	0.59/0.78	0.56/0.75
WL7		1.0	0.43/0.56	0.54/0.71	0.53/0.70
WE3			1.0	0.96/0.97	0.68/0.81
WE7				1.0	0.66/0.78
sh-2					1.0
S 25-1					
A. thaliana					
SL-51 kd					
Potato (T O)					
B 22-1					
Rice seed					
bt-2					

Fraction of amino acid residues identical and fraction of amino acid residues identical plus fraction of amino acid residues homologous in pairwise evaluation of ADPGlc pyrophosphorylase proteins from plant sources. The total number of residues, used as the denominator value in any pairwise comparison, is from the smaller of the two proteins. Nomenclature is presented in Table 1 except for the small subunit of spinach leaf enzyme, SL-51 kd; and the large subunit of spinach leaf enzyme, SL-54 kd

Wisconsin package of sequence analysis programs (Devereaux et al. 1984). For the refinement, regions of identical amino acid residues are aligned and these aligned domains are arranged to simultaneously allow the protein with the greatest number of amino acid residues between two adjacent aligned domains sufficient space to retain all residues and to maximize similarity of the intervening sequence. All the proteins do not possess exactly the same number of amino acid residues, and the first amino acid of the primary translate is unknown except for three proteins: the two proteins from enteric bacteria and the protein from rice seed. Thus, no numbering scheme to which all the proteins can be aligned is evident. However, identification of a particular location in an aligned group of linear sequences is most readily accomplished by referring to an integer scale with a one-to-one correspondence between the scale and locations in the primary structure. Because the N-terminus of at least WE7 and *shrunken-2* (*sh-2*) extend beyond the N-terminus of rice seed, the integer scale numbering scheme is arbitrarily set to commence at 100 to allow for addition of future information. For those primary structures that we consider to be incomplete because of lack of information, we have included an ellipsis to denote the fact that the presented sequence could continue. For the spinach leaf large subunit sequences, there is an ellipsis between continuous peptide sequences.

Nucleic Acid. Only the nucleic acid corresponding to the cloned angiosperm plant proteins is subjected to pairwise comparisons. Alignment was accomplished manually with the same word processing software. DNA sequences were aligned so that the codons corresponding to aligned amino acids were aligned. The deletions in the plant enzymes introduced in the region delimited by locations 375 and 460 of Fig. 1 are not included in the nucleic acid alignments because these gaps were introduced in the plant primary structures in the alignment of the primary structures to maximize the local apparent similarity between the two apparent groups, enteric bacterial protein and angiosperm plant protein.

Sequence Comparison

Protein. Protein sequences were compared pairwise at each amino acid residue over the entire primary structure for identical residues, conservatively replaced residues, and radically replaced residues. A deletion in one of the pairs under consideration is considered a radical replacement. The following groups are comprised of amino acid residues considered to be functionally similar and therefore a conservative replacement: (1) lysine, arginine, asparagine, histidine, and glutamine; (2) aspartic and glutamic acid; (3) serine and threonine; (4) leucine, valine, isoleucin, phenylalanine, tyrosine, and tryptophan. Methionine, glycine, and alanine could be functionally hydrophobic and a conservative replacement if the other residues at that location in the proteins exhibit significant evolutionary pressure to conservatively retain hydrophobic residues. The helix destabilization/distortion capabilities of glycine, histidine, and proline could be considered functionally similar, but are considered a nonconservative replacement in this paper.

Nucleic Acid. The nucleic acid comparisons were from either the codon corresponding to the amino acid at location 178 (indicated by a caret in Fig. 2) or the first codon in common between the two sequences to the last codon in common between the two sequences, where the coding region of the nucleic acid is indicated by iterated triplets. This is different from the protocol for primary structure comparisons to reduce the noise in the comparisons. Random pairs of nucleic acids would exhibit approximately 25% identity, whereas random pairs of protein primary structure would exhibit approximately 5% identity. The regions of the proteins N-terminal of location 184 are apparently random using the information available at present. The nucleic acid comparisons were for base substitutions and deletions. The base substitutions were further characterized as to the alteration of the encoded amino acid from the aggregate substitutions in a particular codon and whether a specific substitution was a transition or transversion mutation. The fraction of identical nucleotides is computed as:

$$1 - [(\text{no. of base substitutions} + \text{no. of deletions}) \div \text{no. of nucleotides}] \quad (1)$$

or as:

$$1 - [(\text{no. of base substitutions} + \text{no. of bases deleted}) \div \text{no. of nucleotides}] \quad (2)$$

Weighting of a deletion equal to a base substitution, Eq. 1, follows the sequence analysis programs available from University of Wisconsin (Devereaux et al. 1984), whereas weighting a deletion equal to the number of bases deleted, Eq. 2, follows van den Berg et al. (1978). These should be considered as the bound-

Table 2. Extended

S 25-1	*A. thaliana*	SL-51 kDa	Potato (T O)	B 22-1	Rice seed	*bt-2*
47	240	450	442	444	479	463
	0.62/0.77	0.60/0.77	0.59/0.75	0.60/0.77	0.59/0.76	0.59/0.76
		0.49/0.68	0.50/0.68	0.48/0.68	0.51/0.72	0.47/0.69
		0.47/0.69	0.45/0.64	0.43/0.64	0.43/0.64	0.44/0.66
0.83/0.94	0.53/0.73	0.47/0.69	0.47/0.68	0.48/0.69	0.44/0.65	0.45/0.66
0.76/0.85	0.52/0.72	0.46/0.68	0.49/0.72	0.47/0.69	0.44/0.67	0.45/0.67
1.0		0.59/0.82	0.61/0.83	0.62/0.85	0.61/0.81	0.62/0.85
	1.0	0.90/0.95	0.91/0.94	0.92/0.96	0.88/0.93	0.86/0.93
		1.0	0.95/0.98	0.94/0.98	0.86/0.93	0.87/0.92
			1.0	0.9954	0.90/0.94	0.88/0.94
				1.0	0.89/0.94	0.88/0.94
					1.0	0.85/0.89
						1.0

ary conditions for the evolutionary significance of a deletion. Calculation of the evolutionary distance is with either of two formulae:

$$-\tfrac{1}{2} \ln[(1 - 2P - Q')(1 - 2Q')^{1/2}] \quad (3)$$

where P is the proportion of transition differences and Q′ is the proportion of the transversion differences, or

$$-\tfrac{3}{4} \ln[1 - (\tfrac{4}{3})\lambda] \quad (4)$$

where λ is the proportion of base substitutions.

Equation 3 is from Kimura (1981) and Eq. 4 is from Jukes and Cantor (1969). Because both formulae were derived under the simplified conditions of considering only base substitutions as evolutionary events, both calculations have the number of nucleotides in the region of comparison adjusted for those that are deleted between the pair under examination.

Results

Figure 1 shows the alignment of the protein sequences. The order of the proteins is chosen to maximize the visual impact of the homology between members of apparent groups of proteins. The alignment after location 500 of the enteric bacterial proteins with the angiosperm plant proteins utilizes the alignment of Anderson et al. (1989) as a starting point. The pairwise evaluation of identity and homology for the proteins from plant sources is shown in Table 2. The total number of residues, used as the denominator value in any pairwise comparison, is from the smaller of the two proteins.

First, Fig. 1 shows that the proteins of the bacterial and angiosperm plant enzymes exhibit many conserved domains. These include the regions from location 180 to 240, from location 272 to 282, from location 392 to 410, and from location 462 to 478. Despite this striking conservation, there are regions where the relationship between bacterial and plant enzyme is minimal or nonexistent, whereas either group is comprised of members that show significant identity among themselves. These include the regions from location 245 to 270, from location 286 to 302, from location 363 to 390, and from location 502 to the C-terminus. This suggests that these two groups are paradigms for two classes of enzyme, the enteric bacterial and the angiosperm plant.

Second, the N-terminal sequence up to the codon at location 178 exhibits similarity in pairwise comparisons for only two groups of proteins. These are the two enteric bacterial proteins and the group of potato (T O), B 22-1, and SL-51 kd. There is no other parameter of the members of the latter group other than empirical observation, which would allow demarcation of a subgroup.

Third, Table 2 and Fig. 1 show that the proteins comprising the angiosperm plant enzyme can be divided into three groups: (1) *brittle-2* (*bt-2*), rice seed, potato (T O), B 22-1, SL-51 kd, and *Arabidopsis thaliana*; (2) *sh-2,* S 25-1, WE3, and WE7; and (3) WL7 and SL-54 kd.

The DNA sequence of the maize *sh-2* locus presented elsewhere (Bhave et al. 1990) can be translated to the reported protein sequence only up to nucleotide 1560, corresponding to location 629 in Fig. 1. The nucleotides from 1621 to the indicated translation termination correspond to the last two amino acid residues of the reported primary structure. The nucleotides corresponding to the intervening 20 amino acid residues have apparently been omitted. The *sh-2* nucleotide sequence in Fig. 2 is the continuous sequence reported.

In keeping with the protein primary structure, the codons prior to the amino acid at location 178 exhibit apparently random nucleotide sequence. In this region there are no obvious pairs of frame-shifts between a pair of cDNA sequences that would allow one protein to be related to another through their related nucleotide sequences. For the remainder of the coding region, the primary results of pairwise comparisons of the nucleic acid sequences are shown

