Pure & Appl. Chem., Vol. 63, No. 4, pp. 535–544, 1991.
Printed in Great Britain. ADONIS 206922209100067P
© 1991 IUPAC

Regulatory mechanisms involved in the biosynthesis of starch

Jack Preiss, K. Ball, J. Hutney, B. Smith-White, L. Li and Thomas W. Okita

Department of Biochemistry, Michigan State University, East Lansing, MI 48824, USA (J.P., K.B, J.H.,L.L., B. S.-W.) and Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, USA (T.W.O.)

Abstract - The reactions involved in starch synthesis have been shown to be catalyzed respectively, by ADPglucose synthetase, EC 2.7.7.27, (ATP + α-glucose 1-P <---> ADPglucose + PP_i) and starch synthase, EC 2.4.1.21, (ADPGlc + α-1,4 glucan ---> α-1,4 glucosyl-α-1,4 glucan + ADP). Mutants isolated from maize endosperm as well as from Arabidopsis thaliana leaf that are deficient or lacking in starch strongly indicate that the ADPGlc pathway is the major if not the sole pathway in plants for synthesis of starch. Synthesis of the α-1,6 glucosidic branch points found in amylopectin are catalyzed by branching enzyme, EC2.4.1.18. Maize endosperm mutants, amylose extender, ae, lacking one of the branching enzyme isoforms contains a higher percentage of amylose and anomalous amylopectin than the normal maize endosperm thus indicating the physiological role of the branching enzyme. It has been shown that the plant ADPGlc synthetase activity, whether isolated from either leaf or non-photosynthetic reserve tissue, is activated by 3-phosphoglycerate (3PGA) and inhibited by orthophosphate (P_i). The activator, 3PGA, at higher concentrations, can reverse or prevent inhibition by P_i. A number of studies have shown that the above biosynthetic enzymes are completely localized in the subcellular compartment where starch is formed; the chloroplasts in leaves and the amyloplasts in non-photosynthetic tissues. Recent data have also shown that the plant ADPGlc synthetase is composed of two different subunits, most probably, the products of two different genes. The subunit masses are 51 and 54 kDa for the spinach leaf and Arabidopsis enzymes, 60 and 55 kDa for the maize endosperm enzyme and 50 and 51 kDa for the potato tuber enzyme. The activator binding site of the smaller subunit of the spinach leaf enzyme has been elucidated with respect to the amino acid, (ε-amino group of lysyl residue) and the amino acid sequence surrounding it. It is localized near the carboxyl-terminal of the peptide. Similar sequences have been observed for subunits of the ADPGlc synthetases of at least 5 different plants, from photosynthetic or non-photosynthetic tissue as determined from amino acid sequences deduced from nucleotide sequences of cDNA clones. Using 8-azido ATP as a photoaffinity analogue, it has been shown that the larger subunit of the spinach leaf enzyme participates in the binding of the substrate. Other experiments are underway to determine the amino acid sequence of the substrate site. cDNA clones of the smaller subunit have been isolated from rice endosperm, spinach leaf and potato tuber and appear to be coding for the smaller subunit. At present a cDNA clone for the larger subunit has not been isolated put partial amino acid sequence analysis of the spinach leaf 54 kDa subunit indicates approximately a 30% identity with the 51 kDa subunit.

INTRODUCTION

The biosynthetic reactions of starch synthesis

A preponderance of evidence strongly indicates that the ADPglucose (ADPGlc) pathway consisting of the enzymatic reactions catalyzed respectively by ADPGlc pyrophosphorylase (EC 2.7.7.27; reaction 1), starch synthase (EC 2.4.1.21; reaction 2) and branching enzyme (EC 2.4.1.18; reaction 3) is the predominant pathway if not the sole pathway towards starch synthesis.

(1) α-glucosyl-1-P + ATP <======> ADPGlc + PP_i
(2) ADPGlc + $(glucosyl)_n$ -------> ADP + $(glucosyl)_{n+1}$
(3) Linear glucosyl chain of α-glucan -----> Branched chain of glucan with α-1->6-linkage branch point

The data supporting this view are from a number of biochemical and genetic studies. Studies of isolated mutants of maize endosperm (refs. 1 & 2) having deficient levels of ADPGlc pyrophosphorylase are also deficient in starch levels. Recently, it has been shown that the rb locus controls the level of ADPglc pyrophosphorylase in developing pea embryos (ref. 3). One pea line having recessive rb genes contained 38- to 72% of the starch found in the pea line having the Rb loci and about 3- to 5% of the ADPGlc pyrophosphorylase activity. Indeed, a mutant of Arabidopsis thaliana (ref. 4), containing less than 0.2% of the leaf starch content observed in the normal strain had only 0.2% of the normal leaf ADPGlc pyrophosphorylase activity. These data in three diverse species of plants strongly indicate that the ADPGlc pathway must play a major role in starch synthesis. Other data demonstrating a

direct relationship between increase in activity of the above starch biosynthetic enzymes and increased starch accumulation in various plants have been previously reviewed (refs. 5 through 7). In many of those studies no relationship was noted between starch synthesis and phosphorylase or UDPglucose pyrophosphorylase activities. Biochemical studies have also shown that the kinetic properties of the enzymes in the ADPGlc pathway (K_m and V_{max} values) are consistent with the known concentrations of substrate and effector metabolites present in the various plant cells (refs. 5 & 6) in contrast to the known properties of the UDPglucose specific starch synthase and plant phosphorylases. For these two enzymes the high K_m values for the respective substrates (UDPglucose and glucose-1-P) compared to the actual cellular levels preclude a significant if any role in starch synthesis. In view that phosphorylase catalyzes an equilibrium reaction in cells having P_i concentrations in excess of glucose-1-P it is most likely that phosphorylase plays a degradative role in starch degradation rather than synthesis.

Regulation of the ADPGlc pathway

There is much in vitro evidence suggesting that ADPGlc synthesis is regulated by activation of the plant ADPglc pyrophosphorylase by 3-phosphoglycerate (3PGA) and inhibition by P_i. In vivo or in situ evidence showing a correlation between the concentrations of 3PGA and starch and inverse correlations between P_i and starch levels were reviewed (refs. 5 & 6). Recently, Pettersson and Pettersson (ref. 8) have applied modern control theory as developed by Kacser and Burns (refs. 9 & 10) to develop a kinetic model to determine the extent that stromal metabolites, known to affect in vitro leaf ADPglc pyrophosphorylase activity, controlled the rate of photosynthetic starch production under conditions of light and CO_2 saturation. The model consists of the 13 enzyme-catalyzed steps of the reductive pentose phosphate pathway with starch synthesis and photosynthate export (P_i/triose-P translocator) from the chloroplast as output processes from the pentose phosphate pathway. Using this model, the steady state concentrations of various stromal metabolites, the corresponding rates of CO_2 fixation and starch production were defined as a function of the P_i concentration external to the chloroplast. The model essentially agreed with reported experimental data with respect to metabolite concentrations and to CO_2 fixation and starch synthesis rates. It showed that ATP and glucose-1-P, substrates of the ADPGlc pyrophosphorylase, 3PGA, fructose-6-P, and P_i provided significant contributions to starch synthesis rate changes induced by increasing P_i concentrations external to the chloroplast. At low P_i concentrations, the most significant contribution to increased starch synthetic rate was ATP and to a smaller extent 3PGA. At P_i concentrations higher than 0.12 mM, 3PGA becomes the predominant regulator of starch synthesis with glucose-1-P and fructose-6-P contributing to regulation to a significant but smaller extent. Thus, using modern mathematical control analysis (refs. 9 & 10) a kinetic model consistent with in vivo data with respect to metabolite concentrations, mass action ratios observed in vivo and taking into account various enzyme properties and equilibria, they reached the conclusion that 3PGA and P_i play an important role in regulating starch sythesis with significant contributions made by ATP, glucose-1-P and fructose-6-P. Since these metabolites are either substrates or effectors of the ADPGlc pyrophosphorylase, the analysis is very consistent with the view that 3PGA is a positve effector and P_i is a negative effector of ADPGlc synthesis and that the 3PGA/P_i ratio therefore regulates starch synthesis by regulation of ADPGlc pyrophosphorylase.

