SYNTHESIS OF THE GLUCOSYL DONOR: ADPglucose PYROPHOSPHORYLASE

I. REGULATORY PROPERTIES

In the biosynthesis of starch and bacterial glycogen, the glucose donor, ADPglucose, is formed from ATP and glucose-1-P via a reaction catalyzed by ADPglucose pyrophosphorylase (ADPGlc PPase; glucose-1-P adenylyltransferase; E.C. 2.7.7.27). This reaction was first described by Espada (1962) in soybean and was subsequently found in many plant tissues and in bacterial extracts. ADPGlc PPases have been isolated from many plants and bacteria, and their regulatory properties have been studied. Although the major activators vary according to the source, they share a common characteristic: The activator specificity of the enzyme is determined by the major pathway of carbon assimilation in the organism. The relationship of activator specificity of the ADPGlc PPase of the various organisms with the pathways is summarized in Table I.

The reaction catalyzed by ADPGlc PPase is reversible, and it should be noted that regulatory properties can be different in the two directions. For example, pyrophosphorolysis is usually much less affected by allosteric activators than is the synthesis of the sugar nucleotide (e.g., see Ghosh and Preiss, 1966; Preiss *et al.*, 1967).

Enteric bacteria, such as *Escherichia coli,* assimilate glucose via glycolysis and regulation of the glycolytic pathway is at the site of fructose-1,6-bis-P synthesis (the phosphofructokinase step), and this is the major activator for the *E. coli* ADPGlc PPase (Preiss, 1984; Preiss and Romeo, 1989, 1994).

For organisms where the predominant pathway is the Entner–Doudoroff pathway, fructose-1,6-bis-P is not a major metabolite in glucose degradation (because glucose-6-P is converted first into 6-P-gluconate and then to 2-keto,3-deoxy,6-P-gluconate); the activators for their ADPGlc PPase are fructose-6-P and pyruvate (Preiss, 1969, 1984; Preiss and Romeo, 1989).

Rhodospirillum rubrum cannot metabolize glucose but grows anaerobically on pyruvate, lactate, or on CO_2. Pyruvate has been shown to be a product of CO_2 fixation, and it is also the sole activator of the *R. rubrum* ADPGlc PPase (Furlong and Preiss, 1969).

TABLE I

ACTIVATOR SPECIFICITIES OF ADPGLUCOSE PYROPHOSPHORYLASES (ADPGlc PPASE) FROM DIFFERENT ORGANISMS[a]

Organisms	Activator specificity	Assimilation pathway
Enterobacteria	Fructose-1,6-bis-P	Glycolysis
Agrobacterium tumefaciens	Fructose-6-P, pyruvate	Entner–Doudoroff pathway
Rhodopseudobacter spheroides	Fructose-1,6-bis-P, pyruvate, fructose-6-P	Glycolysis, Entner–Doudoroff, anaerobic photosynthesis
Rhodospirillum rubrum	Pyruvate	Anaerobic photosynthesis
Cyanobacteria, green algae, higher plants	3-P-Glycerate	Oxygenic photosynthesis

[a]Grouped according to carbon assimilation pathway.

Rhodobacter spheroides, a highly adaptable organism, can metabolize glucose by glycolysis or, under other physiologic conditions, by the Entner–Doudoroff pathway, and it can also assimilate CO_2 during anaerobic photosynthesis. It has an ADPGlc PPase that is effectively activated either by fructose-1,6-bis-P, fructose-6-P, or pyruvate (Greenberg *et al.,* 1983) (i.e., its adaptability in carbon assimilation is associated with an ADPGlc PPase with flexible activation specificity).

Cyanobacteria, green algae, and higher plants assimilate CO_2 during photosynthesis to form 3-P-glycerate (3PGA). By 1982, ADPGlc PPases from several plant species—13 from leaf and 9 from nonphotosynthetic tissues—had been shown to be activated by 3PGA (Preiss, 1982b), which in most cases increases the affinity for the substrates, ATP and glucose-1-P, and reverses the inhibition caused by P_i. Since 1982, ADPGlc PPases from other nonphotosynthetic tissues have been studied (e.g., maize endosperm, potato tuber, cassava root, rice endosperm), and these tissues were highly dependent on the presence of 3PGA and were inhibited by P_i. Some exceptions to this rule have been reported. In the ADPGlc PPases from pea embryos (Hylton and Smith, 1992), barley endosperm (Kleczkowski *et al.,* 1993), and bean cotyledon (Weber *et al.,* 1995), activation by 3PGA is not as high, ranging between 1.5- and 3-fold. However, ADPGlc PPases are usually much less affected by allosteric activators in the pyrophosphorolysis direction than in the synthesis direction (Ghosh and Preiss, 1966; Preiss *et al.,* 1967). Activation for the "anomalous" enzymes would likely be higher if assayed in the synthesis direction, which is, after all, the direction in which the glucose donor is formed.

In the first studies of maize endosperm ADPGlc PPase, it was thought that the enzyme was insensitive to 3PGA activation and P_i inhibition (Dick-

inson and Preiss, 1969a,b). It was found later, however, that if protease inhibitors were added to the maize endosperm extracts, activity was then very sensitive to activation by 3PGA and to inhibition by P_i (Plaxton and Preiss, 1987). It was also shown that if the activity of proteases was not prevented, the size of the 54-kDa subunit was reduced to 53 kDa, a small but reproducible change in size. Thus, partial proteolysis during enzyme isolation can strongly affect ADPGlc PPase regulatory properties, and proteolysis may be one reason behind the allosteric insensitivity found in the atypical ADPGlc PPases.