ALIGNMENT OF CODING REGIONS OF KNOWN PLANT ADPGLUCOSE PYROPHOSPHORYLASES cDNAs

```
sh-2            AAT TCG GGA GGC AAG TGC GAT TTT GAT CTT GCA GCC ACC TTT TTT TGT TCT GTT GTG TAT CTA GTA GTT GGA GGA GAT

sh-2            ATG CAG TTT GCA CTT GCA TTG GAC ACG AAC TCA GGT CCT CAC CAG ATA AGA TCT TGT GAG GGT GAT GGG ATT GAC AGG
WE7                                                     .......CGT GCG TCT CCC CCG TCA GAG TCG AGG GCT CCG CTG CGA

sh-2            TTG GAA AAA TTA AGT ATT GGG GGC AGA AAG CAG GAG AAA GCT TTG AGA AAT AGG TGC TTT GGT GGT AGA GTT GCT GCA
WE7             GCG CCT CAA AGG TCG GCG ACA CGG CAG CAT CAG GCA CGA CAG GGT CCC AGG AGG ATG TGC AAC GGC GGC AGG GGC CCG
Rice seed       TTTGTAATTCTTTAATTTGTTGCAACC ATG AAT GTG TTG GCA TCT AAG ATC TTC CCT TCC CGC TCC AAT GTT GTT AGC GAG CAA
bt-2                                        ATCATAATTCTCGAGTTGCAAACCATGGAC ATG GCT TTG GCG TCT AAA GCC TCC CCT CCG CCA

sh-2            ACT ACA CAA TGT ATT CTT ACC TCA GAT GCT TGT CCT GAA ACT CTT CAT TCT CAA ACA CAG TCC TCT AGG AAA AAT TAT
WE7             CCA TAC TGG ACA GCT GGT GTC ACC TCC GCC CCA GCC CGG CAG ACA CCC TTG TTC TCC GGA CGT CCT TCA GGA GGA TTA
Rice seed       CAA CAA TCG AAG CGC GAG AAA GCA ACT ATT GAT GAC GCT AAG AAC TCG TCC AAG AAC AAA AAT         CTT GAC CGC
Potato (T O)                                                                                ....AAT TCA CAG ACA TGT CTA GAC CCA
B 22-1                                                                                      ......A CAG ACA TGT CTA GAC CCA
SL-51 kDa                                                                           .......AAT TCG CAG ACT TGT CTT GAT CCC
A. thaliana                                                                                         ........GTT AAT TGT
bt-2            TGG AAT GCC ACC GCC GCC GAG CAG CCA ATT CCA AAG CGT GAC AAA GCC GCT GCA AAT GAT TCA ACA TAC CTC AAT CCT
                                    ^

sh-2            GCT GAT GCA AAC CGT GTA TCT GCG ATC ATT TTG GGC GGA GGC ACT GGA TCT CAG CTC TTT CCT CTG ACA AGC ACA AGA
WE7             AGC GAT CCG AAC GAG GTT GCG GCC GTC ATA CTC GGC GGC GGC ACC GGG ACT CAG CTC TTC CCA CTC ACG AGC ACA AGG
S 25 -1             ....CCA AGG GAT GTG GCT GCA GTC ATA CTG GGA GGA GGA GAA GGG ACC AAG TTA TTC CCA CTT ACA AGT AGA ACT
Rice seed       AGT GTC GAT GAG AGT GTG CTT GGA ATC ATT CTT GGA GGT GGT GCA GGG ACT AGA TTG TAT CCC CTC ACC AAG AAG CGT
Potato (T O)    GAT GCT AGC CGG AGT GTT TTG GGA ATT ATT CTT GGA GGT GGA GCT GGG ACC CGA CTT TAT CCT CTA ACT AAA AAA AGA
B 22-1          GAT GCT AGC CGG AGT GTT TTG GGA ATT ATT CTT GGA GGT GGA GCT GGG ACC CGA CTT TAT CCT CTA ACT AAA AAA AGA
SL-51 kDa       GAA GCT AGC CGC AGT GTT CTA GGT ATT ATT CTT GGA GGT GGT GCT GGT ACA CGT TTG TAC CCA CTT ACT AAG AAA AGA
A. thaliana     TGT CAA TTT GCA AGT GTT TTG GGG ATA ATC TTA GGA GGT GGA GCT GGA ACT CGT CTT TAG CCA CTT ACG AAG AAG AGA
bt-2            CAA GCT CAT GAT AGT GTT CTT GGA ATC ATT CTG GGA GGT GGT GCT GGG ACT AGA TTG TAC CCC TTG ACA AAG AAG CGT

sh-2            GCT ACG CCT GCT GTA CCT GTT GGA GGA TGT TAC AGG CTT ATT GAT ATC CCT ATG AGT AAC TGC TTC AAC AGT GGT ATA
WE7             GCC ACA CCT GCT GTT CCT ATT GGA GGA TGT TAC AGG CTC ATC GAC ATT CCC ATG AGC AAC TGC TTC AAC AGT GGC ATC
S 25-1          GCA ACC CCT GCT GTT CCG GTT GGA GGA TGC TAC AGG CTA ATA GAC ATC CCA ATG AGC AAC TGT ATC AAC.....
Rice seed       GCC AAG CCT GCC GTG CCT CTG GGT GCC AAC TAC AGG CTT ATT GAT ATC CCT GTC AGC AAC TGT TTG AAC AGC AAC ATA
Potato (T O)    GCA AAG CCA GCT GTT CCA CTT GGA GCA AAT TAT CGT CTG ATT GAC ATT CCT GTA AGC AAC TGC TTG AAC AGT AAC ATA
B 22-1          GCA AAA CCA GCC GTT CCA CTT GGA GCA AAT TAT CGT CTG ATT GAC ATT CCC GTA AGC AAT TGC TTG AAC AGT AAC ATA
SL-51 kDa       GCA AAG CCT GCG CTG CCG CTT GGT GCC AAT TAT AGG CTT ATT GAC ATC CCA GTG AGC AAC TGT TTA AAC AGT AAC ATT
A. thaliana     GCG AAA CCA GCT GTG CCT CTT GGT GCT AAC TAT AGG CTT ATT GAT ATT CCT GTG AGC AAC TGT TTG AAT AGC AAC ATA
bt-2            GCC AAG CCT GCA GTG CCA TTG GGT GCC AAC TAT AGA CTG ATT GAT ATT CCT GTC AGC AAT TGT CTC AAC AGC AAC ATA

sh-2            AAT AAG ATA TTT GTG ATG AGT CAG TTC AAT TCT ACT TCG CTT AAC CGC CAT ATT CAT CGT ACA TAC CTT GAA
WE7             AAC AAG ATA TTC GTC ATG ACC CAG TTC AAC TCG GCC TCC CTT AAT CGT CAC ATT CAC CGC ACC TAC CTC GGC
Rice seed       TCC AAG ATC TAT GTG CTG ACA CAA TTC AAC TCT GCC TCT CCG AAC CGT CAC CTG TCA AGA GCC TAT GGG AAC AAC ATT
Potato (T O)    TCC AAG ATC TAT GTT CTC ACA CAA TTC AAC TCT GCC TCT CTG AAT CGC CAC CTT TCA CGA GCA TAT GCT AGC AAC ATG
B 22-1          TCC AAG ATC TAT GTT CTC ACA CAA TTC AAC TCT GCC TCC CTG AAT CGC CAC CTT TCA CGG GCA TAT GCT AGC AAT ATG
SL-51 kDa       TCC AAA ATA TAT GTT CTT ACT CAA TTC AAT TCT GCA TCC CTC AAC CGT CAT CTT TCA CGG GCG TAT GCT AGT AAT CTG
A. thaliana     TCC AAG ATC TAT GTT CTT ACT CAG TTC AAT TCC GCG TCT TTG AAT CGT CAT CTT TCA CGA GCT TAT GCT AGT AAC ATG
bt-2            TCC AAG ATC TAT GTG CTA ACG CAA TTT AAC TCT CCT TCC CTC AAC CGT CAC CTC TCA AGA GCC TAC GGG AGC AAC ATT

sh-2            GGC GGG ATC AAC TTT GCT GAT GGA TCT GTA CAG GTA TTA GCG GCT ACA CAA ATG CCT GAA GAG CCA GCT GGA TGG
WE7             GGG GGA ATC AAT TTC ACT GAT GGA TCC GTT GAG GTA TTG GCC GCG ACG CAA ATG CCC GGG GAG GCT GCT GGA TGG
Rice seed       GGC GGG TAC AAG AAT     GAA GGG TTC GTT GAA GTC CTC GCT GCA CAG CAG AGC CCA GAT AAT CCT AAC     TGG
Potato (T O)    GGA GGA TAC AAA AAC     GAG GGC TTT GTG GAA GTT CTT GCT GCT CAA CAA AGT CCA GAG AAC CCC GAT     TGG
B 22-1          GGA GGA TAC AAA AAC     GAG GGC TTT GTG GAA GTT CTT GCG GCT CAA CAA AGT CCA GAG AAC CCC GAT     TGG
SL-51 kDa       GGA GGT TAC AAG AAC     GAA GGG TTT GTA GAA GTT CTT GCC GCT CAG CAA AGT CCT GAG AAT CCT GAC     TGG
A. thaliana     GGA GGT TAT AAG AAT     GAA GGA TTC GTT GAA GTT CTC GCT GCT CAA CAG AGT CCT GAA AAC CC...
bt-2            GGA GGG TAC AAG AAT     GAA GGG TTT GTT GAA GTC TTA GCT GCA CAG CAG AGC CCA GAT AAT CCA AAC     TGG

WL                                                                                                      .........GGG
sh-2            TTC CAG GGT ACA GCA GAC TCT ATC AGA AAA TTT ATC TGG GTA CTC GAG GAT TAT TAC AGT CAC AAA TCC ATT GAC AAC
WE7             TTC CGC GGA ACA GCG GAC GCG TGG AGA AAA ATT ATC TGG GTG CTT GAG GAC TAT TAT AAG AAT AAA TCC ATA GAG CAC
Rice seed       TTT CAG GGT ACT GCA GAT GCT GTA AGA CAG TAC TTA TGG CTA TTT GAG GAG          CAT AAT          GTT ATG GAG
Potato (T O)    TTC CAG GGC ACG GCT GAT GCT GTC AGA CAA TAT CTG TGG TTG TTT GAG GAG          CAT ACT          GTT CTT GAA
B 22-1          TTC CAG GGC ACG GCT GAT GCT GTT AGA CAA TAT CTG TGG TTG TTT GAG GAG          CAT ACT          GTT CTT GAA
SL-51 kDa       TTC CAG GGT ACA GCT GAT GCT GTC AGG CAA TAC TTG TGG CTA TTC GAA GAG          CAC AAT          GTC ATG GAA
bt-2            TTT CAG GGT ACT GCA GAT GCT GTA AGG CAG TAC TTG TGG TTG TTT GAG GAG          CAT AAT          GTG ATG GAA

WL              GTG CTG ATT CTT TCT GGC GAT CAC CTC TAC CGT ATG GAC TAC ATG GAT TTT GTT CAG AGT CAT CGG CAG AGA GAC GCG
sh-2            ATT GTA ATC TTG AGT GGC GAT CAG CTT TAT CGG ATG AAT TAC ATG GAA CTT GTG CAG AAA CAT GTC GAG GAC GAT GCT
WE7             ATT TTG ATC TTG TCG GGC GAT CAG CTT TAT CGC ATG GAT TAC ATG GAG CTT GTG CAG AAA CAT GTG GAT GAC AAT GCT
WE3                                                     ........GAG CTT GTG CAG AAA CAT GTG GAT GAC AAT GCT
Rice seed       TTT CTA ATT CTG GCT GGA GAT CAC CTT TAC CGC ATG GAC TAT GAA AAG TTC ATT CAG GCA CAC AGA GAA ACA GAT TCT
Potato (T O)    TAC CTT ATA CTT GCT GGA GAT CAT CTG TAT CGA ATG GAT TAT GAA AAG TTT ATT CAA GCC CAC AGA GAA ACA GAT GCT
B 22-1          TAC CTT ATA CTT GCT GGA GAT CAT CTG TAT CGA ATG GAT TAT GAA AAG TTT ATT CAA GCC CAC AGA GAA ACA GAT GCT
SL-51 kDa       TTC TTG ATT CTT GCT GGT GAT CAT TTA TAT CGT ATG GAT TAT GAA AGA TTT ATC CAA GCT CAT AGA GAA ACT GAT GCA
bt-2            TTT CTA ATT CTT GCT GGC GAT CAC CTG TAC CGG ATG GAC TAT GAA AAG TTC ATT CAG GCA CAC AGA GAA ACA AAT GCT
```

Fig. 2. Alignment of the cDNA sequences of ADPGlc pyrophosphorylase from rice seed, *Zea mays sh-2* and *bt-2* loci, small and large subunits from potato tuber, wheat endosperm, wheat leaf, spinach leaf small subunit, and *Arabidopsis thaliana*. Coding regions are displayed as iterated triplets. Codons corresponding to aligned amino acids of Fig. 1 are aligned. The gaps corresponding to introduced deletions in the plant primary structures in the alignment in Fig. 1, which only allow correlation of the plant primary structures with the enteric bacterial primary structures, are not included in the alignment of Fig. 2. Nomenclature is presented in Table 1 except for the small subunit of spinach leaf enzyme, SL-51 kd. Continued on pages 455–456.

in Table 3, and these comparisons are summarized in Table 4.