The availability of chloroplast mutants of phosphoglucose isomerase of Clarkia xantiana (refs. 11 & 12) of phosphoglucomutase (ref. 13) and of ADPGlc pyrophosphorylase of Arabidopsis thaliana (refs. 14 & 15), has allowed the analysis of the extent of control that these enzymes exert on chloroplast starch synthesis. Mutant plants with reduced activity of both cytosolic (64%, 36%, 18% of wildtype) and chloroplastic (75%, 50% of wildtype) phosphoglucoisomerase were used to determine the effect of the enzyme level on fluxes towards starch and sucrose synthesis as well as on photosynthetic rate and control coefficents (ref. 9) of these enzymes on these pathways (ref. 11). Saturating or limiting light intensities were used in the various experiments. The plastid P-glucoisomerase exerted very little control over starch or sucrose synthesis in low light but did exert control of starch synthesis in saturating light. Lowering the cytosolic enzyme activity had little effect on either starch or sucrose synthesis in saturating light but increased starch synthetic rate in low light and decreased sucrose synthetic rate. Thus partitioning of carbon between sucrose and starch was affected by variation of the cytosolic phosphoglucoisomerase levels. Further studies (ref. 12) confirmed that reduction of plastid phosphoglucoisomerase had little effect in low light but reduced starch synthesis by 50% in saturating light with no corresponding increase in sucrose synthesis. The reduced levels of cytosolic enzyme (18% of wildtype) lowered the sucrose synthetic rates and increased the rate of starch synthesis. Metabolite levels were also affected in these mutants. In the mutant containing only 18% of the wildtype cytosolic phosphoglucoisomerase activity both fructose-2,6-bisphosphate and 3PGA levels increase about 100%. Neuhaus et. al. (ref. 12) indicate that their results provide strong evidence that reduced sucrose synthesis rate is due to the increased Fru-2,6-P_2 level which causes increased inhibition of cytosolic fructose-1,6-bisphosphatase (for reviews on sucrose synthesis see refs. 16 & 17) which is on the pathway towards sucrose synthesis They also indicate that the data strongly support the view that increased starch synthesis in the mutants with reduced levels of phosphoglucoisomerase is due to activation of the ADPGlc pyrophosphorylase by the increased 3PGA concentration and 3PGA/P_i ratio. Thus these experiments provide further support for the in vivo regulation of starch synthesis by 3PGA.

These experiments have been extended to the null chloroplast phosphoglucomutase (ref. 13) and the low activity (7% of wildtype) ADPGlc pyrophosphorylase mutants (14) of Arabidopsis thaliana. Neuhaus and Stitt (ref. 15) utilized the alleles to construct hybrid plants containing respectively, 50% of wildtype phosphoglucomutase activity and 50% of wildtype ADPGlc pyrophosphorylase activity. The effects of these reduced activities on starch and sucrose fluxes and on CO_2 fixation in low light and high light intensity were measured. In low light, a 50% decrease in

phosphoglucomutase activity had no significant effect on the above fluxes. However, a 50% and 93% decrease of ADPGlc pyrophosphorylase activity resulted in a 23% and 74% decrease in flux of starch synthesis with a concomitant increase of 17% and 42% increase in sucrose synthetic rate. Thus diminution of ADPGlc synthesis activity not only significantly affected starch synthesis but also affected the partitioning of photosynthetic carbon causing more to be directed towards sucrose biosynthesis. In high light a 50% decrease in phosphoglucomutase activity resulted in a 20% decrease in starch synthesis with little effect on sucrose synthesis rate. However reduction of the the ADPGlc synthesizing activity to 50% and 93% resulted in a 39% and 90% decrease in starch synthesis flux. The flux of photosynthetic carbon under these conditions was not redirected towards sucrose synthesis but rather the photosynthetic rate was inhibited about 46%.

TABLE 1: Estimated flux control coefficients for starch synthesis. The values were obtained from (11,15)

Enzyme	Flux control Coefficient	
	Low Light	High Light
Chloroplast Phosphoglucoisomerase	0.0	0.35
Chloroplast Phosphoglucomutase	0.01	0.21
ADPGlc Pyrophosphorylase	0.28	0.64

The flux control coefficents (ref. 10) for the enzymes for starch synthesis were calculated to determine the distribution of control and compared with previous results obtained with the *Clarkia xantiana* phosphoglucoisomerase and these are seen in Table 1. Of the enzymes under study the flux to starch synthesis is only regulated by ADPGlc pyrophosphorylase in low light. In high light and CO_2, although ADPGlc pyrophosphorylase activity exerts major control, other enzymes such as plastid phosphoglucomutase and phosphoglucoisomerasedo exert control to a small but significant extent.

In summary, various analyses of the starch biosynthetic system in a number of plants or utilizing *in vivo* data obtained from different plants and applying the Kacser and Burns control analysis method (refs. 9 & 10) point out that the major site of regulation of starch synthesis is at ADPGlc pyrophosphorylase and that 3PGA and P_i are important regulatory metabolites of that enzyme.

PROPERTIES OF THE STARCH BIOSYNTHETIC ENZYMES. ADPGLUCOSE PYROPHOSPHORYLASE

Summary of regulatory and structural properties. Regulation by 3-phosphoglycerate and by P_i

The properties of the plant and bacterial ADPGlc pyrophosphorylases have been extensively reviewed and for detailed information on the plant systems the reader is referred to (refs. 5 through 7 and 18). For the bacterial systems the reader is referred to (refs. 19 & 20). As indicated earlier in this chapter, and in the above reviews on plant systems, every plant ADPGlc pyrophosphorylase studied has as its major activator, 3-phosphoglycerate (3PGA) and inhibitor, P_i. For every leaf system studied, whether the leaf source is from a plant utilizing C_3 or C_4 pathways or Crassulacean metabolism, the major activator is still 3PGA and the inhibitor is P_i. In non-photosynthetic systems studied (eg., maize endosperm, potato tuber) the ADPGlc pyrophosphorylase activity is highly dependent on the presence of 3PGA and is inhibited by P_i. However, it appears that the wheat endosperm enzyme is not activated by 3PGA (M.Olive and P. Keeling, personal communication). The wheat endosperm ADPGlc pyrophosphorylase is inhibited by P_i and as observed in other plant systems, photosynthetic or non-photosynthetic, 3PGA overcomes the inhibition caused by P_i. Thus even in wheat endosperm regulation of starch synthesis may occur via fluctuation of the [3PGA]/[P_i] ratio. That the *in vitro* observations of 3PGA activation and P_i inhibition are physiologically important are supported by many *in vivo* and *in situ* experiments (refs. 5 through 7 and 18).

Structural properties of ADPlc pyrophosphorylase

The most studied ADPglucose pyrophosphorylase with respect to structural properties is the spinach leaf enzyme (refs. 21 through 24). This enzyme has a molecular mass of 206,000 and is composed of two different subunits, molecular masses of 51 and 54 kd. These subunits can be distinguished not only by differences in their molecular mass but also with respect to different amino acid composition, different amino-terminal sequences, different peptide patterns on HPLC of tryptic digests and distinct antigenic properties. The polyclonal antibody prepared against the 51 kd subunit reacted very strongly, in Western blotting experiments, with the 51 kilodalton subunit but reacted weakly with the 54 kd subunit. Conversely antibodies raised against the 54 kd subunit reacted only weakly with the 51kd subunit and strongly with the 54 kd protein. Thus, on the basis of the protein chemistry and immunological analyses, the two subunits are quite distinct and most probably are the products of two genes. In contrast, the bacterial ADPGlc pyrophosphorylase is composed of only one subunit, 50 to 55 kd in mass depending upon the source and the native enzymes are homotetrameric (ref. 19).