Figure 1 illustrates how relatively small changes in the 3PGA and P_i concentrations can greatly affect the rate of ADPglucose synthesis, particularly at low concentrations of 3PGA, where the activation is minimal, and in the presence of P_i. At 1.2 m*M* P_i and 0.2 m*M* 3PGA, ADPglucose synthesis is inhibited by more than 95%. However, if the P_i concentration decreases 33% to 0.8 m*M*, and the 3PGA concentration increases 50% to 0.3 m*M*, there is an 8.5-fold increase in the rate of ADPglucose synthesis. Conversely, at 0.4 m*M* 3PGA and 0.8 m*M* P_i, the rate of ADPglucose

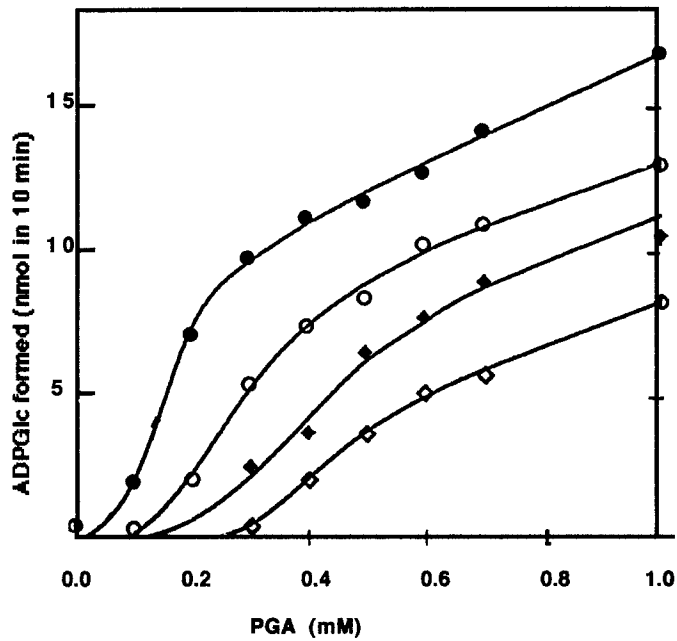


FIG. 1. Effects of P_i and 3PGA on rate of ADPGlucose synthesis catalyzed by potato tuber ADPGlc PPase. ●, 3PGA curve done in the presence of 0.4 m*M* P_i; ○, 0.8 m*M* P_i; □, 1.2 m*M* P_i; ■, 3PGA curve measured in the presence of 1.6 m*M* P_i.

synthesis is 7.5 nmol per 10 minutes. This is reduced to 2.2 nmol (70% decrease) if the 3PGA concentration decreases 50%, to 0.2 m*M*. If the P_i concentration increases to 1.2 m*M*, the synthetic rate is then reduced to 0.65 nmol, which is a reduction in ADPglucose synthesis of 91%. The reason that small changes in the effector concentrations produce such large effects in the synthetic rate is due to the sigmoidal nature of the curves particularly at the low concentrations of 3PGA.

II. PHYSIOLOGIC RELEVANCE OF THE ADPGlc PPase REGULATORY PROPERTIES

In vivo and in situ experiments strongly indicate that the activation by 3PGA and inhibition by P_i observed in vitro are also physiologically important. Many experiments have been cited in reviews (Preiss and Levi, 1980; Preiss, 1982a,b, 1988, 1991, 1996; Sivak and Preiss, 1995; Preiss and Sivak, 1996) showing a direct correlation between the concentration of 3PGA and starch accumulation, and an inverse one between P_i concentration and starch content. This is true for photosynthetic tissues, in which P_i and PGA concentrations within the chloroplast are good indicators of the energy and carbon status, and in this way the ADPGlc PPase provides a good regulatory mechanism for the flux of photosynthate into starch. It has been found that the regulatory properties of the enzyme of nonphotosynthetic tissue, such as potato tuber and maize endosperm, are such that the ADPGlc PPase is almost completely dependent on the presence of the activator, but in these tissues it is still uncertain how 3PGA and P_i can signal the availability of carbon and energy for starch synthesis, since transport of carbon in the amyloplast is via hexose-phosphates rather than by triose-phosphates as seen in chloroplasts (Keeling *et al.*, 1988; Heldt *et al.*, 1991; Hill and Smith, 1991; Viola *et al.*, 1991).

If this activation mechanism is indeed important physiologically, its failure should have important consequences in vivo. This has been confirmed by chemical mutagenesis in bacteria (Preiss, 1969, 1984, 1996), *Arabidopsis thaliana* (Lin *et al.*, 1988a,b), and in the green algae *Chlamydomonas reinhardtii* (Ball *et al.*, 1991). More recently, an allosterically altered ADPGlc PPase has been reported in maize endosperm (Giroux *et al.*, 1996). In the *Chlamydomonas* system, starch-deficient mutants have been isolated and characterized, and have been shown defective in the ADPGlc PPase, which could not be effectively activated by 3PGA. The maize endosperm ADPGlc PPase allosteric mutant is less sensitive to P_i inhibition than the normal enzyme and the mutant endosperm has 15% more dry weight than the normal endosperm (Giroux *et al.*, 1996). The *Chlamydomonas* starch-

deficient and higher dry-weight maize endosperm mutants ADPGlc PPases strongly suggest that the in vitro regulatory effects observed with the photosynthetic and nonphotosynthetic plant ADPGlc PPases are highly functional in vivo, and that ADPGlc synthesis is rate limiting for starch synthesis.

III. SUBUNIT STRUCTURE

To study subunit structure, it is essential to determine the molecular mass of the holoenzyme by gel filtration and/or sucrose density gradient followed by determination of enzymatic activity. The size of the subunits can be determined by sodium dodecyl phosphate–polyacrylamide gel electrophoresis (SDS–PAGE). Put together, this information will show whether the enzyme is a monomer or a polymer and, if the latter, how many subunits make up the holoenzyme and whether there is only one kind of subunit or more than one kind.