The data (Table 3A) indicate that there is no apparent relationship between the frequency of base substitutions and the frequency of deletions in pairwise comparisons. The cDNAs of WE3 and WE7, even though almost devoid of base substitution differences, exhibit a single three-codon deletion. The

```
WL              GGG ATC AGC ATC TGT TGC TTG CCT ATT GAT GGC AGC CGG GCG TCT GAT TTT GGT CTC ATG AAG ATA GAC GAC ACA GGA
sh-2            GAT ATC ACT ATA TCA TGT GCT CCT GTT GAT GAG AGC CGA GCT TCT AAA AAT GGG CTA GTG AAG ATT GAT CAT ACT GGA
WE7             GAC ATT ACT TTA TCA TGT GCC CCT GTT GGA GAG AGC CGG GCA TCT GAG TAC GGG CTA GTG AAG TTC GAC AGT TCA GGC
WE3             GAC ATT ACT TTA TCA TGT GCC CCT GTT GGA GAG AGC CGG GCA TCT GAG TAC GGG CTA GTG AAG TTC GAC AGT TCA GGC
Rice seed       GAT ATT ACT GTT GCT GCC CTG CCA ATG GAT GAG AAA CGT GCA ACT GCA TTC GGC CTC ATG AAA ATT GAC GAG GAA GGG
Potato (T O)    GAT ATT ACC GTT GCC GCA CTG CCA ATG GAC GAG AAG CGT GCC ACT GCA TTT GGT CTC ATG AAG ATT GAC GAA GAA GGA
B 22-1          GAT ATT ACC GTT GCC GCA CTG CCA ATG GAC GAG AAG CGT GCC ACT GCA TTC GGT CTC ATG AAG ATT GAC GAA GAA GGA
SL-51 kDa       GAC ATT ACT GTT GCT GCA CTG CCG ATG GAT GAA AAG CGT GCT ACT GCA TTT GGT TTA ATG AAA ATT GAC GAA GAA GGA
A. thaliana     .....GAT ATC ACA GTA GCT GCA TTA CCA ATG GAC GAG CAA CGA GCC ACT GCT TTT GGG CTG ATG AAG ATT GAT GAG GAA GGA
bt-2            GAT ATT ACC GTT GCT GCC CTA CCG ATG GAT GAG AAA CGT GCA ACT GCA TTT GGC CTC ATG AAA ATT GAT GAA GAA GGG

WL              AGA GTT ATT TCA TTT AGT GAA AAA CCG AGA GGA GCT GAT TTA AAG GAA ATG
sh-2            CGT GTA CTT CAA TTC TTT GAA AAA CCA AAG GGT GCT GAT TTG AAT TCT ATG AGA GTT GAG ACC AAC TTC CTG AGC TAT
WE7             CGT GTG GTC CAG TTT TCT GAG CAG CCA AAG GGT GAC GAT CTG GAA GCG ATG AAA GTG GAC ACC AGT TTT CTC AAT TTC
WE3             CGT GTG GTC CAA TTT TCT GAG AAG CCA AAG GGT GAC GAT CTG GAA GCG ATG AAA GTG GAC ACC AGT TTT CTC AAT TTC
Rice seed       AGA ATA GTT GAA TTT GCA GAG AAA CCA AAA GGA GAA CAA TTG AAA GCA ATG ATG GTT GAC ACT ACC ATA CTC GGC CTT
Potato (T O)    CGC ATT ATT GAA TTT GCA GAG AAA CCG CAA GGA GAG CAA TTG CAA GCA ATG AAA GTG GAT ACT ACC ATT TTA GGT CTT
B 22-1          CGC ATT ATT GAA TTT GCA GAG AAA CCG CAA GGA GAG CAA TTG CAA GCA ATG AAA GTG GAT ACT ACC ATT TTA GGT CTT
SL-51 kDa       AGG ATT ATT GAG TTT GCT GAG AAA CCC AAA GGA GAA CAA TTG CAA GCT ATG AAG CTT GAT ACC ACT ATC TTG GGT CTG
A. thaliana     CGT ATT ATT GAA TTT GCT GAA AAA CCA AAA GGG GAG CAC CTA AAG GCC ATG AAG GTA GAC ACA ACA ATT CTA GGT CTT
bt-2            AGG ATC ATT GAG TTT GCT GAG AAA CCG AAA GGA GAG CAG TTG AAA GCA ATG ATG GTT GAC ACC ACC ATA CTT GGC CTT

WL                    GAG GAA     GCA GAA AAG AAA     CCA TAC ATA GCT TCA ATG GGA GTA TAC ATA TTC AAG AAA GAG ATA CTT
sh-2            GCT ATA GAT GAT         GCA CAG AAA TAT CCA TAC CTT GCA TCA ATG GGC ATT TAT GTC TTC AAG AAA GAT GCA CTT
WE7             GCC ATC GAC GAT     CCT GCT     AAA TAT CCA TAC ATT GCT TCT ATG GGA GTC TAT GTC TTC AAA AGA GAT GTT CTG
WE3             GCC ATC GAC GAC     CCT GCT     AAA TAT CCA TAC ATT GCT TCT ATG GGA GTC TAT GTC TTC AAA AGA GAT GTT CTG
Rice seed             GAC GAC GTG AGG GCA AAG GAA ATG CCT TAT ATT GCT AGC ATG GGT ATC TAT GTG ATT AGT AAA AAT GTA ATG
Potato (T O)          GAT GAC AAG AGA GCT AAA GAA ATG CCT TTC ATT GCC AGT ATG GGT ATA TAT GTC ATT AGC AAA GAC GTG ATG
B 22-1                GAT GAC AAG AGA GCT AAA GAA ATG CCT TTC ATT GCT AGT ATG GGT ATA TAT GTC ATT AGC AAA GAC GTG ATG
SL-51 kDa             GAT GAT GAG AGA GCA AAA GAA ATG CCA TAC ATA GCT AGT ATG GGA ATT TAT GTT ATC AGC AAA GAT GTA ATG
A. thaliana           GAT GAT CAG AGA GCC AAG GAG ATG CCT TTC ATT GCT AGT ATG GGT ATT TAT GTT GTA AGC AGA GAT GTA ATG
bt-2                  GAT GAC GTG AGG GCA AAG GAA ATG CCT TAT ATT GCT AGC ATG GGT ATC TAT GTT TTC AGC AAA GAT GTA ATG

WL              CTA AAT CTT TTG AGA TGG CGT TTT CCC ACT GCA AAT GAT TTT GGA TCT GAA ATA ATT CCA GCT GCA GCA AGA GAG ATT
sh-2            TTA GAC CTT CTC AAG TCA AAA TAT ACT CAA TTA CAT GAC TTT GGA TCT GAA ATC CTC CCA AGA GCT GTA CTA GAT CAT
WE7             CTC AAC CTT CTA AAG TCA AGA TAT GCA GAA CTA CAT GAC TTT GGG TCT GAG ATC CTC CCG AGA GCT CTG CAT GAC CAC
WE3             CTC AAC CTT CTA AAG TCA AGA TAC GCA GAA CTA CAT GAC TTT GGG TCT GAA ATC CTC CCG AGA GCT CTG CAT GAC CAC
Rice seed       CTT CAG CTC CTC CGT GAA CAA TTT CCT GGA GCT AAT GAC TTT GGA AGT GAA GTC ATT CCA GGT GCA ACA AAC ATC GGA
Potato (T O)    TTA AAC CTA CTT CGT GAC AAG TTC CCT GGG GCC AAT GAT TTT GGT AGT GAA GTT ATT CCT GGT GCA ACT TCA CTT GGG
B 22-1          TTA AAC CTA CTT CGT GAC AAG TTC CCT GGG GCC AAT GAT TTT GGT AGT GAA GTT ATT CCT GGT GCA ACT TCA CTT GGG
SL-51 kDa       CTA AAT TTG CTT CGG GAT AAA TTT CCT GGT GCT AAC GAT TTT GGC AGT GAA GTT ATT CCA GGA GCG ACT TCC ATA GGG
A. thaliana     CTA GAC TTA CTA CGG AAT CAG TTT CCT GGA GCT AAT GAC TTT GGA AGT GAA GTC ATT CCC GGT GCC ACT CCC CTT GGA
bt-2            CTT CAG CTC CTC CGT GAA CAA TTT CCT GAA GCC AAT GAC TTT GGA AGT GAG GTT ATT CCA GGT GCA ACC AGC ATT GGA

WL              AAT     GTA AAG GCA TAT CTT TTC AAT GAT TAC TGG GAA GAT ATT GGA ACT ATC AAA TCC TTC TTC GAA GCA AAT CTT
sh-2            AGT     GTG CAG GCA TGC ATT TTT ACG GGC TAT TGG GAG GAT GTT GGA ACA ATC AAA TCA TTC TTT GAT GCA AAC TTG
WE7             AAT     GTA CAG GCT TAT GTC TTC ACT GAC TAC TGG GAG GAC ATT GGA ACA ATC AGA TCC TTC TTC GAT GCA AAC AGG
WE3             AAT     GTA CAG GCT TAT GTC TTC ACT GAC TAC TGG GAG GAC ATT GGA ACA ATC AGA TCC TTC TTC GAT GCG AAC ATG
Rice seed       ATG AGG GTA CAA GCT TAC TTG TAT GAT GGT TAC TGG GAA GAC ATT GGT ACC ATT GAG GCA TTC TAT AAT GCA AAT CTG
Potato (T O)    ATG AGA GTG CAA GCT TAT TTA TAT GAT GGG TAC TGG GAA GAT ATT GGT ACC ATT GAA GCT TTC TAC AAT GCC AAT TTG
B 22-1          ATG AGA GTG CAA GCT TAT TTA TAT GAT GGG TAC TGG GAA GAT ATT GGT ACC ATT GAA GCT TTC TAC AAT GCC AAT TTG
SL-51 kDa       CTA ACA GTT CAA GCT TAC TTG TAT GAT GGT TAT TGG GAA GAT ATC GGT ACA ATT GAG GCT TTT TAC AAT GCA AAC TTG
A. thaliana     CTG AGG GTG CAA GCT TAC CTA TAT GAT GGT TAC TGG GAG AAC ATT GGT ACT ATA GAG GCA TTT TAT AAC GCT AAT CTT
bt-2            AAG AGG GTT CAG GCT TAT CTG TAT CAT GGT TAC TGG GAA GAT ATC GGT ACC ATT GCG GCA TTT TAT AAT GCA AAC TTG

WL              GCC CTT GCT GAA CAG CCT TCA     AAG TTC AGC TTC TAT GAT GCT AGC AAA CCG ATG TAC ACA TCG CGA AGA AAC CTA
sh-2            GCC CTC ACT GAG CAG CCT TCC     AAG TTT GAT TTT TAC GAT CCA AAA ACA CCT TTC TTC ACT GCA CCC CGA TGC TTG
WE7             GCC CTC TGC GAG CAG CCC     CCA AAG TTC GAG TTT TAT GAT CCC AAA ACT CCC TTC TTC ACC TCG CCT CGA TAC TTG
WE3             TCC CTC TGC GAG CAG CCC     CCA AAG TTC GAG TTT TAT GAT CCC AAA ACT CCC TTC TTC ACC TCG CCT CGA TAC TTG
Rice seed       GGA ATA ACG AAA AAG CCT GTA CCA GAT TTC AGT TTC TAT GAC CGG TCT GCT CCA ATT TAT ACA CAA CCT CGA CAC TTG
Potato (T O)    GGC ATT ACA AAA AAG CCG GTG CCA GAT TTT AGC TTT TAC GAC CGA TCA GCC CCA ATC TAC ACC CAA CCT CGA TAT CTA
B 22-1          GGC ATT ACA AAA AAG CCG GTG CCA GAT TTT AGC TTT TAT GAC CGA TCA GCC CCA ATC TAC ACC CAA CCT CGA TAT CTA
SL-51 kDa       GGT ATC ACC AAA AAG CCG GTG CCA GAT TTC AGC TTC TAT GAT CGT TCG TCT CCA ATA TAT ACA CAA CCT CGG TAT TTG
A. thaliana     GGA ATC ACC AAG AAA CCA GTT CCT GAT TTT .......
bt-2            GGA ATA ACC AAG AAG CCA ATA CCA GAT TTC AGC TTC TAT GAC CGT TTT GCT CCA ATT TAT ACA CAA CCT CGA CAC CTG

WL              CCA CCA TCT ATG ATC AGC GGT     AGT AAG ATC ACT GAT TCG ATC ATT TCC CAT GGA TGT TTC TTG GAT AAA TGC AGG
sh-2            CCT CCG ACG CAA     TTG GAC AAG TGC AAG ATG AAA TAT GCA TTT ATC TCA GAT GGT TGC TTA CTG AGA GAA TGC AAC
WE7             CCA CCA ACA AAG     TCA GAC AAG TGC AGG ATC AAA GAA GCG ATC ATT CTG CAC GGC TGC TTC TTG CGT GAA TGC AAA
WE3             CCA CCA ACA AAG     TCA GAC AAG TGC AGG ATC AAA GAA GCG ATC ATT CTG CAC GGC TGC TTC TTG CGT GAA TGC AAA
Rice seed       CCT CCT TCG AAG     GTT CTT GAT GCT GAT GTG ACA GAC AGT GTC ATT GGT GAA GGG TGT GTT ATT AAA AAC TGT AAG
Potato (T O)    CCA CCA TCA AAA     ATG CTT GAT GCT GAT GTC ACA GAT AGT GTC ATT GGT GAA GGT TGT GTG ATC AAG AAC TGT AAG
B 22-1          CCA CCA TCA AAA     ATG CTT GAT GCT GAT GTC ACA GAT AGT GTC ATT GGT GAA GGT TGT GTG ATC AAG AAC TGT AAG
SL-51 kDa       CCT CCT TCA AAG     ATG CTT GAC GCT GAT ATA ACT GAC AGT GTC ATC GGT GAA GGC TGT GTG ATT AAG AAC TGC AAG
bt-2            CCA CCT TCA AAG     GTT CTT GAT GCT GAT GTG ACA GAC AGT GTT ATT GGT GAA GGA TGT GTT ATT AAA AAC TGC AAG

WL              GTA GAG CAC AGT GTC GTT GGA ATC CGT TCT CGA ATA GGC TCC AAC GTA CAC CTC AAG     GAT ACG GTA ATG CTC GGT
sh-2            ATC GAG CAT TCT GTG ATT GGA GTC TGC TCA CGT GTC AGC TCT GGA TGT GAA CTC AAG     GAC TCC GTG ATG ATG GGA
WE7             ATC GAG CAC ACT GCG TTC             TCA CGC CTA AAC TCC GGA AGC GAG CTC AAG AAT GCG     ATG ATG ATG GGC
WE3             ATC GAG CAC TCC ATC ATC GGC GTT CCT TCA CGC CTA AAC TCC GGA AGC GAA CTC AAG AAC GCG     ATG ATG ATG GGT
Rice seed       ATA CAC CAT TCA GTA GTT GGA CTC CGG TCC TGC ATT     TCT GAA GGC GCA ATA ATA GAG GAC TCA TTA CTC ATG GGA
Potato (T O)    ATT CAT CAT TCC GTG GTT GGA CTC AGA TCA TGC ATA     TCA GAG GGA GCA ATT ATA GAA GAC TCA CTT TTG ATG GGG
B 22-1          ATT CAC CAT TCC GTG GTT GGG CTC AGA TCA TGC ATA     TCA GAG GGA GCA ATT ATA GAA GAC TCA CTT TTG ATG GGG
SL-51 kDa       ATT CAT CAT TCT GTT ATT GGT CTC CGA TCT TGC ATT     TCG GAG GGT GCA ATT ATT GAG GAT ACG CTG TTA ATG GGA
bt-2            ATA AAC CAT TCT GTA GTT GGA CTC CGA TCT TGC ATA     TCT GAA GGT GCT ATC ATA GAG GAC AGT TTA CTA ATG GGT
```