Recently, other plant enzymes have been studied in greater detail and they have been shown to be composed of two dissimilar subunits. The maize endosperm ADPGlc pyrophosphorylase, which has a molecular mass of 230,000, could react with the antibody prepared against the native spinach leaf enzyme in Western blot experiments (ref. 25). In endosperm extracts, as well as with the highly purified enzyme, two immunoreactive proteins are found of molecular masses of 55 and 60 kd after SDS gel electrophoresis. Antibody prepared against the spinach leaf 54 kd subunit cross-reacts mainly with the endosperm 60 kd subunit and to a small extent with the 55 kd subunit while the spinach leaf 51 kd subunit antibody cross-reacts very well with the endosperm 55 kd subunit and very weakly with the 60 kd subunit. Of interest was the analysis of the maize endosperm mutants, shrunken 2 and brittle 2 . In Western blotting experiments and using the native and subunit antibodies of the spinach leaf enzyme, it is found that the mutant bt_2 endosperm lacks the 55 kd subunit and the mutant sh_2 endosperm lacks the 60 kd subunit. These results (ref. 25) strongly suggest that the maize endosperm ADPGlc pyrophosphorylase is composed of two immunologically distinctive subunits and that the sh_2 and bt_2 mutations cause reduction in ADPGlc pyrophosphorylase activity through the lack of one of the subunits. Thus the most likely possibility is that the Sh_2 gene is the structural gene for the 60 kd protein while the Bt_2 gene is the structural gene for the 55 kd protein. Consistent with this is the isolation of an ADPGlc pyrophosphorylase cDNA clone from a maize endosperm library (ref. 25) and showing that it hybridized with the small subunit cDNA clone from rice (ref. 26). This maize ADPGlc pyrophosphorylase cDNA clone is found to hybridize to a transcript present in maize endosperm but absent in bt_2 endosperm. Thus, the bt_2 mutant appears to be the structural gene of the 55 kd subunit of the ADPGlc pyrophosphorylase. Most important is that the above data certainly point out that the non-photosynthetic tissue ADPGlc pyrophosphorylase is also composed of two subunits and on the basis of immunoreactivity there is corresponding homology between the subunits in the leaf enzyme with the subunits of a reserve tissue enzyme.

The potato tuber ADPGlc pyrophosphorylase has been highly purified and by 2-dimensional polyacrylamide gel electrophoresis, two different proteins could be distinguished by their slight differences in molecular mass, 50 and 51 kd and in net charge (refs. 27 & 28). The smaller tuber subunit is reactive with the antibody prepared against the smaller spinach leaf (51 kd) subunit.The antiserum prepared against the spinach leaf larger (54 kd) subunit however, is not reactive with either potato tuber enzyme subunit. Thus, as found with the spinach leaf and maize endosperm enzyme, the potato tuber enzyme is composed of two distinct subunits and not one as previously thought (ref. 29). It appears that other plant ADPGlc pyrophosphorylases may also be composed of 2 subunits on the basis of Western blotting experiments. Using anti-sera toward native spinach leaf ADPGlc pyrophosphorylase the presence of two immunoreactive bands in gel electrophoresis of extracts of wheat, rice and maize leaves was observed (ref. 30). Similar data was obtained with Arabidopsis thaliana ADPGlc pyrophosphorylase (refs. 4 & 14). It is composed of two subunits, molecular masses of 51 and 54 kd. Indeed, one of the Arabidopsis thaliana ADPGlc pyrophosphorylase mutants, TL25, is lacking both subunits and it is thought that the mutation is affecting a regulatory locus (ref. 4) while another mutant, TL46, lacks the larger, 54 kd subunit (ref. 14). This mutation provides further evidence that the larger subunit is a necessary component of the native ADPGlc pyrophosphorylase for optimal activity since the mutant has only 7% of the wildtype activity. The mutant synthesizes starch at 9% the rate seen for the normal strain in high light and only at 26% of the rate at low light (ref. 15).

The allosteric 3PGA activator binding site of ADPglucose pyrophosphorylase: determination of the amino acid sequence at the 3PGA allosteric binding site

With the findings that the plant native ADPGlc pyrophosphorylases are tetrameric and composed of two different subunits it was of interest to know why the the two subunits were required for optimal catalytic activity. The bacterial ADPGlc pyrophosphorylase from at least 7 organisms contains only one subunit and the native structure is homotetrameric (refs. 19 & 20). Since the enzyme must contain ligand binding sites for the activator, 3PGA, and inhibitor, P_i, sites for the two substrates, ATP and glucose-1-P, as well as a catalytic site, it is easy to conceive that these sites could be located on different subunits.

Initial studies were designed to determine the nature and location of the 3PGA binding site. These studies were facilitated by the finding that pyridoxal-5-phosphate was an activator of spinach leaf ADPGlc pyrophosphosphorylase (refs. 21 & 22). Pyridoxal-P is not as effective as 3PGA with regard to activation of velocity; only a 6-fold stimulation is observed compared to a 25-fold stimulation caused by 3PGA. Three other ADPGlc pyrophosphorylases (maize endosperm, Arabidopsis thaliana and from the cyanobacterium, Synecohcystis 6803) are also stimulated by pyridoxal-5-P (unpublished results). The apparent affinity, or $A_{0.5}$ value (concentration of activator giving 50% of maximal stimulation), for the spinach leaf enzyme for pyridoxal-5-P is about 15 μM and is somewhat lower than the $A_{0.5}$ of 45 μM for 3PGA. Thus, there is a higher apparent affinity for pyridoxal-5-P even though there is a lesser stimulation of maximal velocity.

The view that pyridoxal-5-P is binding to the same site as 3PGA is based on the following; pyridoxal-5-P inhibits the activation caused by 3PGA (refs. 21 & 22) Pyridoxal-5-P can also reverse inhibition caused by P_i in the same manner as 3PGA (21). In addition, pyridoxal-5-P can be covalently bound to the spinach leaf ADPGlc pyrophosphorylase by reduction with $NaBH_4$, (refs. 21,22,24,31), producing a modified enzyme that is highly active in the absence of activator (Table 2). The activator, 3PGA, as well as the inhibitor, P_i, prevents or inhibits the reductive covalent binding of the pyridoxal-5-P and reduces the increase in activity caused by the reductive phosphopyridoxylation (refs. 21,22, 24,31). The modified enzyme is also quite insensitive to P_i inhibition(21,22,31). All the above results are consistent with the view that pyridoxal-5-P binds to the activator site and that the covalent modification with pyridoxal-5-P places the enzyme in a conformational state where it is resistant to P_i inhibition.

Table 2. Chemical modification of spinach leaf ADPglucose pyrophosphorylase by pyridoxal-P and $NaBH_4$. +3PGA/-3PGA is the ratio of activity obtained in the presence and absence of the activator, 3PGA in the assay of ADPGlc pyrophosphorylase actvity. [^{3}H]-Pyridoxal-P was used to measure incorporation of Pyridoxal-P into Enzyme (31).

Conditions	+3PGA/-3PGA	mol Pyridoxal-P/ mol Enzyme	% Protection
None, no $NaBH_4$	25	---	---
None	5.4	1.3	0
ADPGlc, 1mM + 6mM $MgCl_2$	7.4	1.1	10
3PGA, 1mM	18	0.1	91
2PGA, 1mM	8	1.1	10
Pi, 1mM	32	0.3	75
Glucose 6-P, 1mM	4.5	1.28	0

Incorporation of [^{3}H]-pyridoxal-5-P into the spinach leaf enzyme ADPGlc pyrophosphorylase has been observed and there is a relationship between the amount of pyridoxal-5-P incorporated and the increase in activity caused by the modification (refs. 22 & 31). Many metabolites and substrates were tested and only P_i and 3PGA inhibited incorporation of the labelled pyridoxal-5-P into the enzyme (Table 2). Labelled pyridoxal-5-P is incorporated about equally into both the 54 and 51 kilodalton subunits (ref. 31). As seen in Table 2, the activator, 3PGA, prevented incorporation of the pyridoxal-5-P to the same extent into both subunits (ref. 31). The labelled 51 kd subunit has been subjected to trypsin digestion. The labelled peptide isolated using reverse-phase HPLC was sequenced (refs. 24 & 31). The sequence obtained is shown in Fig. 1 and is compared to the fructose 1,6-bis P activator site sequence of the Escherichia coli ADPGlc pyrophosphorylase (ref. 32). It is not surprising that the amino acid sequences corresponding to the activator sites are different. There are many more basic amino acid residues in the E. coli sequence than in the spinach leaf 51 kd activator sequence. The amino acid sequence of the spinach leaf 54 kd subunit activator site has not been elucidated as yet. At present, up to 2 or 3 radioactive peptides have been isolated from tryptic digests and appear to be different from the peptide isolated from the 51 kilodalton peptide (ref. 31). It would be of interest to know if the subunit activator sites are independent in causing the enzyme to go into an active conformation upon binding of 3PGA or if both peptides interact to provide the active conformation upon 3PGA binding.