Many bacterial ADPGlc PPases have been purified and in many their subunit structure has been determined. Invariably the native enzymes are tetrameric with only one kind of subunit, with a molecular mass ranging from 49,000 to 54,000, according to the species.

In contrast, the plant enzyme consists of two related but different subunits with masses in the 50,000 to 60,000 range. The "small" subunits have molecular masses of about 50,000 to 54,000, whereas the other, "large" subunits have molecular masses of 51,000 to 60,000. Although the difference in mass between the two subunits in one enzyme can be small, it is still convenient to designate them as small and large; they differ in many other characteristics, and this is discussed as follows.

The potato tuber, spinach leaf, and maize endosperm enzymes have small subunit masses of 50,000, 51,000, and 54,000, respectively, and large subunit masses of 51,000, 54,000, and 60,000, respectively. The small and large subunits have about 50 to 60% identity with each other and have about 30 to 40% identity with the procaryotic ADPGlc PPases.

An ADPGlc PPase that is well studied with respect to structural properties is the spinach leaf enzyme (Morell *et al.,* 1987, 1988; Ball and Preiss, 1994). This enzyme has a molecular mass of 206,000 and is composed of two different subunits, with molecular masses 51,000 and 54,000. These subunits, which can be separated by chromatography after denaturating the holoenzyme with urea, can be distinguished not only by their molecular masses but also with respect to amino acid composition, amino-terminal sequences, peptide patterns on high-performance liquid chromatography (HPLC) of their tryptic digests, and antigenic properties. The polyclonal antibody prepared against the 51-kDa subunit reacted very strongly, in

immunoblots, with the 51,000 subunit, but weakly with the 54,000 subunit. Conversely, antibodies raised against the large subunit reacted only weakly with the small subunit and strongly with the large. Thus, on the basis of the protein chemistry and immunologic analyses, the two subunits are distinct and probably are the products of two genes.

Preiss *et al.* (1990) showed that the maize endosperm ADPGlc PPase, which has a molecular mass of 230,000, could react with the antibody prepared against the native spinach leaf enzyme in immunoblot experiments. In SDS gel electrophoresis of endosperm extracts or of the highly purified enzyme, two polypeptides of 55 and 60 kDa reacted with the antiserum raised against the spinach holoenzyme. The results were different when antibodies raised against the separate subunits (large or small) were used. The antibody prepared against the spinach leaf large subunit cross-reacted mainly with the endosperm large subunit and to a small extent with the 55-kDa subunit. The antibody against the spinach leaf small subunit antibody cross-reacted well with the endosperm 55-kDa subunit and weakly with the 60-kDa subunit. The maize endosperm starch-deficient mutants, *shrunken* 2 (*sh* 2) and *brittle* 2 (*bt* 2), were also studied. In immunoblotting experiments and while using antibodies against the native or subunit antibodies of the spinach leaf enzyme, the mutant *bt* 2 endosperm lacked the 55-kDa subunit and the mutant *sh* 2 endosperm lacks the 60-kDa subunit. These results indicate that the maize endosperm ADPGlc PPase is composed of two immunologically distinctive subunits, and that the *sh* 2 and *bt* 2 mutations cause reduction in ADPGlc PPase activity (and the consequent deficiency in starch content) through the lack of one of the subunits. Thus, the *sh* 2 gene would be the structural gene for the 60-kDa, large subunit, whereas the *bt* 2 gene would be the structural gene for the 55-kDa, small subunit.

An ADPGlc PPase cDNA clone, isolated from a maize endosperm library (Barton *et al.*, 1986), hybridized with the small subunit cDNA clone from rice (Anderson *et al.*, 1989). This maize ADPGlc PPase cDNA clone hybridizes to a transcript that is present in maize endosperm but absent in *bt* 2 endosperm. Thus, the *bt* 2 mutant appears to be the structural gene of the 55-kDa subunit of the ADPGlc PPase. These data also indicate that the nonphotosynthetic tissue ADPGlc PPase is also composed of two subunits and, on the basis of immunoreactivity, there is homology between the large and small subunits in the leaf enzyme with the subunits of a reserve tissue enzyme, respectively.

The potato tuber ADPGlc PPase has been highly purified and, by two-dimensional polyacrylamide gel electrophoresis, two polypeptides could be distinguished by their slight differences in molecular mass, 50,000 and 51,000, and in net charge (Okita *et al.*, 1990). The tuber small subunit is

reactive with the antibody prepared against the spinach leaf small subunit. The antiserum prepared against the spinch leaf large subunit, however, does not react with either potato tuber enzyme subunit. The potato tuber enzyme is composed of two distinct subunits and is not a homomer as initially thought (Sowokinos and Preiss, 1982).

The ADPGlc PPase of *A. thaliana* is composed of two subunits, with molecular masses of 51,000 and 54,000. One *A. thaliana* mutant, TL25, lacks both subunits of the ADPGlc PPase (it is thought that the mutation affects a regulatory locus), whereas another mutant, TL46, lacks the large, 54-kDa subunit only. The TL46 mutation provides further evidence that the larger subunit is a necessary component of the native ADPGlc PPase for optimal activity since the mutant has only 7% of the wild-type activity. The mutant synthesizes starch at 9% of the rate displayed by the wild type in high light, and at 26% of the wild-type rate measured at low light (Neuhaus and Stitt, 1990).

IV. STRUCTURE–FUNCTION RELATIONSHIPS

The researcher who wants to elucidate the mechanism of action and the regulation of an enzyme has many methodological tools at his disposal, and more become available every year.