Fig. 2. Continued.

group of rice seed, potato (T O), B 22-1, SL-51 kd, *A. thaliana,* and *bt-2* are without a deletion among them, whereas there are many base substitutions in all pairwise comparisons. Because the four-codon terminus extension of the rice seed protein is the result of a single base frame-shift mutation in the codon for amino acid residue 639, we do not indicate a deletion between rice seed and the group

```
WL          GCT GAT TTC TAT GAA ACT GAC ATG GAA AGA GGC GAC CAG CTG GCC GAA GGA AAG GTT CCG ATT GGG ATC GGG GAG AAC
sh-2        GCG GAC ATC TAT GAA ACT GAA GAA GAA GCT TCA AAG CTA CTG TTA GCT GGG AAG GTC CCG ATT GGA ATA GGA AGG AAC
WE7         GCG GAC TCG TAC GAG ACC GAA GAC GAG ATG TCG AGG CTG ATG TCG GAG GGC AAG GTC CCC ATC GGC GTC GGG GAG AAC
WE3         GCG GAT TCG TAC GAG ACC GAG GAT GAG ATC TCG AGG CTG ATG TCC GAG GGC AAG GTC CCC ATC GGC GTC GGG GAG AAC
Rice seed   GCT GAC TAC TAC GAG ACT GAA GCA GAC AAG AAA CTC CTT GGT GAA AAA GGT GGC ATT CCC ATT GGT ATT GGG AAG AAT
Potato (T O) GCA GAT TAC TAT GAG ACT GAT GCT GAC AGG AAG TTG TTG GCT GCA AAG GGC AGT GTC CCA ATT GGC ATC GGC AAG AAT
B 22-1      GCA GAT TAC TAT GAG ACT GAT GCT GAC AGG AAG CTG CTG GCT GCA AAG GGC AGT GTC CCA ATT GGC ATC GGC AAG AAT
SL-51 kDa   GCT GAT TAT TAT GAG ACT GAT GCT GAT CGG AAA CTC CTA GCT GCT AAG GGT AGC GTA GTA CTT GGC ATT GGG CAG AAT
bt-2        GCG GAC TAC TAT GAG ACA GAA GCT GAT AAA AAA CTC CTT GCC GAA AAA GGT GGC ATT CCT ATT GGT ATT GGG AAA AAT

WL          ACT TCG ATT CAA AAC TGC ATC ATT GAC AAG AAT GCG AGG ATA GGG AAG AAT GTG ACC ATT GCT AAC GCC GAG GGT GTA
sh-2        ACA AAG ATA AGG AAC TGT ATC ATT GAC ATG AAT GCT AGG ATT GGG AAG AAC GTG GTG ATC ACA AAC AGT AAG GGC ATC
WE7         ACA AAG ATC AGC AAC TGC ATC ATC GAC ATG AAC GCG AGG ATA GGA AGG GAC GTG GTC ATC TCA AAC AAG GAG GGA GTG
WE3         ACA AAG ATC AGC AAC TGC ATC ATC GAC ATG AAC GCG AGG ATA GGA AGG GAC GTG GTC ATC TCA AAC AAG GAG GGA GTG
Rice seed   TGC CAC ATT AGA AGG GCA ATC ATC GAC AAG AAT GCT CGT ATT GGA GAT AAT GTG AAG ATA ATC AAT GTT GAC AAT GTC
Potato (T O) TGT CAC ATT AAA AGA GCC ATT ATC GAC AAG AAT GCC CGT ATA GGG GAC AAT GTG AAG ATC ATT AAC AAA GAC AAC GTT
B 22-1      TGT CAC ATT AAA AGA GCC ATT ATC GAC AAG AAT GCC CGT ATA GGG GAC AAT GTG AAG ATC ATT AAC AAA GAC AAC GTT
SL-51 kDa   AGC CAT ATA AAA AGA GCA ATA ATT GAT AAG AAT GCT CGG ATT GGG GAC AAT GTC AAG ATT ATC AAC AGT GAC AAT GTA
bt-2        TCA TGC ATC AGG AGA GCA ATC ATT GAC AAG AAT GCT CGA ATT GGA GAC AAT GTT AAG ATA CTC AAT GCT GAC AAT GTT

WL          CAG GAA GCG GAC AGG GCG TCA GAA GGC TTC CAC ATC CGG TCC GGT ATC ACG GTT GTG CTG AAG AAC TCG GTG ATT GCG
sh-2        CAA GAG GCT GAT CAC CCG GAA GAA GGG TAC TAC ATA AGG TCT .....
WE7         CAA GAA GCC GAC AGG CCG GAG GAG GGG TAC TAC ATC AGG TCC GGG ATC GTG GTG ATC CAG AAG AAC GCG ACC ATC AAG
WE3         CAA GAA GCC GAC AGG CCG GAG GAG GGG TAC TAC ATC AGG TCC GGG ATC GTG GTG ATC CAG AAG AAC GCG ACC ATC AAG
Rice seed   CAA GAA GCT GCA AGA GAG ACT GAT GGA TAC TTC ATC AAA AGT GGC ATT GTT ACC GTG ATC AAG GAT GCT TTG CTC CTA
Potato (T O) CAA GAA GCG GCT AGG GAA ACA GAT GGA TAC TTC ATC AAG AGT GGG ATT GTC ACC GTC ATC AAG GAT GCT TTG ATT CCA
B 22-1      CAA GAA GCG GCT AGG GAA ACA GAT GGA TAC TTC ATC AAG AGT GGG ATT GTC ACC GTC ATC AAG GAT GCT TTG ATT CCA
SL-51 kDa   CAA GAA GCA GCA AGA GAA ACA GAC GGT TAC TTC ATC AAG AGT GGA ATT GTA ACT GTT ATC AAA GAC GCC TTG ATT CCG
bt-2        CAA GAA GCT GCA ATG GAG ACA GAC GGG TAC TTC ATC AAA GGT GGA ATT GTC ACA GTG ATC AAG GAT GCT TTA CTC CCT

WL          GAT GGA TTA GTC ATA TGA GCTGAAAAAAGGCGGTTCTCCAGTCCAGCAAGAGAAATAAA
WE7         GAC GGC ACC GTC GTG TAG TAC3G3CCGGCGCGACG4TTCCGCGACAACCTCTCTGCTG ATCG TCGTCGTCGGCTTCTCG4CCG3A....
WE3         GAC GGC ACC GTC GTG TAG TAC3G3TCGGCGTGACG3TTCTGCGACAACCTCTC GCTGCGTTGATCGTCGTCGTCGTCT CGAGG ....
Rice seed   GCG GAA CAG TTA TAT GAA GTA GCT GCG TAA TATATGATG4CATCGGCGACGAGCACCAGGCG3CATTATAAGAAGAATA4GCA....
Potato (T O) AGT GGA ATC ATC ATC TGA AGGAATGCGT4AACTTGGTTG CCTCCAAGAT4GGCTAAACAGCCATGAGGTAGAAACGTGCTGAACT4....
B 22-1      AGT GGA ATC GTC ATC TGA AGGAATGCGT4AACTTGGTTGTCCTCCTAGAT4GGCTAAACAGCCATGAGGTAGAAACGTGCTGAACT4.....
SL-51 kDa   AGT GGA ACC GTA ATC TAA AATGGTGCAAATACCAT4CCTCTTTCT11CT11CCAAGT
bt-2        AGT GGA ACA GTT ATA TGA AGTGAACGTGCGACATGCAGCTGTGTGT CTCGACATTCGACAATGATCAGCACCTGGTTACATGCGTG....
```

Fig. 2. Continued.

comprised of *bt-2,* SL-51 kd, potato (T O), B 22-1, and *A. thaliana.*

Also, the assertion made earlier in the text of a grouping of related proteins is confirmed. The fraction of identical nucleotides (Table 4A) independent of the weighting of a deletion indicates that the cDNAs of *bt-2,* SI-51 kd, rice seed, potato (T O), B 22-1, and *A. thaliana* are related and the cDNAs of WE3, WE7, S 25-1, and *sh-2* are related. The *sh-2*: WE3:WE7 group was first reported elsewhere (Bhave et al. 1990), and the latter group of three cDNAs was suggested previously (Müller Röber et al. 1990). The fraction of base substitutions resulting in (1) no change, (2) conservative change, and (3) radical change show clear and consistent similarity among *A. thaliana,* SL-51 kd, rice seed, potato (T O), B 22-1, and *bt-2* (Table 3B and C). Except for the pair from wheat endosperm, no other pairs of cDNA sequences demonstrate such a pronounced similarity. For the other cDNA sequences, the fraction of base substitutions resulting in no change is slightly higher for S 25-1, WE3, WE7, and *sh-2,* whereas the fraction of base substitutions resulting in a radical replacement at a particular location is slightly less for the same group (Table 3B and C). The cDNA analysis confirms the exclusion of WL7 from either group.