ADPglucose pyrophosphorylase	Peptide sequence of activator site
Spinach leaf	SGIVTVIK*DALIPS
Escherichia coli	RLKDLTNK*RAKPAV

Fig. 1. The amino acid sequence of the spinach leaf activator (3PGA) site and the E. coli activator (fructose 1,6-P) site of ADPglucose pyrophosphorylase. * indicates the lysine residue the activator, pyridoxal-P, covalently modifies.

Studies on the catalytic site of the spinach leaf ADPGlc pyrophosphorylase using the photoaffinity substrate analog, 8-azido-ATP

8-azido-ATP (8-N_3-ATP) has been shown to be a substrate for the spinach leaf ADPGlc pyrophosphorylase with a K_m larger than that of the natural substrate, ATP (ref. 33). The K_m is 0.81 μM for 8-N_3-ATP and 0.12mM for ATP. Moreover, the Vmax with 8-N_3-ATP is only 1% of that observed with ATP. The product of the reaction is 8-N_3-ADPglucose. 8-N_3-ADPGlc has a K_m for the spinach leaf ADPGlc pyrophosphorylase of 80 μM whereas the K_m for ADPGlc is 225 μM. The maximal velocity rate for 8-N_3-ADPGlc is only 0.3% of that for ADPGlc. When UV light (257 nm) irradiates 8-azido compounds nitrene radical is formed which can react with electron-rich residues. Indeed, it is so reactive that it can form a secondary amine reacting with a C-H linkage. Preliminary experiments (ref. 33) have shown that labelled 8-N_3-ADPGlc can covalently link to the spinach leaf ADPGlc pyrophosphorylase and inactivate it upon UV irradiation. The substrate, ADPGlc, was very effective in inhibiting the chemical modification by the azido compound while UDPglucose, a non-substrate, did not prevent the inactivation. The labelled 8-azido-ADPGlc is incorporated mainly, if not solely, in the 54 kilodalton subunit (ref. 33). ADPGlc inhibits the incorporation by at least 67%. It thus appears that the substrate binding site is on the 54 kd subunit. It is probable, however, that the substrate/catalytic site is shared between the two subunits and that the nitrene radical reacts with a strong nucleophilic residue on the larger subunit causing the incorporation to be mainly on the larger subunit.

Isolation of the ADPglucose pyrophosphorylase gene. Cloning of ADPglucose pyrophosphorylase genes and their sequence comparisons

A cDNA clone of a rice seed ADPGlc pyrophosphorylase gene corresponding to the small subunit has been isolated and its DNA nucleotide sequence has been determined (ref. 34). The nucleotide sequence of the 1647 bp of the rice seed cDNA clone has an open reading frame of 483 amino acids which begins at bp 1 with ATG and ends at bp position 1460 with the termination codon, TAA. The predicted protein molecular weight is about 52,000 and consistent with the SDS gel polyacrylamide gel electrophoresis molecular mass seen for the in vitro translated product of the clone (ref. 30). The cDNA has after the termination codon a 3'-untranslated end of 201 bp and a poly (A) tail. The concensus polyadenylation signal, AAUAAA, typical of most nonhistone eukaryotic mRNAs (ref. 35), is present but is 117 bp upstream from the polyadenylation addition site. However, an alternative form, AAUGAU is observed within the usual 10 to 30 bp of the poly(A) site and may serve as the signal. The additional 27 nucleotides at the 5' end was determined by using a synthetic 16-base oligonucleotide corresponding to nucleotides 20 to 35 of the mRNA transcript annealed to poly(A^+)-RNA and subjected to chain termination reactions using reverse transcriptase (ref. 34). Two putative nonsense codons are found in this upstream area with one of them 15 bases away and in-frame with the coding sequence. The presence of this upstream nonsense codon indicates that the open reading frame present in the insert DNA represents a complete ADPGlc pyrophosphorylase coding sequence.

```
W.E.7                                            RASPPSESRA   PLRAPQRSAT   RQHQ          -24

RICE      MNVLASKIFP    SRSNVVSEQQ    QSKREKATID   DAKNSSKNKN   LDRSVDESVL    -50
W.E.7     ARQGPRRMCN    GGRGPPYWTA    GVTSAPARQT   PLFSGRPSGG    LSDPNEVAA    -73
W.E.3
W.L.
SL-51k                       VSDSQ    NSQDGLDPE      CETCLDP       EASRSVL    -28
Potato                                NSQTCLDPD                      ARSVL    -14
          :.::::: ::    .:.:::  ::    :::::... :   :::::: :::   .:: : ::..
R         GIILGGGAGT    RLYPLTKKRA    KPAVPLGANY   RLIDIPVSNC   LNSNISKIYV   -100
W.E 7      VILGGGTGT    QLFPLTSTRA    TPAVPIGGCY   RLIDIPMSNC   FNSGINKIFV   -122
W.E. 3
W.L
51kd      GIILGGGA T    RLYPLTKKRA    KPAVPLGANY   RLIDIPVSNC   LNSNISKIYV    -77
Potato    GIILGGGAGT    RLYPLTKKRA    KPAVPLGANY   RLIDIPVSNC   LNSNISKIYV    -64
           ::::::: :    ::. :.:       .:      .:    ::::: : :      . ::::
R         LTQFNSASPN    RHLSRAYGNN    IGGYKNEGFV   EVLAAQQSPD   NPNWFQGTAD   -150
W.E. 7    MTQFNSASLN    RHIHRTYLGG     GINFTDGSV   EVLAATQMPG   EAAGWFRGTAD  -172
W.E. 3
W.L.
51kd      LTQFNSASLN    RHLSRAYASN    LGGYKNEGFV   EVLAAQQSPE   NPDWFQGTAD   -127
Potato    LTQFNSASLN    RHLSRAYASN    MGGYNNEGFV   EVLAAQQSPE   NPDWFQGTAD   -114
          :.: ..:..:           .  .   .::.:::::.     ..| |        |.... |
R         AVRQYLWLFE    EHNVMEFLIL    AGDHLYRMDY   EKFIQAHRET   DSDITVAALP   -200
W.E. 7    AWRKIIWVLED   YYKNKSIEHILIL SGDQLYRMDY   MELVQKHVDD   NADITLSCAP   -226
W.E. 3                                              ELVQKHVDD   NADITLSCAP    -19
W.L.                         GVLIL SGDHLYRMDY   MDFVQSHRQR   DAGISICCLP    -35
51kd      AVRQYLWLFE    EHNV            DHLYRM-I   MKIIQAHRET   DADITVAALP   -168
Potato    AVRQYLWLFE    EHTVLEYLIL    AGDHLYRMDY   EKFIQAHRET   DADITVAALP   -164
Arabid.                                                           DITVAALP     -8
             ||. .|     |.|.|  ||.     . | |.|      |  | :::.   |    .   |
R         MDEKRATAFG    LMKIDEEGRI    VEFAEKPKGE   QLKAMMVDTT   ILGLDDVRAK   -250
W.E. 7    VGESRASEYG    LVKFDSSGRV    VQFSEQPKGD   DLEAMKVDTSF  LNFAIDDPAK   -277
W.E. 3    VGESRASEYG    LVKFDSSGRV    VQFSEKPKGD   DLEAMKVDTSF  LNFAIDDPAK    -70
W.L.      IDGSRASDFG    LMKIDDTGRV    ISFSEKPRGAD   LKEM              EEAEK    -75
51kd      MDEKRATAFG    LMKIDEEGRI    IEFAEKPKGE   QLQAMKVDTT   ILGLDDERAK   -218
Potato    MDEKRATAFG    LMKIDEEGRI    IEFAEKPQGE   QLQAMKVDTT   ILGLDDKRAK   -214
Arabid.   MDEQRATAFG    LMKIDEEGRI    IEFAEKPGEH     LKAMKVDTT   ILGLDDERAK    -58

             |||||||.   |.. . ..|     ||.  .      |||||..| |   :     |.||
R         EMPYIASMGI    YVISKNVMLQ    LLREQFPGAN   DFGSEVIPGA   TNIGMRVQAY   -300
W.E 7      YPYIASMGV    YVFKRDVLLN    LLKSRYAELH   DFGSEILPRA     LHDHNVQAY  -325
W.E.3      YPYIASMGV    YVFKRDVLLN    LLKSRYAELH   DFGSEILPRA     LHDHNVQAY  -118
W.L.      K PYIASMGV    YIFKKEILLN    LLRWRFPTAN   DFGSEIIPAA     AREINVKAY  -123
51kd      EMPYIASMGI    YVISKDVMLN    LLRDKFPGAN   DFGSEVIPGA   TSLGLRV ---  -265
Potato    EMPFIASMGI    YVISKDVMLN    LLRDKFPGAN   DFGSEVIPGA   TSLGMRVQAY   -264
Arabid.   EMPFIASMGI    YVVSRDVMLD    LLRNQFPGAN   DFGSEVIPGA   TPLGLRVQAY   -108
          ..   |||      |||| .|..|    | .              .       :.|  | |||
R         LYD YWED      IGTIEAFYNA    NLGITKKPVP   DFSFYDRSAP   IYTQPRHLPP   -348
W.E 7     VFTDYWED      IGTIRSFFDA    NRALCEQPPK   FEFYDPKTPF    FTSPRYLPP   -372
W.E. 3    VFTDYWED      IGTIRSFFDA    NMSLCEQPPK   FEFYDPKTPF    FTSPRYLPP   -165
W.L.      LFNDYWED      IGTIKSFFEA    NLALAEQPSK   FSFYDASKPM    YTSRRNLPP   -170
51kd      --- YWED      IGTIEAFYNA    NLGI TKKPVP  DFSFYDRSSP   IYTQPRYLPP   -309
Potato    LYDGYWED      IGTIEAFYNA    NLGITKKPVP   DFSFYDRSAP   IYTQPRYLPP   -312
Arabid.   LYDGYWEN      IGTIEAFYNA    NLGITKKPVP   DF---                     -138
          .             . . ....           |
R         SK   VLD      ADVTDSVIGE    GCVIKNCKIH   HSVVGLRSCI   SEGAIIEDS    -392
W.E. 7    TK   SDK      CRIKEAIILH    GCFLRECKIE   HT AFSRLNSG  SELKNAMMMG   -417
W.E. 3    TK   SDK      CRIKEAIILH    GCFLRECKIE   HSIIGVPSRLN  SGSELKNAMMMG -213
W.L.      SM   ISG      SKITDSII      SHGCFLDKCR   VEHSVVGIRSR  IGSNVHLKDTV  -215
51kd      SK   MLD      ADITDSVIGE    GCVIKNCKIH   HSVIGLRSSI   SEGAIIEDT    -353
Potato    SK   MLD      ADVTDSVIGE    GCVIKNCKIH   HSVVGLRSCIE  SEGAIIEDS    -357
                                                   . :. :.:     || |||||
R         LLMGADYY      ETEADKKLLG    EKGG  IPIGI  GKNCHIRRAI   IDKNARIGD    -438
W.E.7     ADSYETEDEMS   RLMSEGKVPIG   VGENTKISNCI               IDMNARIGR    -459
W.E.3     ADSYETEDEIS   RLMSEGKVPIG   VGENTKISNCI               IDMNARIGR    -255
W.L.      MLGADFYETD    MERGDQLAEGK   V PIGIGENTSI  QNCI        IDKNARIGK    -260
51kd      LLMGADYY      ETDADRKLLA    AKGS VSIGI   GQNSHIKRAI   IDKNARIGD    -399
Potato    LLMGADYY      ETDADRKLLA    ALGS VPIGI   GKNCHIKRAI   IDKNARIGD    -403
            . | |   |   ||| |  .|.     |.||||  .   * . .        .
R         NVKIINVDNV    QEAARETDGY    FIKSGIVTVI   KDALLLAEQL   YEVAA        -483
W.E. 7    DVVISNKEGV    QEADRPEEGY    YIRSGIVVIQ   KNATIKDGT    VV           -500
W.E. 3    DVVISNKEGV    QEADRPEEGY    YIRSGIVVIQ   KNATIKDGT    VV           -296
W.L.      NVTIANAEGV    QEADRASEGF    HIRSGITVVL   KNSVIADGL    VI           -301
51KD      NIKIINSDNV    QEAARETDGY    FIKSGIVTVI   KDALIPSGT    VI           -457
Potato    NVKIINKDNV    QEAARETDGY    FIKSGIVTVI   KDALIPSGI    II           -444
                                                   *
```