Chemical modification can supply information on the amino acids involved in the active and regulatory sites. The amino acid sequences obtained by Edman degradation of the proteins purified from different tissues and species, and/or by cloning followed by deduction of amino acid sequences, can be compared. This exercise will point out the amino acid sequences well conserved in enzymes from different sources, which are likely to be essential for enzyme function. Using site-directed mutagenesis, the amino acids deemed to be crucial are replaced by others, and the effect of these changes on the properties of the enzyme are studied. To achieve this objective, *E. coli* is transformed with the mutated gene in a suitable vector, the overexpressed enzyme is purified, and its properties are compared with those of the enzyme obtained from bacteria transformed with the nonmutated gene.

Chemical mutagenesis, followed by screening for starch with iodine reagent, can help identify amino acids crucial for binding or catalysis in an approach similar to that used for the ADPGlc PPase of *E. coli.* In plants, chemical mutagenesis has been used with *A. thaliana,* (Lin *et al.,* 1988a,b), with *C. reinhardtii* (Ball *et al.,* 1991), and with the potato enzyme expressed in *E. coli* (Greene *et al.,* 1996). As for any methodology intending to identify a crucial amino acid, the effect of the mutation in a single amino acid must

be specific for a particular substrate or modulator. A generalized effect indicates that the amino acid in question affects the general conformation of the enzyme.

V. FUNCTION OF THE HIGHER PLANT ADPGlc PPase SUBUNITS

After discovering that the plant native ADPGlc PPases were tetrameric and composed of two different subunits, the next step was to determine why the two subunits were required for optimal catalytic activity. Since the enzyme must contain ligand binding sites for the activator (3PGA), inhibitor (P_i), sites for the two substrates (ATP and glucose-1-P), as well as a catalytic site, it is possible that these sites could be located on different subunits.

Two cDNAs encoding the mature large subunit and small subunits of the potato tuber (*Solanum tuberosum* L.) ADPGlc PPase have been expressed in *E. coli* (Iglesias *et al.*, 1993; Ballicora *et al.*, 1995). The large subunit and small subunits could be expressed separately as well as together. As seen in Table II, considerable activity of ADPGlc PPase is obtained when the cDNA of the large subunit is expressed along with the cDNA of the small subunit enzyme in an *E. coli* mutant devoid of ADPGlc PPase activity. The purified recombinant enzyme, containing both the large and small subunits, has a specific activity of 64 μmol · min^{-1} · mg^{-1} when measured in the presence of the activator (3 m*M* 3PGA). If the large subunit is expressed alone, little activity is observed. However, expression of the small subunit alone leads to significant ADPGlc PPase activity (Ballicora *et al.*, 1995). This homomeric (four small subunits) enzyme has been puri-

TABLE II

COMPARISON OF THE PROPERTIES OF TRANSGENIC ADPGlc PPASES WITH THE PROPERTIES OF THE POTATO TUBER ENZYME[a]

Enzyme source	$A_{0.5}$ (m*M*)	$I_{0.5}$ (m*M*) at 0.25 m*M*, 3PGA	$I_{0.5}$ (m*M*) at 3.0 m*M*, 3PGA
Potato tuber (Sowokinos and Preiss, 1982)	0.40	0.12	0.33
pMLaugh10 + pMON17336 (large and small subunits)	0.16	0.07	0.63
pMLaugh10 (small subunit only)	2.40	—	0.08

[a] The kinetic constants of the recombinant enzyme purified from *E. coli* were measured (Ballicora *et al.*, 1995) and they coincided with the data obtained with the native potato tuber enzyme (Sowokinos and Preiss, 1982). $A_{0.5}$ and $I_{0.5}$ are concentration of activator PGA needed for 50% of maximal activation and concentration of inhibitor P_i giving 50% inhibition, respectively.

fied almost to homogeneity with a specific activity of 50 μmol · min^{-1} · mg^{-1} when measured in the presence of a high concentration (4 m*M*) of 3PGA. As shown in Table II, the enzyme composed exclusively of small subunits has a lower apparent affinity ($A_{0.5}$ = 2.4 m*M*) for the activator, 3PGA, than the heterotetramer. The enzyme with only the small subunit is also more sensitive to P_i inhibition ($I_{0.5}$ of 0.08 m*M* in the presence of 3 m*M* 3PGA) as compared with the heteromeric enzyme ($I_{0.5}$ value of 0.63 m*M*). The K_m values for the substrates and Mg^{+2} are essentially the same whether the enzyme is composed of only one subunit, the small subunit, or two subunits, small and large. In every case the native enzyme is a tetramer—a homotetramer in the case of the small subunit alone and a heterotetramer in the case of the large and small subunits (Ballicora *et al.*, 1995).

These data suggest that the small subunit is primarily involved in catalysis; it has substantial activity in the absence of the large subunit if the concentration of 3PGA, the activator, is high. The large subunit, when expressed alone, has little activity, but if expressed with the small subunit, the resulting enzyme has similar regulatory kinetic constants as does the native potato enzyme. This suggests that the prime function of the large subunit would be to regulate the activity of the small subunit, increasing the apparent affinity for the activator, and decreasing the affinity for the inhibitor P_i. This information agrees with results obtained with *A. thaliana,* in which the mutant ADPGlc PPase lacking the large subunit had activity but its affinity for the activator, 3PGA, was lower and the affinity for P_i was higher than for the wild-type heterotetrameric enzyme (Li and Preiss, 1992).