The three ratios of types of base substitutions again confirm the same grouping of proteins. In addition, this secondary analysis suggests that there is greater selective pressure to retain the primary structure of the protein represented by rice seed, SL-51 kd, *bt-2,* potato (T O), B 22-1, and *A. thaliana.*

The calculated evolutionary distances are shown in Table 4B. Kimura (1981) derived the equation for calculation of evolutionary distance when the proportions of transition and transversion mutations between two nucleic acid sequences are known. In his derivation, he showed that the formula of Jukes and Cantor (1969) is pertinent when the ratio of transition mutations to transversion mutations is 0.5. Clearly, this is not found exactly with any pair of the mRNAs; but some pairs begin to approximate this condition, whereas some are clearly not near this condition. Apparently, the precursor gene of the coding regions of the genes for the *A. thaliana,* potato (T O), *bt-2,* B 22-1, and rice seed proteins diverged from the precursor of the coding regions of the genes for the other proteins slightly prior to the divergence of the coding regions of the gene for the WL7 protein from the precursor for the coding regions of the genes of the proteins of the group comprised of WE3, WE7, S 25-1, and *sh-2.*

Observation of regions exhibiting identity and conservative replacement of the amino acid residues would be striking enough by itself. However, structure–function studies of the *E. coli* enzyme make the conservation of these regions even more striking.

Even though the enzymes from almost all sources examined to date exhibit allosteric regulation of catalysis by an enzyme with a tetrameric structure, the only property of these enzymes that shows identity

AZIDO ADP-GLUCOSE AND AZIDO AMP BINDING SITE IN *E. coli*

```
                  260        270        280        290        300
E. coli           FVDLLPAQQRMKGEN WYRGTADAVTQNLDIIRRYKA     EY
SL-54kDa          GY*EV*A*... ...S *FQ******R*FGWLFEDQ....EI*D
sh-2              GS*QV*A*T*MPEEPAG*FQ**Q*SIRKFIWVLED*YSHKSIDN
WE7               GS*EV*A*T*MPGEAAG*FR*****WRK IWVLED*YKNKSI*H
rice seed         G**EV*A*Q*SPDNPN *FQ******R*Y*WLFEE  HN  VME
potato(T O)       G**EV*A*Q*SPENPD *FQ******R*Y*WLFEE  HT  VLE
B 22-1            G**EV*A*Q*SPENPD *FQ******R*Y*WLFEE  HT  VLE
SL-51kDa          G**EV*A*Q*SPENPD *FQ******R*Y*WLFEE  HN  VME
A. thaliana       G**EV*A*Q*SPENP.....
bt-2              G**EV*A*Q*SPDNPN *FQ******R*Y*WLFEE  HN  VME
S. typhimurium    **************   ********************    **
```

E. coli SUBSTRATE-BINDING SITE

```
                330        340        350        360
E. coli           ACMPVPIEEAS FGVMAVDENDKIIEFVEKPANP
SL-54kDa          SQL*MDDSSA*D**L*K......VLS*S****GD
WL7               C*L*IDGSR**D**L*KI*DTGRV*S*S***RGA
sh-2              S*A**DESR**KN*LVKI*HTGRVLQ*F***KGA
WE7               S*A**GESR**EY*LVKF*SSGRVVQ*S*Q*KGD
WE3               S*A**GESR**EY*LVKF*SSGRVVQ*S***KGD
rice seed         *AL*MDEKR*TA**L*KI**EGR*V**A***KGE
potato(T O)       *AL*MDEKR*TA**L*KI**EGR****A***QGE
B 22-1            *AL*MDEKR*TA**L*KI**EGR****A***QGE
SL-51kDa          *AL*MDEKR*TA**L*KI**TGR****A***KGE
A. thaliana       *AL*MDEQR*TA**L*KI**EGR****A***KGE
bt-2              *AL*MDEKR*TA**LIKI**EGR****A***KGE
S. typhimurium    *******K**TA********S****D********
```

Fig. 3. Comparison of the primary structures of all of the known ADPGlc pyrophosphorylases relative to that of the peptide of the *Escherichia coli* protein covalently modified by probes of the substrate binding sites. The integers atop each group correspond to those in Fig. 1. The residues of the other ADPGlc pyrophosphorylase proteins that are identical to that found in the *E. coli* protein are indicated as a star. The residue identified as being covalently modified in the *E. coli* protein is emboldened. Nomenclature is the same as in Fig. 1.

between all prokaryotes and all eukaryotes is the catalyzed reaction. The extended regions of identity are assumed to be involved in the fundamental processes of binding of substrates and/or catalysis. Figure 3 shows the peptide regions associated with azido-ADPGlc binding sites of the *E. coli* enzyme (Lee and Preiss 1986) as well as the peptide region protected against reductive phosphopyridoxylation by ADPGlc (Parsons and Preiss 1978b). The emboldened lysine at location 270 and the emboldened arginine at 277 are the residues that become trypsin resistant after photoaffinity labeling by azido-ADPGlc (Lee and Preiss 1986), and the emboldened lysine at location 361 is the residue labeled by reductive phosphopyridoxylation (Parsons and Preiss 1978b). Also as emboldened letters in Fig. 3 are two sites that have been studied by site-directed mutagenesis. First, tyrosine 114 of the *E. coli* enzyme (in Fig. 3, location 276) is a conservative replacement between angiosperm plants and bacteria, whereas this substitution is conserved in identity among all plants examined. Enzyme from the allele of *E. coli glgC* created in vitro, which has phenylalanine at this location, shows altered kinetic and allosteric properties (Kumar et al. 1988). The significance of the concomitant adjacent conservative replacement of an arginine with glutamine in all plants has not been examined. Second, lysine 195 of the *E. coli* enzyme (in Fig. 3, location 361) is conserved for all proteins except WE7, where it is conservatively replaced by glutamine. Site-directed change of this residue in *E. coli* to match this evolutionarily observed change results in a dramatic loss of kinetic affinity for glucose-1-phosphate (Hill et al. 1991). An identical effect can be obtained by changing this residue to the hydrophobic residue isoleucine, whereas a lesser effect can be achieved by replacement with either a histidine or an arginine (Hill et al. 1991).

Allosteric inhibition of the bacterial enzyme has been examined by use of the photoaffinity analog azido-AMP (Larsen and Preiss 1986). Figure 4 shows one of the peptides labeled by azido-AMP binding to the enzyme from *E. coli* (Larsen et al. 1986). It is significantly different between angiosperm plant and bacterial enzymes. There are regions of the primary structures that are conserved as a linear peptide sequence. However aligning the enteric bacterial protein and angiosperm plant protein requires the use of a flanking insertion : deletion pair, with the size of each component mutation being approximately equal. Thus, the location of this apparently conserved region within both the subregion and the entire primary structure is different between the enteric bacterial enzyme and the plant enzyme. This divergence is striking when considering the regions of identical and conservatively replaced residues immediately preceding and following. The implication is that part of the change in allosteric inhibitor specificity between the plant and bacterial enzyme is achieved by alteration of the protein in this region. This leads to the suggestion that this component of the AMP binding site is extremely pertinent to the specificity of the allosteric inhibitor for

Table 3. Primary analysis of nucleic acids of plant ADP-glucose pyrophosphorylase cDNAs

	WL7	WE3	WE7	*sh-2*	S 25-1	*A. thaliana*	Potato (TO)	B 22-1	SL-51 kDa	Rice seed	*bt-2*
Part A											
No. of nucleotides	903	915	1338	1281	135	288/423	1338	1338	1338	1338	1338
WL7	—	314	330	310	NA	157	382	373	385	398	371
WE3	7/20	—	21	230	NA	179	384	392	473	437	414
WE7	10/20	1/3	—	313	31	312	567	552	580	588	581
Maize *sh-2*	7/17	8/8	9/10	—	37	278	537	526	519	538	528
S 25-1	NA	NA	0	0	—	53	50	52	47	47	50
A. thaliana	6/15	6/8	8/10	7/10	0	—	140	149	153	148	165
Potato (TO)	10/19	9/11	15/22	18/19	0	0	—	14	256	287	303
B 22-1	10/19	9/11	15/22	18/19	0	0	0	—	252	288	301
SL-51 kDa	10/19	9/11	15/22	18/19	0	0	0	0	—	294	285
Rice seed	10/19	9/11	15/22	18/19	0	0	0	0	0	—	166
Maize *bt-2*	6/15	7/8	12/14	13/17	0	0	0	0	0	0	—
Part B											
WL7	—	0.29	0.29	0.23	NA	0.29	0.31	0.31	0.31	0.28	0.27
WE3	0.25	—	0.61	0.43	NA	0.28	0.25	0.23	0.27	0.20	0.23
WE7	0.25	0.29	—	0.48	0.68	0.29	0.26	0.27	0.28	0.23	0.26
Maize *sh-2*	0.19	0.23	0.22	—	0.65	0.30	0.28	0.28	0.26	0.26	0.26
S 25-1	NA	NA	0.12	0.13	—	0.49	0.38	0.40	0.36	0.38	0.50
A. thaliana	0.26	0.30	0.33	0.29	0.28	—	0.87	0.88	0.83	0.78	0.79
Potato (TO)	0.21	0.25	0.24	0.28	0.36	0.07	—	0.94	0.77	0.77	0.74
B 22-1	0.25	0.27	0.28	0.30	0.29	0.07	0.06	—	0.86	0.78	0.74
SL-51 kDa	0.23	0.27	0.28	0.32	0.36	0.11	0.10	0.10	—	0.73	0.74
Rice seed	0.25	0.26	0.27	0.25	0.30	0.09	0.09	0.12	0.13	—	0.70
Maize *bt-2*	0.28	0.27	0.29	0.30	0.24	0.08	0.13	0.13	0.15	0.08	—
Part C											
WL7	—	0.46	0.45	0.57	NA	0.44	0.48	0.45	0.45	0.46	0.45
WE3	0.85	—	0.10	0.33	NA	0.41	0.50	0.50	0.46	0.54	0.48
WE7	0.88	2.0	—	0.29	0.19	0.38	0.50	0.45	0.44	0.49	0.45
Maize *sh-2*	0.79	1.48	1.26	—	0.21	0.41	0.44	0.42	0.43	0.49	0.44
S 25-1	NA	NA	0.48	0.65	—	0.24	0.26	0.31	0.28	0.23	0.28
A. thaliana	0.65	0.74	0.84	0.67	0.71	—	0.05	0.05	0.06	0.11	0.13
Potato (TO)	0.65	0.76	0.73	0.73	0.66	1.81	—	0.00	0.04	0.14	0.12
B 22-1	0.63	0.76	0.73	0.74	0.73	1.78	13	—	0.04	0.11	0.12
SL-51 kDa	0.76	0.76	0.74	0.70	0.34	1.49	1.4	1.35	—	0.13	0.12
Rice seed	0.75	0.67	0.71	0.67	0.56	0.88	1.14	1.26	1.08	—	0.22
Maize *bt-2*	0.74	0.82	0.81	0.71	0.72	0.99	1.28	1.21	1.09	1.37	—
Part D											
WL7	—	1.16	1.16	1.21	NA	1.09	1.48	1.26	1.34	1.13	0.97
WE3	0.41	—	2.17	1.83	NA	0.91	1.0	0.88	1.03	0.75	0.87
WE7	0.41	1.56	—	2.15	5.25	0.86	1.08	0.95	0.97	0.86	0.91
Maize *sh-2*	0.30	0.76	0.94	—	4.8	1.01	0.98	0.94	0.80	1.02	0.86
S 25-1	NA	NA	2.10	1.85	—	1.73	1.06	1.4	1.00	1.00	2.08
A. thaliana	0.41	0.38	0.40	0.42	0.93	—	11.09	13.2	7.47	8.40	9.36
Potato (TO)	0.45	0.33	0.35	0.39	0.61	6.78	—	15	8.15	8.19	5.6
B 22-1	0.44	0.30	0.36	0.39	0.68	4.25	15	—	8.68	6.5	5.89
SL-51 kDa	0.45	0.38	0.38	0.34	0.57	4.88	5.94	6.2	—	5.54	5.0
Rice seed	0.39	0.25	0.30	0.35	0.62	4.72	3.35	3.29	2.77	—	9.0
Maize *bt-2*	0.37	0.31	0.35	0.35	0.96	3.74	2.95	3.07	2.76	2.34	—