Fig. 2. Alignment of the deduced amino acid sequences of rice seed (R), wheat seed (W.E.7 and W.E.3), wheat leaf (W.L.), spinach leaf 51 kd subunit (S.L.-51k), Arabidopsis and potato tuber ADPGlc pyrophosphorylase polypeptides. The sequences are derived from the respective cDNA and genomic clones as indicated in the text. Conserved amino acids in the alignment are indicated on top as : , the homologous amino acids are indicated as · and amino acids conserved for all sequences are indicated as | . * indicates the presumed 3PGA binding site. Numbers on the right indicate the amino acid residue from the translational start site.

Seen in Fig. 2 is the deduced amino acid sequence. As expected, because of the localization of the ADPGc pyrophosphorylase in either the amyloplast or chloroplast, and it being encoded by nuclear genes, the subunit posseses a putative leader sequence at its N-terminus. The actual N-terminal sequence of the mature subunit is not known at the present time but alignment of the rice seed primary sequence with the sequence of the 51 kd subunit of the spinach leaf ADPGlc pyrophosphorylase suggests that cleavage of the leader sequence occurs between residues 28 and 29, Thr-Ile (ref. 34). This putative leader sequence however, shares very little sequence conservation particularly with the 3 blocks of homology suggested (ref. 36) necessary for chloroplast transport and processing. Nevertheless, the subunit does contain peptide domains similar in properties to those observed in other chloroplast transit peptides. The N-terminus has a 6-residue hydrophobic domain and a highly basic region at amino acid residues, 23-26 and a hydrophillic domain at amino acid residues, 17-22. These features in the sequence may serve as signals necessary for plastid targeting and processing.

At present, cDNA or genomic clones for the small subunit ADPGlc pyrophosphorylase gene of rice endosperm (refs. 30,34,37), maize endosperm (ref. 26), spinach leaf (ref. 33), Arabidopsis thaliana (A. Lonneborg, private communication) and potato tuber (ref. 27) have been isolated. In addition, a cDNA clone for the maize endosperm ADPGlc pyrophosphorylase larger molecular weight subunit (Shrunken 2 locus) has also been isolated (ref. 26). cDNA clones have been isolated from wheat leaf and wheat endosperm (ref. 38). As will be shown later, the deduced amino acid sequence derived from the wheat endosperm cDNA would suggest that it is more homologous with the deduced amino acid sequence of the maize endosperm cDNA of the larger (Shrunken-2) subunit than the rice seed cDNA of the smaller pyrophosphorylase subunit gene. While no other large molecular weight subunit cDNA clone has been reported, the major portion of the spinach leaf large molecular weight, 54 kilodalton, subunit has been sequenced by Edmann degradation technique (K. Ball, J. Hutny, J. Leykam and J. Preiss, unpublished results).