The small subunit of the higher plant ADPGlc PPases is highly conserved (85–95% identity), whereas the large subunit is less conserved (50–60% identity; Smith-White and Preiss, 1992). The higher heterogeneity seen in the large subunit sequence probably reflects different demands in the modulation of the small subunit sensitivity to allosteric activation and inhibition posed by different demands of the tissue and species. Expression of large subunits would differ during development or in different plants and tissues (e.g., leaf, stem, guard cells, tuber, endosperm, root, embryo), providing the resulting ADPGlc PPases with differing sensitivities to regulators.

VI. IDENTIFICATION OF THE SUBSTRATE BINDING SITES

Chemical modification can be used to obtain information on the catalytic mechanism and on the catalytic site of the enzyme of interest. One goal in the design of affinity labels for enzymes is to determine the catalytically important residues. First, the affinity label has to behave as an analogue of the substrate (or of the activator or inhibitor) by competition experiments.

Second, the enzyme is covalently bound to the affinity label in conditions chosen according to the enzyme in question and the chemical nature of the analogue, so as to decrease nonspecific labeling. Third, the labeled enzyme is subjected to proteolysis and the radioactive peptide(s) are isolated by HPLC. The labeled peptide(s) are then sequenced, providing information about the domains of the enzyme involved in the interaction with the substrate (or with the modulators).

Chemical modification studies on ADPGlc PPase have involved the use of the following affinity labels:

1. Pyridoxal-5-phosphate (PLP), an analog of 3PGA or phosphorylated sugars that can be covalently bound to the enzyme by reduction with $NaBH_4$
2. The photoaffinity substrate analogs, 8-azido-ATP and 8-azido-ADPglucose. When ultraviolet (UV) light (257 nm) irradiates 8-azido compounds, a nitrene radical is formed, which can react with electron-rich residues and inactivate the enzyme.
3. Phenylglyoxal, for the identification of arginine residues

These studies have provided information on the catalytic and regulatory sites of the spinach and cyanobacterial ADPGlc PPases, and on the role of the large and small subunits (Morell *et al.*, 1988; Smith-White and Preiss, 1992; Ball and Preiss, 1994; Charng *et al.*, 1994).

In addition, residues that chemical modification suggested were involved in substrate binding have been subjected to site-directed mutagenesis (Kumar *et al.*, 1989; Hill *et al.*, 1991; Charng *et al.*, 1994, 1995; Sheng *et al.*, 1996). These studies have provided information on the catalytic and regulatory sites of the spinach ADPGlc PPase and on the role of the large and small subunits. They have also shown that many of the studies initiated with the bacterial ADPGlc PPases are highly relevant for studies on the higher plant enzyme (Kumar *et al.*, 1988; Hill *et al.*, 1991; Charng *et al.*, 1994; Sheng *et al.*, 1996).

In the ADPGlc PPase from *E. coli*, the Lys residue 195 has been identified as the binding site for the phosphate of glucose-1-P (Hill *et al.*, 1991), and tyrosine (Tyr) residue 114 has been identified as involved in the binding of the adenosine portion of the other substrate, ATP (Lee and Preiss, 1986). When the amino acid sequence of the *E. coli* enzyme is aligned with those from the plant and cyanobacterial ADPGlc PPases, the identity ranges from 30 to 33% (Smith-White and Preiss, 1992). Sequence identity is much higher when only the ATP and glucose-1-P binding sites (Table III) are compared with the corresponding sequences of the plant and cyanobacterial enzymes, suggesting that those sequences are still important in the plant enzyme, probably having the same function.

TABLE III

CONSERVATION OF THE SEQUENCE OF THE *E. coli* ADPGlc PPase BINDING SITES FOR Glc-1-P[a] AND ATP[b] IN THE ENZYMES FROM OTHER ORGANISMS[c]

Organism	Glc-1-P site	ATP site
Prokaryotes		
E. coli	IIEFVEKP-AN	WYRGTADAV
S. typhimurium	**D*****-**	*********
Anabaena	V*D*S***KGE	*FQ******
Synechocystis	*TD*S***QGE	*FQ******
Plant small subunit		
Spinach leaf, 51 kDa	****A***KGE	*FQ******
Potato tuber, 50 kDa	****A***QGE	*FQ******
Maize endosperm, 54 kDa	****A***KGE	*FQ******
Rice seed	*V**A***KGE	*FQ******
A thaliana	****A***KGE	*FQ******
Wheat endosperm	****A***KGE	*FQ******
Plant large subunit		
Spinach leaf, 54 kDa	VLS*S***KGD	*FQ**•****
Potato tuber, 51 kDa	VVQ*A***KGF	*FQ******
Maize endosperm, 60 kDa	VLQ*F***KGA	*FQ****SI
A. thaliana	V*SFS***KGD	*FQ*****L
Wheat endosperm	VVQ*S*Q*KGD	*FR*****W

[a] Data from Hill *et al.* (1991).

[b] Data from Kumar *et al.* (1988).

[c] For references to sequences, see Smith-White and Preiss (1992) for the plant enzymes; Charng *et al.* (1992) for *Anabaena;* Kakefuda *et al.* (1992) for *Synechocystis;* and Ainsworth *et al.* (1993) for the wheat endosperm small subunit. Lys-195 and Tyr-114 of the *E. coli* enzyme belong to the Glc-1-P and ATP binding sites, respectively. * signifies the same amino acid as in the *E. coli* enzyme.