Tabulation of pairwise nucleic acid sequence analysis of the aligned cDNA sequences. Part A presents the number of the two types of mutations. Above the diagonal arrangement of horizontal dashes (referred to as the diagonal) is the number of base substitutions, and below the diagonal is the number of in-frame deletions divided by the number of codons removed by these deletions. Above the diagonal of B, below the diagonal of B, and above the diagonal of C is the fraction of base substitutions yielding codons encoding, respectively, the same amino acid (silent mutations), a conservative replacement (muffled mutations), and a nonconservative replacement. Below the diagonal of C is the ratio of transition mutations to transversion mutations. Part D is the ratio of base substitution mutations yielding different types of resultant codons. Above the diagonal is the ratio of silent mutations to muffled mutations, and below the diagonal is the ratio of silent mutations to all other mutations

Table 4. Similarity matrix and evolutionary distance for plant ADPGlc pyrophosphorylase cDNAs

	WL7	WE3	WE7	*sh-2*	*A. thaliana*	Potato (T O)	B 22-1	SL-51 kDa	Rice seed	*bt-2*
Part A. Similarity matrix										
WL7	—	0.59	0.59	0.62	0.43	0.55	0.55	0.55	0.54	0.58
WE3	0.63	—	0.98	0.74	0.35	0.56	0.55	0.46	0.50	0.53
WE7	0.65	0.97	—	0.72	0.23	0.55	0.56	0.58	0.53	0.54
sh-2	0.67	0.72	0.70	—	0.32	0.51	0.52	0.53	0.51	0.52
A. thaliana	0.31	0.31	0.20	0.28	—	0.66	0.64	0.63	0.64	0.60
Potato (T O)	0.60	0.54	0.52	0.48	0.66	—	0.99	0.80	0.78	0.77
B 22-1	0.54	0.53	0.54	0.49	0.64	0.99	—	0.81	0.78	0.77
SL-51 kDa	0.53	0.44	0.51	0.50	0.63	0.80	0.81	—	0.77	0.78
Rice seed	0.51	0.48	0.51	0.48	0.64	0.78	0.78	0.77	—	0.87
bt-2	0.54	0.51	0.51	0.49	0.60	0.77	0.77	0.78	0.87	—
Part B. Evolutionary distance										
WL7	—	0.507	0.511	0.456	1.49	0.602	0.596	0.614	0.635	0.647
WE3	0.508	—	0.024	0.316	1.96	0.780	0.701	0.798	0.812	0.776
WE7	0.505	0.024	—	0.366	NC	0.715	0.685	0.619	0.733	0.753
sh-2	0.452	0.324	0.356	—	2.18	0.783	0.797	0.776	0.814	0.795
A. thaliana	1.44	1.71	NC	2.15	—	0.473	0.516	0.529	0.511	0.582
Potato (T O)	0.599	0.670	0.709	0.816	0.449	—	0.011	0.232	0.262	0.283
B 22-1	0.646	0.692	0.679	0.786	0.490	0.011	—	0.229	0.263	0.281
SL-51 kDa	0.680	0.975	0.737	0.768	0.509	0.229	0.224	—	0.268	0.262
Rice seed	0.718	0.836	0.754	0.818	0.485	0.262	0.263	0.269	—	0.141
bt-2	0.641	0.759	0.739	0.791	0.568	0.279	0.277	0.259	0.140	—

Summary of pairwise nucleic acid sequence analysis. Part A is the computation of the fractional identity between two nucleotide sequences. Above the diagonally arranged horizontal dashes (hereafter referred to as the diagonal) a deletion is weighted equal to a base substitution (Devereaux et al. 1984), whereas below the diagonal a deletion is weighted equal to the number of bases deleted (van den Berg et al. 1978). Part B is the calculated evolutionary distance between pairs of nucleic acid sequences. Above the diagonal, the formula of Kimura (1981) was used, whereas below the diagonal the formula of Jukes and Cantor (1969) was used. Nomenclature is the same as in Tables 2 and 3

the bacterial enzyme. Figure 5 shows another peptide labeled by azido-AMP binding to the *E. coli* enzyme, which overlaps the region of the allosteric activator binding site (Larsen et al. 1986). This region is highly conserved in higher plants. When examined, AMP does not modulate the enzyme activity of enzymes from plant sources (Ghosh and Preiss 1966), and there are no reports of AMP being an allosteric effector of ADPGlc pyrophosphorylase from plant sources. Thus, it is not clear how much this conserved region is involved in specificity of the allosteric inhibitor.

Regions of the protein involved in allosteric activation have been studied in enzymes from both higher plant and bacterial sources using pyridoxal-5′-phosphate (PLP), an in vitro activator analog (Gentner et al. 1969). As shown by the emboldened letter in Fig. 5, the lysine that binds PLP at the activator binding site in the *E. coli* enzyme is conserved in identity in all proteins known except those from WE7, S 25-1, and *sh-2*. The region in both directions surrounding this residue exhibits rather remarkable conservation in identity between bacterial and plant enzymes (31 of 50 residues), suggesting that there has been an evolutionary constraint against changing this portion of the enzyme. In conjunction with the fact that 3PGA is 3–10-fold more potent than FBP as an activator of the angiosperm enzyme (Ghosh and Preiss 1966) and protects against specific reductive phosphopyridoxylation of the small subunit of the spinach leaf enzyme by PLP at a location significantly distant (Morell et al. 1988), the extensive conservation suggests that this region may also be involved in a fundamental aspect of the enzyme mechanism other than specificity of the allosteric activator. A candidate function for this region is analogous to the P-loop of the nucleotide-binding site (Walker et al. 1982). The nucleotide-binding P-loop motif is $GX_4GK[TS]$ (Saraste et al. 1990), where residues exclusively selected are indicated by their one-letter mnemonic, residues having no evolutionary bias are shown as "X," and residues having some evolutionary bias have the interchangeable residues enclosed in brackets. There is nothing exactly conforming to this motif; however, the group comprised of rice seed, potato (T O), B 22-1, SL-51 kd, *A. thaliana*, and *bt-2* have a large number of conserved glycines, locations 181, 185, 186, 187, and 188, and two conserved lysines, locations 198 and 201. Also 4 of the 12 conserved proline locations are within this region. The differences between the nucleotide-binding P-loop and

AZIDO-AMP BINDING SITE IN *E. coli*

```
                  370        380        390        400        410        420        430        440
E. coli           PSMPNDP SK                    SLASMGIYVFDADYLYELLEEDDRDENSSHDFGKDLIPKITEAG L   AYAHPFPL SCVQSDPDAEPYWRD
SL-54kDa          DLKAMAVDTTVLGLSK....                                ...FPTA*   ***SEI** A..    ....**LFND               **ER
WL7               DLKEM            EE AEKK  PYI****V*I*KKEI*LN**RWRFPTA*   ***SEI**AAAREIN  VK**LFND               **E*
sh-2              DLNSMRVETNFLSYAIDD  AQKYPYL********KK*A*LD**KSKYTQLH   ***SEIL*RAVLDHS  VQ*CIFTG               **E*
WE7               DLEAMKVDTSFLNFAIDD PAK YPYI****V***KR*V*LN**KSRYAELH   ***SEIL*RALHDHN  VQ**VFTD               **E*
WE3               DLEAMKFDTSFLNFAIDD PAK YPYI********KR*V*LN**KSRYAELH   ***SEIL*RALHDHN  VQ**VFTD               **E*
rice seed         QLKAMMVDTTILGL  DDVRAKEMPYI*******ISKNVMLQ**REQFPGA*   ***SEV**GA*NI* MRVQ**LYDG               **E*
potato(T O)       QLQAMKVDTTILGL  DDKRAKEMPFI*******ISK*VMLN**RDKFPGA*   ***SEV**GA*SL* MRVQ**LYDG               **E*
B 22-1            QLQAMKVDTTILGL  DDKRAKEMPFI*******ISK*VMLN**RDKFPGA*   ***SEV**GA*SL* MRVQ**LYDG               **E*
SL-51kDa          QLQAMKVDTTILGL  DDERAKEMPYI*******ISK*VMLN**RDKFPGA*   ***SEV**GA*SL* *RVQ**LYDG               **E*
A. thaliana       HLKAMKVDTTILGL  DDERAKEMPFI*******VSR*VMLD**RNQFPGA*   ***SEV**GA*PL* *RVQ**LYDG               **EN
bt-2              QLKAMMVDTTILGL  DDVRAKEMPYI********SK*VMLQ**REQFPEA*   ***SEV**GA*SI*K RVQ**LYHG               **E*
S. typhimurium     A*LGDA **                 ******************AA**K*DA********I******E* M   ********L*******Q*******
```

Fig. 4. Comparison of the primary structures of all of the known ADPGlc pyrophosphorylases relative to that of the peptide of the *Escherichia coli* protein covalently modified by azido-AMP, a small molecule probe of the allosteric inhibitor binding site of the *E. coli* enzyme. The integers atop the group correspond to those in Fig. 1. The residues of the other ADPGlc pyrophosphorylase proteins, which are identical to those found in the *E. coli* protein, are indicated as a star. The peptide identified as being covalently modified in the *E. coli* protein is emboldened. Nomenclature is the same as in Fig. 1.