At the DNA level the isolated genes are quite dissimilar. For wheat leaf and wheat endosperm there is only 55.7% identity (ref. 38) and on the basis of Southern hybridization analyses and restriction enzyme mapping, it is concluded that there are at least two distinct gene families in wheat. For spinach leaf and rice endosperm there is only about a 50% identity (D. Cress and J. Preiss, unpublished results). Fig. 2 compares the deduced amino acid sequence of the clones from rice and wheat endosperm and spinach, wheat and Arabidopsis thaliana leaf and from potato tuber. Two wheat endosperm cDNA clones were isolated (ref. 38) and they represented two closely related gene sub-familes in wheat endosperm. One, AGA.3, had a DNA insert of 1272 bp in length with an open reading frame of 888 bp encoding 296 amino acids. The other, AGA.7, is 1798 bp in length comprising of an open reading frame of 1500 bp encoding 500 amino acids. The clones were considered incomplete but it was estimated that the AGA.7 insert was nearly complete, possibly lacking some of the transit peptide portion. These wheat endosperm clones are 96.3 % identical in shared regions of their open reading frames and differ in only 10 of the 296 amino acid residues giving an identity of 96.7%. In contrast, there is only a 55.3% identity between the wheat leaf, AGA.1 DNA insert deduced amino acid sequence and AGA.7. AGA.1 is 947 bp in size and encodes for 301 amino acids. Comparisons of the various clones were made with the deduced amino acid sequence derived from the rice endosperm cDNA clone which was a full length clone of 1647 bp and coded for 483 amino acids (34). The derived amino acid sequences are seen in Fig. 2. AGA.1, AGA.3 and AGA.7 showed only about 31%, 42% and 39% identity with respect to amino acids, respectively, with the rice seed amino acid sequence. If homologous amino acids were considered, the homology between the sequences to the rice sequence is about 42, 61 and 54%, respectively. Greater identity was seen with the potato, spinach and Arabidopsis leaf cDNA clones with respect to amino acid sequences. The cDNA clone of potato tuber is almost a full length clone which codes for 444 amino acids (ref. 27) while the spinach leaf clones are incomplete with two of them encoding for a total of 457 amino acids (ref. 33). The Arabidopsis genomic clone is incomplete and its sequence encodes 138 amino acids (A. Lonneborg, private communication). Of interest is that for the potato tuber, spinach leaf 51 kilodalton and Arabidopsis leaf deduced amino acid sequences there are about 75%, 69% and 86% identity, respectively, with the rice seed amino acid sequence. If homologous amino acids are considered, the homology is then, respectively, 76%, 72% and 90%. All of these 3 last mentioned clones have been shown to be representative genes of the smaller subunit of the ADPGlc pyrophosphorylase. In contrast as seen in Fig. 3, less identity and homology is seen with the derived amino acid sequences of the cDNA clone of the maize Shrunken -2 (higher molecular mass) subunit of the ADPglc pyrophosphorylase with the rice seed and spinach leaf cDNA clones derived amino acid sequences. The Shrunken-2 sequence (C. Hannah, unpublished results) codes for 329 amino acids. Included in the comparison were the sequences of the 54 kilodalton, high molecular mass subunit of the of the spinach leaf enzyme obtained by automated Edmann degradation procedures (K. Ball, J.Leykam, T. Jones and J. Preiss, unpublished results) and then aligning the peptide sequences to the derived sequences of the rice seed and spinach leaf lower molecular mass subunit. There is only a 36% and 37% identity seen in the sequences of Shrunken-2 with those of rice seed and spinach leaf, respectively. If homologous amino acids are considered, then the homology is 50% and 54%, respectively. Similarly, about 42% identity is observed for the spinach leaf peptides (total of 227 amino acids) of the 54 kilodalton subunit compared to the amino acid sequences of rice seed and spinach leaf. If homologous amino acids are considered then the homology is about 50 to 54% (Fig. 3). Comparison of the Sh-2 sequence with the spinach leaf 54 kilodalton subunit shows a higher identity. About 157 amino acids can be compared and 52% identity is seen and homology is 65% when homologous amino acids are considered. Also of interest is the comparison of the amino acid sequences seen for the wheat endosperm cDNA clone, AGA.7 and the Shrunken-2 cDNA clone. There is a 63% identity in sequence and homology is 71% when homologous amino acids are considered.

```
                                       .                          .
RICE     MNVLASKIFP  SRSNVVSEQQ  QSKREKATID  DAKNSSKNKN  LDRSVDESVL  -50
SL-51kd         ---  VSDSQ       NSQDGLDPE   CETCLDP      EASR  SVL  -28
SL-54kd                               SVTAD  NASETKVRDI  GQDK    SS  -21
                        :.::                  ::::.: ::   .:: : :
R        GIILGGGAGT  RLYPLTKKRA  KPAVPLGANY  RLIDIPVSNC  LNSNISKIYV  -100
51kd     GIILGGGA T  RLYPLTKKRA  KPAVPLGANY  RLIDIPVSNC  LNSNISKIYV  -77
54kd                  LFPL                    LIDVPMSNQ  INSGINK     -41
                          ::      ::   .:.:   :::::        :::::::
R        LTQFNSASPN  RHLSRAYGNN  IGGYKNEGFV  EVLAAQQSPD  NPNWFQGTAD  -150
51kd     LTQFNSASLN  RHLSRAYASN  LGGYKNEGFV  EVLAAQQSPE  NPDWFQGTAD  -127
54kd                      AYNFG  SGGNFGDGYV  EVLAA         SWFQGTAD  -69
           .     ..      .  ..::  .:| :::: .    ..| |     .||...   |
R        AVRQYLWLFE  EHNVMEFLIL  AGDHLYRMDY  EKFIQAHRET  DSDITVAALP  -200
51kd     AVRQYLWLFE  EHNV          DHLYRM-I  MKIIQAHRET  DADITVAALP  -168
54kd     AV            EIEDILIL  SGD         MDFLQNH QS  GADISISQLP  -101
Sh-2     IWVLEDYYSH   KSIDNIVIL  S DQLYRMNY  MELVQKHVED  DADITISCAP  -47
         .|.  |.  |  |.|::  |:.  . | |||||   | .| |.|   .|..     :
R        MDEKRATAFG  LMKIDEEGRI  VEFAEKPKGE  QLKAMMVDTT  ILGLDDVRAK  -250
51kd     MDEKRATAFG  LMKIDEEGRI  IEFAEKPKGE  QLQAMKVDTT  ILGLDDERAK  -218
54kd     MDDSSASDFG  LMK      V  LSFSEKPKGD  DLKAMAVDTT  VLGLSK      -141
Sh-2     VDESRASKNG  LVKIDHTGRV  LQFFEKPKGA  DLNSMRVETN  FLSYAIDDAQ  -97
           ::.:::::  ::. :  .:   ::.   .     |||||.|| |        ::::
R        EMPYIASMGI  YVISKNVMLQ  LLREQFPGAN  DFGSEVIPGA  TNIGMRVQAY  -300
51kd     EMPYIASMGI  YVISKDVMLN  LLRDKFPGAN  DFGSEVIPGA  TSLGLRV     -265
54kd                                 FPTAN  DFGSEIIP A          AY  -157
Sh-2     KYPYLASMGI  YVFKKDALLD  LLKSKYTQLH  DFGSEILPRA  VLDHSCVQA   -136
         ..     |||  .||  .|. |  ||. |  |     : |||   |  ..|  : :::
R        LYDG  YWED  IGTIEAFYNA  NLGITKKPVP  DFSFYDRSAP  IYTQPRHLPP  -348
51kd     ----  YWED  IGTIEAFYNA  NLGATKKPVP  DFSFYDRSSP  IYTQPRYLPP  -309
54kd     LFND  YWER  IGT   SFFEA  NLALTIHPSK   FSFYDADKP  MYTSR       -197
Sh-2     IFTG  YWED  VGTACSFFDA  NLALTEQPSK   FDFYDPKTP  FFTAPRCLPP  -183
                             .   ::... : :   :::.:. : .  :  . .
R        SK   VLD    ADVTDSVIGE  GCVIKNCKIH  HSVVGLRSCI  SEGAIIEDS   -392
51kd     SK   MLD    ADITDSVIGE  GCVIKNCKIH  HSVIGLRSSI  SEGAIIEDT   -353
54kd                                                    VNSNVHLK   -205
Sh-2        TQLD     KCKMKYAFISD GCLLRECNIE  HSVIGVCSRV  SSGCELKDS   -228
           ..:::::.:    . .    :        . :::  :.:.:.:...:  :: :::::
R          LLMGADYY  ETEADKKLLG  EKGG IPIGI  GKNCHIRRAI  IDKNARIGD   -438
51kd       LLMGADYY  ETDADRKLLA  AKGS VSIGI  GQNSHIKRAI  IDKNARIGD   -399
54kd             YY  TK                                             -209
Sh-2       VMMGADIY  ETEEEASKLL  AGKV VPIGI  GRNTKIRNCI  IDMNARIGK   -274
         |. | |   .  |||   .|.   .:.:          .  .
R        NVKIINVDNV  QEAARETDGY  FIKSGIVTVI  KDALLLAEQL  YEVAA       -483
51KD     NIKIINSDNV  QEAARETDGY  FIKSGIVTVI  KDALIPSGTV  I           -457
54KD     NVIISNSEGV  QEAD   DGV  I                                  -227
Sh-2     NVVITNSKGI  QEADHPEEGY  YIRSESCDPE  ECNITVSSYR  SLRLRLQTDY  -324

Sh-2     MVLHR                                                      -329
```

Fig. 3. Alignment of the deduced amino acid sequences of rice seed (R), spinach leaf 51 kd subunit (51kd), maize endosperm Sh -2 subunit and the peptide fragments obtained from sequencing the spinach leaf 54 kd subunit (54kd) of ADPglucose pyrophosphorylase. Conserved amino acids in the alignment are indicated on top as : , the homologous amino acids are indicated as · and amino acids conserved for all 4 sequences are indicated as | . Numbers on the right indicate the amino acid residue from the translational start site.