The binding site for pyridoxal phosphate in the small subunit was isolated, revealing a lysine (Lys) residue close to the C terminus, which may be important for 3PGA activation (Morell *et al.,* 1988). When PLP is covalently bound (Fig. 2), the plant ADPGlc PPase no longer requires 3PGA for activation; and the binding of PLP is prevented by the allosteric effectors, 3PGA and P_i. These observations indicate that the activator analog, PLP, is binding at the activator site. In addition, Preiss *et al.* (1992) and Ball and Preiss (1994) showed that three Lys residues of the spinach leaf large subunit are also involved or are close to the binding site of pyridoxal-P and, presumably, to the activator, 3PGA (Table IV). The chemical modification of these Lys residues by pyridoxal-P was prevented by the presence

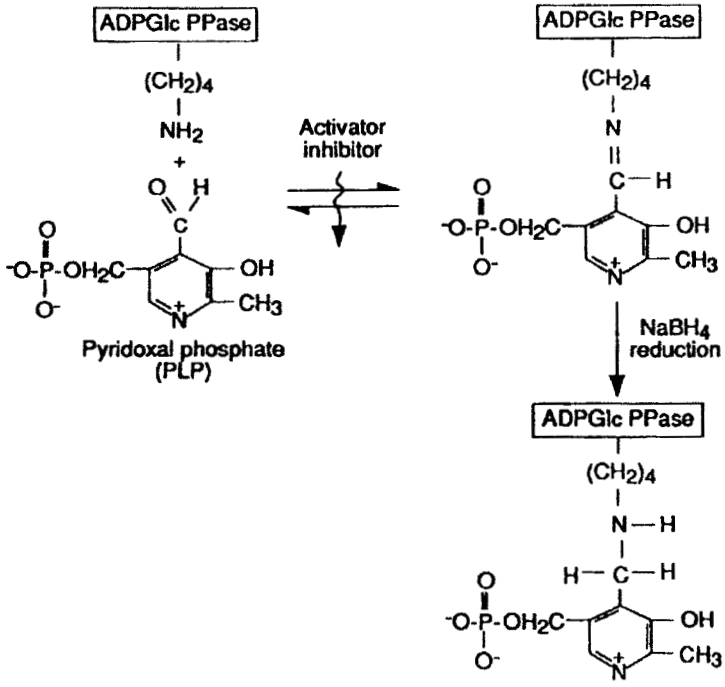


FIG. 2. Chemical modification is one of the tools used to identify the amino acid residues involved in the binding of a substrate, activator, or inhibitor. In the case of the ADPglucose pyrophosphorylase (ADPGlc PPase), the allosteric sites can be modified using pyridoxal-5-phosphate (PLP). PLP forms a Schiff base with an ε-amino group of a Lys residue. This Schiff base is converted to a stable secondary amine by reduction with $NaBH_4$. The modified enzyme no longer requires activator for catalysis, indicating that a Lys residue participates in the binding of the activator. This evidence is supported by the fact that modification of the enzyme with PLP can be prevented if an allosteric effector (i.e., 3PGA or P_i) is present when the enzyme is incubated with PLP.

of 3PGA during the reductive pyridoxylation process and, in the case of the Lys residue of site 1 of the small subunit and site 2 of the large subunit, P_i also prevented them from being modified by reductive pyridoxylation. Thus, it is believed that the most important sites involved are sites 1 and 2. Similar results were obtained with the *Anabaena* ADPGlc PPase (Charng *et al.*, 1994). Chemical modification of the enzyme with PLP caused the cyanobacterial enzyme no longer to require activator for maximal activity; chemical modification was prevented by 3PGA and P_i. The modified Lys residue was identified as Lys-419 and the sequence adjacent to that residue is similar to that observed for site 1 sequences in the higher plants. Site-directed mutagenesis of Lys-419 to either Arginine (Arg), Alanine (Ala), Glutamine (Gln), or glutamic acid (Glu) produced mutant enzymes (ex-

TABLE IV

PLANT AND CYANOBACTERIAL ADPGlc PPase ACTIVATOR BINDING SITES[a]

	Activator site 1	Activator site 2
Potato tuber, 50 kDa	SGIVTVIKDALIPSGIII	IKRAIIDKNAR
Spinach, 51 kDa (small)	SGIVTVIKDALIPSGTVI	IKRAIIDKNAR
Maize, 54 kDa	GGIVTVIKDALLPSGTVI	IRRAIIDKNAR
Wheat seed (small)	SGIVTVIKDALLPSGTVI	IKRAIIDKNAR
Anabaena	SGIVVVLKNAVITDGTII	QRRAIIDKNAR
Synechocystis	NGIVVVIKNVTIADGTVI	IRRAIIDKNAR
Spinach, 54 kDa (large)	SGITVIFKQATIKDGVV	IKDAIIDKNAR
Potato, 51 kDa (large)	SGIIIILEKATIRDGTVI	IRKCIIDKNAK
Maize, 60 kDa (large)	SGIVVILKNATINECLVI	IRNCIIDMNAR
Wheat seed (large)	SGIVVIQKNATIKDGTVV	IQNCIIDKNAR
Barley endosperm (large)	SGIVVIQKNATIKDGTVV	ISNCIIDMNAR

[a] The sequences listed in one-letter code are from Smith-White and Preiss (1992). The sequences of the barley endosperm enzyme are from Villand *et al.* (1992). The Lys residues underlined indicate they are covalently modified by pyridoxal-P and the chemical modification of the Lys residue is prevented by 3PGA and P_i, or site-directed mutagenesis has identified them to be involved in binding the activator. The numbers 441 and 417 correspond to the Lys residues in the potato tuber ADPGlc PPase small subunit. Site 1 is present both in the large and in the small subunits of the plant ADPGlc PPase, whereas site 2 is only in the large subunit even though similar sites are observed in the small subunit.

pressed in *E. coli*) with lowered affinities, 25- to 150-fold lower than that of the wild-type enzyme. No other kinetic constants, such as affinity for substrates and the inhibitor, P_i, were affected, nor was the heat stability or the catalytic efficiency of the enzyme affected. These mutant enzymes, however, were still activated to a great extent at higher concentrations of 3PGA, suggesting that an additional site was involved in the binding of the activator. The Lys-419 in the Arg mutant was chemically modified with the activator analog, PLP, and Lys 382 was the amino acid that was reductively phosphopyridoxylated. Modification of Lys-382 in the Arg mutant also caused a dramatic alteration in the allosteric properties of the enzyme, which could be prevented by the presence of 3PGA or P_i during the chemical modification process. Therefore, Lys-382 was identified as the additional site involved in the binding of the activator and, as seen in Table IV, the adjacent sequence about Lys-382 in the *Anabaena* enzyme is similar to that seen for site 2.