FBP BINDING PEPTIDE OF *E. coli*

```
                  180        190        200        210        220        230
E. coli             LILAGGRGTRLKDLTNKRAKPAVHFGGKFRIIDFALSNCINSGIRRMGV
SL-54kDa          AI**G**A****FP*.....***PL**AY L**VPM**Q*****NK...
sh-2              AI**G**T*SQ*FP**ST**T***PV**CY*L**IPM***F****NKIF*
WE7               AV**G**T**Q*FP**ST**T***PI**CY*L**IPM***F****NKIF*
S 25-1            AV**G**E**K*FP**SRT*T***PV**CY*L**IPM*****.....
rice seed         GI**G**A****YP**********PL*ANY*L**IPV***L**N*SKIY*
potato(T O)       GI**G**A****YP**********PL*ANY*L**IPV***L**N*SKIY*
B 22-1            GI**G**A****YP**********PL*ANY*L**IPV***L**N*NKIY*
SL-51kDa          GI**G**A****YP*********LPL*ANY*L**IPV***L**N*SKIY*
A. thaliana       GI**G**A****YP**********PL*ANY*L**IPV***L**N*NKIY*
bt-2              GI**G**A****YP**********PL*ANY*L**IPV***L**N*SKIY*
S.typhimurium     **************A***************V********L******I**
```

Fig. 5. Comparison of the primary structures of all of the known ADPGlc pyrophosphorylases relative to that of the peptide of the *E. coli* protein covalently modified by PLP at the FBP activator binding site. The integers atop the group correspond to those in Fig. 1. The residues of the other ADPGlc pyrophosphorylase proteins, which are identical to those found in the *E. coli* protein, are indicated as a star. The residue identified as being covalently modified in the *E. coli* protein is emboldened. Nomenclature is the same as in Fig. 1.

3 PGA BINDING SITE OF SPINACH LEAF ENZYME

```
                  620        630        640
SL-51kDa          ETDGYFIKSGIVTVIKDALIPSGTVI
SL-54kDa          ..**VI..***TVIFKN*T*KD*V*
WL7               ASE*FH*R***TV*L*NSV*AD*L**
sh-2              PEE**Y*R****VIL*N*T*NECL**
WE7               PEE**Y*R****VIQ*N*T*KD***V
WE3               PEE**Y*R****VIQ*N*T*KD***V
riceseed          ******************LLAEQLYEVAA
potato(T O)       ************************II*
B 22-1            *************************
bt-2              ********G**********L******
```

Fig. 6. Comparison of the primary structure of all of the known plant ADPGlc pyrophosphorylases relative to that of the peptide of the small subunit of the spinach leaf enzyme covalently modified by PLP at the 3PGA binding site. The integers atop the group correspond to those in Fig. 1. The residues of the other angiosperm ADPGlc pyrophosphorylase proteins, which are identical to those found in the spinach leaf small subunit, are indicated as a star. The residue identified as being covalently modified is emboldened. Nomenclature is the same as in Fig. 1.

this region may be due to a structural distinction. The proteins that have the P-loop use it to bind a phosphate diester–divalent cation complex, whereas enteric bacterial ADPGlc pyrophosphorylase binds FBP. Because 3PGA is structurally similar to half of FBP, this region is possibly involved in binding the phosphate of 3PGA, which corresponds to the phosphorylated sixth position of FBP. However, the failure to modify this region in the spinach leaf enzyme is without explanation. Finally, the region suggests that the two broad groupings of the plant proteins are subject to differential selective pressure upon the member proteins, with the group comprised of rice seed, potato (T O), B 22-1, SL-51 kd, *bt-2,* and *A. thaliana* being more like the enteric bacterial protein. The significance of this is unknown.

The emboldened letters in Fig. 6 are the lysines labeled by PLP at the activator binding site in the small subunit (Morell et al. 1988) and large subunit (K. Ball and J. Preiss, unpublished) of the spinach leaf enzyme. This residue is conserved among all angiosperm plants. There are two observations concerning this region. First there has been selective manipulation of this region during evolution. Almost all the residues of rice seed, *A. thaliana,* SL-51 kd, *bt-2,* potato (T O), and B 22-1 are identical whereas WL7, WE3, WE7, S 25-1, and *sh-2* have some residues identical and some residues conservatively replaced. Second, the WE3 and WE7 proteins have a glutamine side chain in the hydrophobic region immediately N-terminal to this lysine. The biochemical and functional effects of these distinctions are not known.

Discussion

The large-scale conservation between enteric bacterial enzymes and plant enzymes through the regions that are implicated for the bacterial enzyme in substrate binding suggests that the catalytic mechanism for this enzyme has undergone little modification through evolution. However, the mechanism(s) of allosteric modulation of this catalytic mechanism has(have) been subjected to many attempts at refinement during evolution, presumably to accommodate different metabolic and physiologic conditions.

The question of whether there are representatives of paradigms for the two subunits of the plant enzyme leads to equivocation at best. Figure 1 and Tables 2 and 3 indicate that rice seed, potato (T O), B 22-1, *bt-2,* SL-51 kd, and *A. thaliana* show substantial identity to each other while showing essentially the same lessened relationship to all other proteins of ADPGlc pyrophosphorylase. Biochemical genetic studies with maize have shown that the *sh-2* locus encodes the large subunit of the maize endosperm enzyme and the *bt-2* locus encodes the small subunit of the endosperm enzyme (Preiss et al. 1990). In lieu of the exclusion of *sh-2* from the group that includes two members which are identified explicitly as the small subunit of the respective enzyme, we consider rice seed, potato (T O), B 22-1, SL-51 kd, *bt-2,* and *A. thaliana* as members of a small subunit consensus group.

Unfortunately, a binary dichotomy fails beyond this point. Comparison of the other primary structures shows them to be as related among themselves as to any of the small subunit consensus group. A weak grouping is suggested by the slightly greater similarity of *sh-2,* S 25-1, WE3, and WE7. This is confirmed in Fig. 1 with a large number of positions in the primary structure of those proteins showing the conserved residues being unique to those proteins. Likewise, there are a number of positions in the primary structure of the WL7 and SL-54 kd that show conserved residues that are unique to that pair. It is suggested that the proteins not of the consensus small subunit group exhibit the property that amino acid residues at some locations in the primary structure are conserved among all members of the group, whereas amino acid residues at other locations show clear evidence that there are at least two different subclasses of these proteins. Examples of this are locations 438–445, 470–490, and 545–552. The region from location 510 to 540 may also conform to this pattern, but awaits determination of the primary structure of the spinach leaf 54-kd subunit. As aligned nucleotides, the cDNAs of S 25-1, WE3, WE7, and *sh-2* are a group, whereas WL7 is as related to any one of these as it is to a member of the consensus small subunit group.

In that the dichotomy of the large subunit seems to correlate with tissue type from which mRNA was isolated, photosynthetic versus nonphotosynthetic, we would like to suggest that there may be minimally two sorts of enzyme in the plant—that found in photosynthetic tissue and that found in nonphotosynthetic tissue. Because SL-51 kd is from leaf, *bt-2* is from endosperm, potato (T O) and B 22-1 are from tuber, and rice seed is from seed, the two groups of plant enzymes would possess a small subunit with a primary structure almost identical to the small subunit consensus protein group, whereas the other subunit would have a much more plastic primary structure that would be specific for tissue metabolism in an as yet unknown manner. Immunologic comparisons of enzymes from different tissues of a single plant (Krishnan et al. 1986) support this idea. Also, immunologic examinations of purified enzyme from different tissues from various plants support this idea. Antibody prepared against the large subunit of spinach leaf enzyme reacts weakly, if at all, with the large subunit from nonphotosynthetic tissue compared to that from photosynthetic tissue (Plaxton and Preiss 1987; Anderson et al. 1990; Okita et al. 1990; Preiss et al. 1990; Preiss 1991). However, antibody prepared against the small subunit of spinach leaf enzyme reacts very well against the small subunit from both tissues (Plaxton and Preiss 1987; Anderson et al. 1990; Okita et al. 1990; Preiss et al. 1990; Preiss 1991). This suggestion can be further tested with the acquisition of clones that will allow a one-to-one correspondance between the expressed genes and the observed proteins from *A. thaliana* leaf (Lin et al. 1988a,b), rice leaf (Krishnan et al. 1986), and maize leaf (Krishnan et al. 1986).

Attempts to subdivide the small subunit group are difficult because the members of the group are so similar. However two points should be noted. First, the few base substitutions in the cDNAs and the resulting even fewer residue changes in the deduced primary structures between potato (T O) and B 22-1 probably represent allelic variation between cultivars. Second, rice seed and *bt-2* are slightly more related to each other than they are to any other member of the small subunit group. Of the 51 residues of the primary structure that are different between the two, irrespective of conservative or nonconservative replacement, 13 occur prior to location 178, the apparent demarcation between the random N-terminus and the conserved body. The Table 2 entry for the comparison of these two proteins can be amended for a comparison of only the body of the two proteins, adjusting the denominator for the shortened basis of comparison. The amended fractional similarity for these two proteins is 0.91/0.94, which is greater than the similarity between either of these proteins and any one of the other small subunit group members. This fractional identity could be even greater, in that the frame-shift at the rice seed C-terminus introduces 10 differences in the primary structure and the frame-shifted DNA in the rice seed cDNA is 100% identical to the *bt-2* cDNA. Performing this amended comparison for the other pairs is irrelevant either because one of the proteins does not extend sufficiently far N proximal of location 178 (other small subunit proteins) or the bodies of the two proteins are sufficiently different that reduction of the noise will not signif-

icantly affect the fractional similarity (large subunit proteins WE7 and *sh-2*). That rice and maize are monocot plants, whereas the other representatives of the small subunit are from dicot plants, leads us to suggest that the small subunit of the enzyme from monocot plants may have a distinctive primary structure relative to the small subunit of the enzyme from dicot plants. Again, the comparison of nucleic acids confirms this subgrouping with the cDNA corresponding to the small subunit from a monocot source being slightly more similar to small subunit cDNA from another monocot than small subunit cDNA from a dicot, and the cDNA corresponding to the small subunit from a dicot source being slightly more similar to small subunit cDNA from another dicot than small subunit cDNA from a monocot. Whether there are concomitant biochemical distinctions devolving from this grouping is not known. The analysis for this distinction cannot be performed upon the large subunit proteins.

The retrieval of multiple clones suggesting multiple mRNA species in wheat endosperm is of interest. These could represent evolution of identical loci in a polyploid organism. Wheat is the only polyploid organism in this group of plants. However, there is no evidence to exclude the possibility of multiple proteins of such similarity being present in the pertinent tissue of any of the other plants. These results could arise from differences of style of operation in different laboratories. Only the wheat clones were examined in detail for subtle differences of structure prior to nucleotide sequence determination. DNA sequences of all the others are the result of nucleotide sequence determinations of the clone with the largest insert.

Conclusive elucidation of the evolutionary relationship between the various genes awaits isolation of genomic clones representing the different expressed mRNA species. However this initial comparison suggests that the genes encoding proteins not of the small subunit consensus group are related by gene duplication with subsequent divergence. Thus, the pairwise comparisons between either the proteins or the nucleic acids corresponding to *sh-2*, S 25-1, WE3, and WE7 are orthologous comparisons as is the comparison between either the proteins or the nucleic acids from WL7 and SL-51 kd. A parologous comparison occurs when the proteins or nucleic acids are from a different group. Paralogous comparisons in Fig. 1 and Tables 3 and 4 suggest that the two subgroups of the large subunit group separated at about the same time as the large subunit diverged from the small subunit.

The N-terminus region shows only a random relationship between any pair of proteins from plant sources. This is made more acute in light of the extensive identity and functional similarity of residue side chains over the remainder of the protein. This variability of the N-terminal sequences may be similar to that found among the chloroplast transit sequences of the small subunit of ribulose bisphosphate carboxylase/oxygenase (RUBISCO) (Mazur and Chui 1985), or it may merely reflect a lack of information; but this abrupt change from a random relationship to an ordered relationship between the primary structures allows an explanation of the underlying basis of tissue-specific isoforms of the subunits of ADPGlc pyrophosphorylase (Krishnan et al. 1986). There are two general ways to achieve a significant difference of mass while retaining similar antigenicity. One is to interpose among the retained epitopes lengths of polypeptide that show a clear pattern of size difference. The other is to append to one or both ends of an invariant region a length of polypeptide. For ADPGlc pyrophosphorylase one of the two possibilities in the latter method is applicable. Taking the proteins from wheat ADPGlc pyrophosphorylase, the only organism for which some of the proteins from two different tissues are known, sufficient length variation to produce the observed tissue-specific differences of molecular mass can occur in only one region of the primary structure, the N-terminal region. The large-scale similarity between proteins of wheat and the other plant species, independent of tissue source, suggests that this observation can be generalized to all plant species.