Thus, very good identity is observed in comparing similar subunits of the ADPGlc pyrophosphorylase from the different plants and this is expected as the spinach leaf lower molecular weight subunit antibody reacts very well with the equivalent subunits of maize endosperm (ref. 25), rice seed (refs. 30 & 34), Arabidopsis (refs. 4 & 14) and potato tuber (ref. 28) enzymes. The lower molecular weight antibody does not react well with the higher molecular mass subunit of the ADPGlc pyrophosphorylase of these various plants. Therefore it was not expected to see much homology between the lower and higher molecular weight subunits. However, there appears to be some identity of about 40% between the 54 and 51 kd subunits of the spinach leaf ADPGlc pyrophosphorylase. The sequence analyses point out a greater relationship in identity between the spinach leaf 54 kd subunit and the maize Sh-2 subunit and most certainly between the wheat endosperm subunit encoded by the cDNA insert, AGA.7 (Fig 2,3). Most probably, the isolated cDNA clones of wheat endosperm are representative of the large molecular weight subunit of the wheat endosperm ADPGlc pyrophosphorylase. Because of the low but certain homology between the two subunits of the ADPGlc pyrophosphorylase it can be speculated that they may have arisen originally from the same gene. The bacterial ADPGlc pyrophosphorylase has been purified to homogeneity from 7 different and diverse species and in each case has been shown to be homotetramers composed of only one subunit (refs. 19 & 20). Recent results with cyanobacterial ADPGlc pyrophosphorylase, which has 3PGA as an allosteric activator and P_i as an inhibitor, similar to the higher plant enzyme (ref. 39), suggests that it is homotetrameric in structure unlike the higher plant enzymes (A.Iglesias, G. Kakefuda and J. Preiss, unpublished results). Thus it is quite possible that during evolution, there was gene duplication in the higher plant photosynthetic systems of the pyrophosphorylase gene and then divergence of the genes to produce two different peptides for the native ADPGlc pyrophosphorylase. Yet, both subunits are required for optimal activity of the native plant enzyme. The reason for this gene duplication is presently unknown but conservation of amino acid sequences may point out the particular importance for those sequences.

Comparison of the plant and bacterial ADPglucose pyrophosphorylase amino acid sequences

Fig. 4 compares the derived amino acid sequences from the rice seed and pyrophosphorylase cDNA clones corresponding to the small molecular mass, 51 kilodaton subunit with the deduced amino acid sequence of the E. coli glg C gene, the structural gene for ADPGlc pyrophosphorylase. There is an overall identity of about 29% at the amino acid level between the the plant and bacterial subunits suggesting a common origin for these two genes. Divergence appears to have occurred in a non-random manner for when the peptides are matched in residues 48 to 227 of rice seed to 19 to 196 of E. coli and rice seed residues 297 to 394 to E. coli amino acid residues 264 to 368, the degree of identity is higher, about 43%. Regions of largest divergence are at the N-terminal, amino acid residues

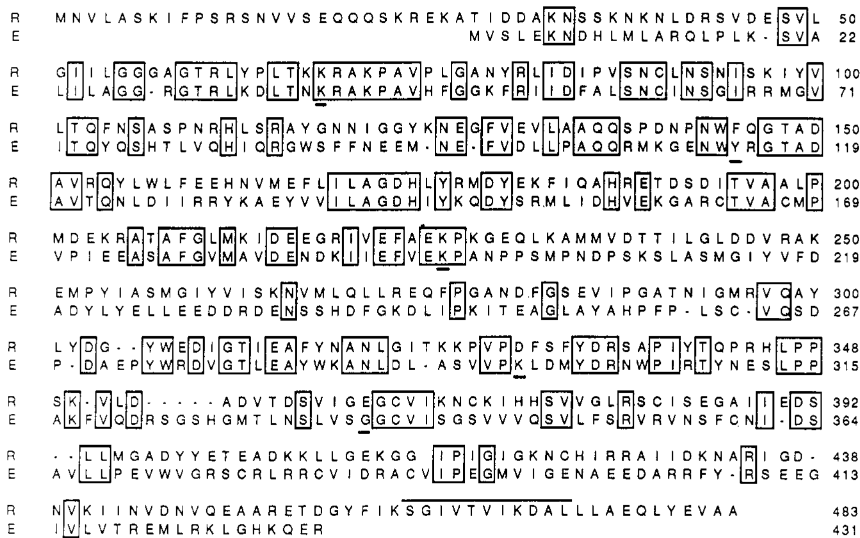

Fig. 4. Comparison of the primary sequences of the rice seed (R) and E. coli ADPglucose pyrophosphorylases.Conserved residues are noted in enclosed boxes and gaps in the sequences as indicated by dashes were done to obtain maximum homology. Numbers on the right indicate the amino acid residue from the translational start of either the rice or bacterial subunit. The underlined amino acids in the *E. coli* sequences are those residues indicated in the text to be essential for allosteric and/or catalytic function for the bacterial enzyme (K39, Y114, K195, K296 and G336). The allosteric site for the 51 kd subunit of the spinach leaf enzyme, SGIVTVIKDAL, is overlined.

29 to 47 of rice seed, a central portion between rice seed residues 228 through 296 and at the C-terminal beginning with Met[395]. Where significant identity occurs in region 48 to 227 for the rice seed is of high interest in that for the E. coli enzyme, equivalent area specific amino acids have been identified to be essential for substrate binding as well as for allosteric regulation (32). Lys 39 and Lys 195 of the *E. coli* primary sequence have been shown to be involved in the binding of the activator, fructose 1,6 bis-P, and the substrate, ADPglucose. Recent results (M.A. Hill and J. Preiss, unpublished) using site-directed mutagenesis at residue 195 and substituting the amino acids Gln, Glu or Leu indicate that the most affected kinetic parameter is the affinity for the substrate, glucose-1-P. Thus Lys 195 is involved in the binding of the β-phosphate of ADPGlc or of glucose 1-P. Six of the eight amino acids in that region are identical when E. coli and the rice seed sequences are compared. Indeed, Fig. 2 shows there is almost complete identity in the rice sequence, D[214]EEGRIVEFAEKP[226] with the equivalent sequences seen in the small subunits of the spinach leaf, potato tuber and Arabidopsis enzymes. For the 3 last mentioned enzymes the sequence is DEEGRIIEFAEKP. Similar conservation is seen for the wheat ADPGlc pyrophosphorylase subunits where the sequences, DSSGRVVQFSEQP, DSSGRVVQFSEKP or DDTGRVISFSEKP are present. The conservation of these sequences in the plant and bacterial enzymes suggests that it is important for binding of the substrate. There is also great identity at the activator site of the bacterial enzyme, Lys 39, where 9 of the 10 amino acids are the same in the equivalent rice seed sequence, residues 65 to 74. The potato tuber and spinach leaf subunits have the same identical sequences as the rice seed subunit. Fructose 1,6 bis-P is an activator of the various higher plant ADPGlc pyrophosphorylases (refs. 29,40,41). However, it is not as potent as an activator as 3PGA. The tyrosine residue 114 in E. coli has been shown to be involved in binding of the substrates and activator as changing it to Phe via site-directed mutagenesis caused a lowering of the apparent affinities for substrates and activator (ref. 42). Although there is a change to Phe for the rice seed, spinach leaf and potato tuber subunits in their equivalent sequences, 9 of the 12 amino acids in these sequences are identical with the E. coli sequence (Fig. 2). The conservation of amino acids in this region as well in the region equivalent to rice seed residues 65-74 in the plant enzymes suggest that these amino acids play a major role in maintenance of protein conformation and of the regulatory and catalytic functions of the plant enzyme having 3PGA as the major activator and P_i as the inhibitor. Whereas the activator for the E. coli enzyme is at the N-terminal, the activator site for spinach leaf enzyme is at the C-terminal (refs. 24 & 31). The reactive residue is Lys 430 in the 51 kd subunit of the spinach leaf enzyme and the equivalent lysine in the rice seed enzyme is residue 469 (Fig. 2). The sequence around the reactive lysine is observed in the derived amino acid sequences from the cDNA and genomic clones of either the high or low molecular mass subunits. Thus the extended portion of the plant peptide is not observed for the bacterial enzyme. The change in activator specificity could also be due to other amino acid replacements in other regions of the plant enzyme. Kumar et al. (42) have shown that mutations of the bacterial gene where Lys[296] and Gly[336] are substituted by Glu and Asp, respectively, yield an enzyme whose activity is less dependent on the presence of the activator, fructose 1,6 bis-P. It is of interest that in the amino acid sequences derived for the rice seed, spinach and Arabidopsis leaf and potato tuber small subunits, which directly align with Lys[296] and Gly[336] are the acidic amino acids, aspartate and glutamate, respectively (Fig. 4). Thus mutations to acidic amino acids at the rice seed subunit residues 329 and 363 as well as the extension of the C-terminal peptide bearing the activator binding site may result in the lessening of the activation by fructose 1,6 bis-P and creation of