In the ADPGlc PPases of *Anabaena* and higher plants, there are five highly conserved Arg residues that are not present in the enteric bacterial ADPGlc PPases. As discussed previously, the regulatory characteristic of enteric bacteria are different from those of cyanobacteria and higher plants;

for example, the enteric ADPGlc PPases are not inhibited by P_i, but by 5-AMP. Phenylglyoxal inactivation of the spinach enzyme can be prevented by 3PGA or by P_i, which is evidence that one or more Arg residues are present in the allosteric sites of the spinach leaf enzyme. Both subunits of the spinach leaf enzyme were labeled when [^{14}C]phenylglyoxal was used (Ball and Preiss, 1992). Thus, Arg residues may also be involved in the binding of the allosteric ligands, particularly P_i. Site-directed mutagenesis was used to find out whether these five Arg residues were in some way responsible for the different regulatory properties. All five conserved Arg residues in the *Anabaena* ADPGlc PPase—that is, Arg 66, 105, 171, 294, and 385 were mutagenized to Ala (Sheng and Preiss, 1998). As shown in Table V, the Arg 294 Ala mutation resulted in a mutant enzyme with a much lower affinity for the inhibitor, phosphate, measured in the absence or presence of 3PGA. This mutation had no (or little) effect on the kinetic constants for the substrates or for the activator, 3PGA (Sheng and Preiss, 1998), and it can be concluded that Arg 294 of the *Anabaena* enzyme is involved in the binding of P_i. The activator, 3PGA, and the inhibitor, P_i, probably bind to different sites, although there could be some overlapping.

Another effect of the site-directed mutagenesis was that the purified mutant enzyme Arg 294 Ala had a 3-fold higher specific activity than the wild-type enzyme, suggesting that with disappearance of the inhibitor binding site there was also a conformational change, resulting in an enzyme with a higher catalytic efficiency. These results not only clarified another aspect of the structure–function relationships of the ADPGlc PPase, but also resulted in the creation of an enzyme that might be useful in the development of transgenic crops with higher starch production.

TABLE V

EFFECT OF SITE-DIRECTED MUTAGENESIS OF SEVERAL AMINO ACIDS ON THE RESPONSE OF THE *Anabaena* ADPGlc PPase TO 3-PGA

	3-P-glycerate	WT	R66A	R105A	R294A	R385A
$I_{0.5}$ P_i (mM)	—	0.055	0.26	0.077	5.2	0.062
	+	1.0	0.58	0.89	38	0.87
V_{max} (unit[a]/mg)	−	6.9	4.8	4.8	11	0.63
	+	60	44	79	170	13

[a] One unit of enzyme activity is defined as the amount of enzyme required to form 1 μmol of ADP-glucose/min at 37°C (assay in the direction of synthesis).

As discussed previously, cDNA clones encoding the putative mature forms of the large and small subunits of the potato tuber ADPGlc PPase have been expressed together, using two different compatible vectors, in an *E. coli* mutant deficient in ADPGlc PPase activity (Iglesias *et al.*, 1993; Ballicora *et al.*, 1995; Table II). This expression system was then used for site-directed mutagenesis experiments aiming to test whether the Lys residues in the potato tuber ADPGlc PPase have a role in activation, as suggested by the chemical modification (with pyridoxal-P) experiments of the spinach enzyme.

As shown in Table VI, site-directed mutagenesis of Lys 441 of the potato ADPGlc PPase small subunit to Glu and Ala results in mutant enzymes with lower affinity, 30- to 83-fold, respectively, for 3PGA (Ballicora *et al.*, 1996; Preiss *et al.*, 1996). A conservative mutation to arginine resulted in only a two-fold increase in $A_{0.5}$, indicating that the positive charge of the cationic amino acid is important for the binding of the activator. Mutagenesis of Lys residue 417 in the large subunit (the residue homologous to the *Anabaena* Lys residue 382 and to site 2 of the spinach leaf large subunit Lys residue modified by PLP) was also done. When Lys 417 was replaced by either Ala or Glu, the affinity for 3PGA decreased (Table V) but the increase in $A_{0.5}$ was only 3- to 13-fold and not as high as seen with the mutations of the small, 50-kDa subunit Lys 441 residue. When both Lys residues in the large (51-kDa) and small subunits were mutated, the decrease in affinity or increase in $A_{0.5}$ was additive. Thus, Lys residues in both subunits seem to contribute to the binding of the activator.

TABLE VI

SITE-DIRECTED MUTAGENESIS OF LYS RESIDUES AT THE BINDING SITE FOR THE ALLOSTERIC ACTIVATOR IN THE SUBUNITS OF THE POTATO TUBER ADPGlc PPASE. EFFECT ON THE SENSITIVITY OF THE HOLOENZYME TO THE ACTIVATOR, 3PGA[a]

ADPGlc PPase subunits			
Large	Small	3PGA $A_{0.5}$ (m*M*)	Ratio of $A_{0.5}$ mutant/$A_{0.5}$ wt
Wild-type	Wild-type	0.10	1
K417A	Wild-type	0.3	3
K417E	Wild-type	1.3	13
K417A	K441A	6.0	60
K417E	K417E	No activation	0
Wild-type	K441R	0.18	1.8
Wild-type	K441A	3.2	32
Wild-type	K441E	8.3	83

[a] Data from Preiss *et al.* (1996) and unpublished results of M. A. Ballicora and J. Preiss.