For a measure of the evolutionary constraint upon the allowed alteration of primary structure of the encoded protein, we submit comparisons to analysis performed by others upon proteins that are considered to have been conserved through evolution—from plants, the small subunit of RUBISCO; from vertebrates, the proteins encoded by the globin genes; and from all organisms, cytochrome c. For RUBISCO, Broglie et al. (1983) report an apparent relationship between representatives from monocot and dicot sources similar to ADPGlc pyrophosphorylase with some residues invariant, some residues somewhat constrained, and some residues random. They report 75–80% identity between the deduced primary structures. For orthologous globin genes, Miyata et al. (1980) report that the deduced primary structures are 78–95% identical. For cytochrome c, Dickerson and Timkovich (1975) report that the proteins from plants show some residues invariant, some residues somewhat constrained, and some random rare residues. They report from 80 to 95% identity, depending upon the particular pair chosen for comparison. The data presented here suggest that the primary structure of at least the small subunit of the ADPGlc pyrophosphorylase enzymes

from angiosperm plants is as equally conserved as the primary structures of other proteins considered to be evolutionarily conserved.

Acknowledgments. The alignment from location 370 to 460 in Fig. 1 of the enteric bacterial proteins relative to the plant proteins is the work of Yee-Yung Charng. This work was supported in part by National Science Foundation grant DMB86-10319 and USDA/DOE/NSF Plant Science Center Program 88-37271-3964.

References

Anderson JM, Hnilo J, Larsen R, Okita TW, Morell M, Preiss J (1989) The encoded primary sequence of a rice seed ADPglucose pyrophosphorylase subunit and its homology to the bacterial enzyme. J Biol Chem 264:12238–12242

Anderson JM, Okita TW, Preiss J (1990) Enhancing carbon flow into starch: the role of ADPglucose pyrophosphorylase. In: Vayda ME, Park WD (eds) The molecular and cellular biology of the potato. C.A.B. International Wallingford, Oxon, UK, chapter 12

Bae JM, Giroux J, Hannah L (1990) Cloning and characterization of the *brittle-2* gene of maize. Maydica 35:317–322

Baecker PA, Furlong CE, Preiss J (1983) Biosynthesis of bacterial glycogen: primary structure of *Escherichia coli* B ADPglucose synthetase as deduced from the nucleotide sequence of the *glgC* gene. J Biol Chem 258:5084–5088

Barton C, Yang L, Galvin M, Sengupta-Gopalan C, Borelli T (1986) Isolation of the *shrunken-2* and *brittle-2* genes from maize. In: Shannon JC, Knievel DP, Boyer CD (eds) Regulation of carbon and nitrogen reduction and utilization in maize. American Society of Plant Physiologists, pp 363–365

Bhave MR, Lawrence S, Barton C, Hannah LC (1990) Identification and molecular characterization of *shrunken-2* cDNA clones of maize. Plant Cell 2:581–588

Broglie R, Coruzzi G, Lamppa G, Keith B, Chua N-H (1983) Monocot and dicot genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase: structural analysis and gene expression. Stadler Symp 15:59–71

Devereaux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–407

Dickerson RE, Timkovich R (1975) Cytochromes c. In: Boyer PD (ed) The enzymes, vol. 11. Academic Press, New York, pp 397–548

Espada J (1962) Enzymic synthesis of adenosine diphosphate glucose from glucose-1-phosphate and adenosine triphosphate. J Biol Chem 237:3577–3581

Gentner N, Greenberg E, Preiss J (1969) TPNH and pyridoxal-5'-phosphate: activators of ADP-glucose pyrophosphorylase of *Escherichia coli* B. Biochem Biophys Res Commun 36: 373–380

Ghosh HP, Preiss J (1966) Adenosine diphosphate glucose pyrophosphorylase: a regulatory enzyme in the biosynthesis of starch in spinach leaf chloroplasts. J Biol Chem 241:4491–4504

Haugen TH, Ishaque A, Preiss J (1976) Biosynthesis of bacterial glycogen: characterization of the subunit structure of *Escherichia coli* B glucose-1-phosphate adenyltransferase (EC 2.7.7.27). J Biol Chem 251:7880–7885

Hill MA, Kaufmann K, Otero J, Preiss J (1991) Biosynthesis of bacterial glycogen: mutagenesis of a catalytic site residue of ADP-glucose pyrophosphorylase from *Escherichia coli.* J Biol Chem 266:12455–12460

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454–458

Krishnan HB, Reeves CD, Okita TW (1986) ADPglucose pyrophosphorylase is encoded by different mRNA transcripts in leaf and endosperm of cereals. Plant Physiol 81:642–645

Kumar A, Tanaka T, Lee YM, Preiss J (1988) Biosynthesis of bacterial glycogen: use of site-directed mutagenesis to probe the role of tyrosine 114 in the catalytic mechanism of ADP-glucose synthetase from *Escherichia coli.* J Biol Chem 263: 14634–14639

Larsen CE, Preiss J (1986) Covalent modification of the inhibitor binding site(s) of *Escherichia coli* ADP-glucose synthetase: specific incorporation of the photoaffinity analogue 8-azidoadenosine 5'-monophosphate. Biochemistry 25:4371–4376

Larsen CE, Lee YM, Preiss J (1986) Covalent modification of the inhibitor-binding site(s) of *Escherichia coli* ADP-glucose synthetase. J Biol Chem 261:15402–15409

Lee YM, Preiss J (1986) Covalent modification of substrate-binding sites of *Escherichia coli* ADP-glucose synthetase. J Biol Chem 261:1058–1064

Leung PSC, Preiss J (1987) Biosynthesis of bacterial glycogen: primary structure of *Salmonella typhimurium* ADPglucose synthetase as deduced from the nucleotide sequence of the *glgC* gene. J Bacteriol 129:4355–4360

Lin T-P, Caspar T, Somerville C, Preiss J (1988a) Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADPglucose pyrophophorylase activity. Plant Physiol 88:1131–1135

Lin T-P, Caspar T, Somerville C, Preiss J (1988b) A starch deficient mutant of *Arabidopsis thaliana* with low ADPglucose pyrophosphorylase activity lacks one of the two subunits of the enzyme. Plant Physiol 88:1175–1181

Mazur BJ, Chui C-F (1985) Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase–oxygenase from tobacco. Nucleic Acids Res 13:2373–2386

Miyata T, Yasunaga T, Nishida T (1980) Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc Natl Acad Sci USA 77:7328–7332

Morell MK, Bloom M, Knowles V, Preiss J (1987) Subunit structure of spinach leaf ADPglucose pyrophosphorylase. Plant Physiol 85:182–187

Morell M, Bloom M, Preiss J (1988) Affinity labeling of the allosteric activator site(s) of spinach leaf ADPglucose pyrophosphorylase. J Biol Chem 263:633–637

Müller-Röber BT, Koßmann J, Hannah LC, Willmitzer L, Sonewald U (1990) One of two different ADP-glucose pyrophosphorylase genes from potato responds strongly to elevated levels of sucrose. Mol Gen Genet 224:136–146

Okita TW, Nakata PA, Anderson JM, Sowokinos J, Morell M, Preiss J (1990) The subunit structure of potato tuber ADPglucose pyrophosphorylase. Plant Physiol 93:785–790

Olive MR, Ellis RJ, Schuch WW (1989) Isolation and nucleotide sequences of cDNA clones encoding ADPglucose pyrophosphorylase polypeptides from wheat leaf and endosperm. Plant Mol Biol 12:525–538

Parsons TF, Preiss J (1978a) Biosynthesis of bacterial glycogen: incorporation of pyridoxal phosphate into the allosteric activator site and an ADP-glucose-protected pyridoxal phosphate binding site of *Escherichia coli* B ADP-glucose synthetase. J Biol Chem 253:6197–6202

Parsons TF, Preiss J (1978b) Biosynthesis of bacterial glycogen: isolation and characterization of the pyridoxal-P allosteric

activator site and ADP-glucose-protected pyridoxal-P binding site of *Escherichia coli* B ADP-glucose synthetase. J Biol Chem 253:7638–7645

Plaxton WC, Preiss J (1987) Purification and properties of nonproteolytically degraded ADPglucose pyrophosphorylase from maize endosperm. Plant Physiol 83:105–112

Preiss J (1973) Adenosine diphosphoryl glucose pyrophosphorylase. In: Boyer PD (ed) The enzymes, vol 8. Academic Press, New York, pp 73–119

Preiss J (1982) Regulation of the biosynthesis and degradation of starch. Annu Rev Plant Physiol 33:431–454

Preiss J (1984) Bacterial glycogen synthesis and its regulation. Annu Rev Microbiol 38:419–458

Preiss J (1988) Biosynthesis of starch and its regulation. In: Stumpf PK, Conn EE (eds) The biochemistry of plants, vol 14. Academic Press, New York, chapter 6

Preiss J (1991) Biology and molecular biology of starch synthesis and its regulation. In: Miflin BJ (ed) Oxford survey of plant molecular and cellular biology, vol. 7. Oxford University Press, Oxford, UK, pp 59–114

Preiss J, Romeo T (1989) Physiology, biochemistry and genetics of bacterial glycogen synthesis. In: Rose AR, Tempest DH (eds) Advances in microbial physiology, vol 30. Academic Press, New York, pp 183–238

Preiss J, Shen L, Greenberg E, Gentner N (1966) Biosynthesis of bacterial glycogen. IV. Activation and inhibition of adenosine diphosphate glucose pyrophosphorylase of *E. coli* B. Biochemistry 5:1833–1845

Preiss J, Bloom M, Morell M, Knowles VL, Plaxton WC, Okita TW, Larsen R, Harmon AW, Putnam-Evans C (1987) Regulation of starch synthesis: enzymological and genetic studies. In: Bruening G, Harada J, Kosuge T, Hollaender A (eds) Tailoring genes for crop improvement. Plenum, New York, pp 132–152

Preiss J, Danner S, Summers PS, Morell M, Barton CR, Yang L, Neider M (1990) Molecular characterization of the *brittle-2* gene effect on maize endosperm ADPglucose pyrophosphorylase subunits. Plant Physiol 92:881–885

Sanwal GG, Greenberg E, Hardie J, Cameron EC, Preiss J (1968) Regulation of starch biosynthesis in plant leaves: activation and inhibition of ADPglucose pyrophosphorylase. Plant Physiol 43:417–427

Saraste M, Sibbald PR, Wittinghofer W (1990) The P-loop—a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 15:430–434

Sowokinos JR, Preiss J (1982) Pyrophosphorylases in *Solanum tuberosum* III. Purification, physical and catalytic properties of ADPglucose pyrophosphorylase in potatoes. Plant Physiol 69:1459–1466

van den Berg J, van Ooyen A, Mantei N, Schambock A, Grosveld G, Flavell RA, Weissmann C (1978) Comparison of cloned rabbit and mouse beta-globin genes showing strong evolutionary divergence of two homologous pairs of introns. Nature 276:37–44

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1:945–951

Received May 6, 1991/Revised and accepted November 11, 1991