the sensitivity to activation by 3PGA. This possibility can be tested via construction of a recombinant enzyme where the DNA fragment transcribing the higher plant C-terminus is fused to the to the mutant and wild type E. coli enzyme gene and analyzing the regulatory kinetics of the hybrid enzyme.

REFERENCES

1. C.Y. Tsai, and O.E. Nelson, Science 151, 341-343 (1966).
2. D.B. Dickinson and J. Preiss, Plant Physiol. 44, 1058-1062 (1969).
3. A.M. Smith, M. Bettey, and I.D. Bedford, Plant Physiol. 89, 1279-1284 (1989).
4. T.P. Lin, T. Caspar, C. Somerville J. Preiss, Plant Physiol., 86, 1131-1135 (1988).
5. J. Preiss and C. Levi, in (J. Preiss, ed.), The Biochemistry of Plants, Vol. 3. Academic Press, Inc., pp. 371-423 (1980).
6. J. Preiss, in (J. Preiss, ed.), The Biochemistry of Plants, Vol 14. Academic Press Inc., New York, pp. 181-254 (1988).
7. J. Preiss, (1990). in (B. Miflin, ed.) Oxford Survey of Plant Molecular and Cellular Biology, Vol. 7 In Press.
8. G. Pettersson and U. Ryde-Pettersson, Eur. J. Biochem., 179, 169-172 (1989).
9. H. Kacser and J.A. Burns, Symp. Soc. Exp. Biol., 27, 65-107 (1973).
10. H.Kacser, in (Davies, D.D., ed.), The Biochemistry of Plants, Vol. 11. Academic Press Inc., New York, pp. 39-67 (1987).
11. A.L. Kruckeberg, H.E. Neuhaus, R. Feil, L.D. Gottlieb and M. Stitt, Biochem. J., 261, 457-467 (1989).
12. H.E. Neuhaus, A.L. Kruckeberg, R. Feil, and M. Stitt, Planta, 178, 110-12 (1989).
13. T. Caspar, S.C. Huber, and C. Somerville, Plant Physiol., 79, 1-7 (1986).
14. T.P. Lin, T. Caspar, C. Somerville and J. Preiss, Plant Physiol., 88, 1175-1181 (1988).
15. H.E. Neuhaus, and M. Stitt, Planta, In Press (1990).
16. T. ap Rees, in (D.D. Davies, ed.) The Biochemistry of Plants, Vol. 12 Academic Press, New York. pp. 87-115 (1987).
17. M. Stitt, S. Huber, and P. Kerr, in (M.D. Hatch and N.K. Boardman, eds.) The Biochemistry of Plants, Vol. 10, Academic Press, New York. pp. 328-409 (1987).
18. J. Preiss, Ann. Rev. Plant Physiol., 54, 431-454 (1982).
19. J. Preiss, Ann. Rev. Microbiol., 38, 419-458 (1984).
20. J. Preiss and T. Romeo, in (A.H. Rose and D.W. Tempest, eds.), Advances in Microbial Physiology, Vol. 30, pp. 183-238 (1989).
21. J. Preiss, M. Bloom, M. Morell, V. Knowles, W.C. Plaxton, T.W. Okita, R. Larsen, A.C. Harmon and C. Putnam-Evans, in (G. Bruening, J. Harada, T. Kosuge, and A. Hollaender, eds.), Tailoring Genes for Crop Improvement. An Agricultural Perspective. Plenum Press, New York. pp. 133-152 (1987).
22. J. Preiss, M. Morell, M. Bloom, V.L. Knowles, and T.P. Lin, in (J. Biggens, ed.), Progress in Photosynthesis Research,Vol. III. Nijhoff, The Hague, pp. 693-700 (1987).
23. M.K. Morell, M. Bloom, V. Knowles, and J. Preiss, Plant Physiol., 85, 185-187 (1987).
24. M. Morell, M. Bloom, R. Larsen, T.W. Okita, and J. Preiss, in (J.L. Key, and L. McIntosh, eds.), Plant Gene Systems and their Biology., Alan R. Liss, Inc. New York, pp. 227-242 (1987).
25. J. Preiss, S. Danner, P.S. Summers, M. Morell, C.R. Barton, L. Yang, and M. Nieder, Plant Physiol., 92, 881-885 (1990).
26. C. Barton, L. Yang, M. Galvin, C. Sengupta-Gopalan, and T. Borelli, in (J.C. Shannon, D.P. Knievel, and C.D. Boyer, eds.), Regulation of Carbon and Nitrogen Reduction and Utilization in Maize, American Society of Plant Physiologists, Rockville MD, pp 363-365 (1986).
27. J.M. Anderson, T.W. Okita, and J. Preiss, J. in W.C. Park, and M.E. Vayda, (eds.), Molecular and Cellular Biology of the Potato., A.B. International, Wallinford, UK pp. 159-180 (1990).
28. T.W. Okita, P.A. Nakata, J.M. Anderson, J. Sowokinos, M. Morell, and J. Preiss, Plant Physiol., 91, 221-227 (1990).
29. J.R. Sowokinos and J. Preiss, (1982). Plant Physiol., 69, 1459-1466 (1982).
30. H.B. Krishnan, C.D. Reeves, and T.W. Okita, Plant Physiol., 81, 642-645 (1986).
31. M. Morell, M. Bloom, and J. Preiss, J. Biol. Chem., 263, 633-637 (1988).
32. T.F. Parsons, and J. Preiss, J. Biol. Chem., 253, 7638-7645 (1978).
33. J. Preiss, D. Cress, J. Hutny, M. Morell, M. Bloom, T. Okita, and J. Anderson, in J.R. Whitaker and P.E. Sonnet, (eds.), Biocatalysis in Agricultural Biotechnology. Amer. Chem. Soc. Washington, D.C. pp. 84-92 (1989).
34. J.M. Anderson, J. Hnilo, R. Larson, T.W. Okita, M. Morell, and J. Preiss, J. Biol. Chem., 264, 12238-12242 (1989).
35. N.J. Proudfoot, and G.G. Brownlee, Nature, 263, 211-214 (1976).
36. G.A.Karlin-Neumann, and E.M. Tobin, EMBO J., 5, 9-13 (1986).
37. J.M. Anderson, R. Larsen, D. Laudencia, W.T. Kim, D. Morrow, T.W. Okita, and J. Preiss, Molec. Gen. genetics, In Press (1990).
38. M.R. Olive, R.J. Ellis and W.W. Schuch, Plant Mol. Biol. 12: 525-538 (1989).
39. C. Levi, and J. Preiss, Plant Physiol., 58, 753-756 (1976).
40. H.P. Ghosh and J. Preiss, J. Biol. Chem., 241, 4491-4505 (1966).
41. G.G. Sanwal, E. Greenberg, J. Hardie, E. Cameron and J. Preiss, Plant Physiol., 43, 417-427 (1968).
42. A. Kumar, T. Tanaka, Y.M. Lee, and J. Preiss, J. Biol. Chem., 263, 14634-14639 (1988).