Random mutagenesis has also been used to determine whether other sequence regions or amino acids in the large subunit are important for the allosteric function (Greene *et al.*, 1996a,b). In one study (Greene *et al.*, 1996b), the Asp residue 416 (413 in the special notation used by Greene *et al.*) was mutated to an Ala residue and the affinity for 3PGA decreased about 6-fold, similar to the decrease observed when Lys 417 was mutated to Ala (Ballicora *et al.*, 1996; Table IV). In a second mutant isolated via random mutagenesis, Leu had replaced the proline residue 52 (Greene *et al.*, 1996a). The mutant enzyme's affinity for 3PGA was substantially decreased; the $A_{0.5}$ being increased 45-fold in mutant P52L, suggesting that a region of the large subunit *N*-terminal may also be involved in the formation of the allosteric activator binding site.

Giroux *et al.* (1996) described the effect of a single gene mutation in the *sh* 2 locus of maize (coding for the large subunit of the ADPGlc PPase), which increases seed weight by 11 to 18% without changing the proportion of the seed weight taken by starch. The direct effect of the mutation is the addition of two amino acids, tyrosine and serine, that seem to decrease the sensitivity of the ADPGlc PPase to inhibition by phosphate. This change in regulatory properties was found in the ADPGlc PPase measured in the seed extract and in the enzyme expressed in *E. coli.* When the researchers placed the two extra amino acids in the corresponding position of the potato tuber ADPGlc PPase, expressed in *E. coli,* they observed a similar decrease in sensitivity to P_i.

VII. CLONING OF THE ADPGlc PPase GENES AND COMPARISON OF THEIR SEQUENCES

Many cDNA or genomic clones for the small subunit ADPGlc PPase gene of rice endosperm (Krishnan *et al.*, 1986; Anderson *et al.*, 1989, 1990), maize endosperm (Barton *et al.*, 1986), spinach leaf (Preiss *et al.*, 1989), *A. thaliana* (B. Smith-White and J. Preiss, unpublished results, 1998), and potato tuber (Anderson *et al.*, 1990; Nakata *et al.*, 1991) have been isolated. In addition, a cDNA clone for the maize endosperm ADPGlc PPase large molecular subunit (*Sh* 2 locus) has also been isolated (Barton *et al.*, 1986). Olive *et al.* (1989) isolated cDNA clones from wheat leaf and wheat endosperm, which are now considered to represent the large subunit gene of the ADPGlc PPase, as suggested by the deduced amino acid sequence.

Although the isolation of the spinach leaf large subunit cDNA clone has not been reported, the major portion of the spinach leaf large subunit (54 kDa) has been sequenced by the Edmann degradation technique (B. Smith-White and J. Preiss, 1992). Since 1991, many other ADPGlc PPase

genes, either genomic or represented by a cDNA, have been isolated from many plants and different tissues, and they are too numerous to cite here. Figure 3 shows the deduced amino acid sequences of 45 subunits of ADPGlc PPases obtained from the EMBL nucleotide sequence library and GenBank.

At the DNA level, the isolated genes are dissimilar. For example, in wheat leaf and wheat endosperm, there is only 55.7% identity (Olive *et al.,* 1989) and, on the basis of Southern blot 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 approximately a 50% identity (B. S. White and J. Preiss, unpublished results, 1998).

Good identity is observed in comparing amino acid sequences of similar subunits of the ADPGlc PPase from the different plants, and this is expected since the spinach leaf lower-molecular-weight subunit antibody reacts well with the equivalent subunits of maize endosperm (Plaxton and Preiss, 1987; Preiss *et al.,* 1990), rice seed (Krishnan *et al.,* 1986; Anderson *et al.,* 1989), *Arabidopsis* leaf (Lin *et al.,* 1988a,b), and potato tuber (Okita *et al.,* 1990) enzymes. The lower-molecular-weight antibody does not react well with the higher-molecular-mass subunit of the ADPGlc PPase of these various plants. Therefore, it was not expected that much homology would be seen between the lower- and higher-molecular-weight subunits. However, there appears to be some identity (approximately 40–60%) between the large and small subunits of the higher plant ADPGlc PPase (Fig. 3).

Because of the relatively low but certain homology between the two subunits of the ADPGlc PPase, it can be speculated that they may have arisen originally from the same gene. The bacterial ADPGlc PPase is a homotetramer composed of only one subunit (Preiss, 1984). The cyanobacterial ADPGlc PPase has 3PGA as an allosteric activator and P_i as an inhibitor, similar to the enzyme from higher plants (Levi and Preiss, 1976), and unlike the bacterial enzymes (e.g., fructose-1,6-biphosphate is the activator in enteric bacteria). Both bacterial (Preiss, 1984; Preiss and Romeo, 1989) and cyanobacterial (Iglesias *et al.,* 1991) ADPGlc PPases are homotetrameric, unlike the higher plant enzymes, indicating that regulation by 3PGA and P_i (a good signaling system for a photosynthetic organism) is not related to the heterotetrameric nature of the higher plant enzyme. It is possible that during evolution there was duplication of the ADPGlc PPase gene, and divergence of the genes then produced two different genes coding for the two peptides, both of which were required for optimal activity of the native higher plant enzyme.

As indicated in the preceding, one can tentatively assign catalytic function to the small subunit of the ADPGlc PPase. The extensive identity and

a

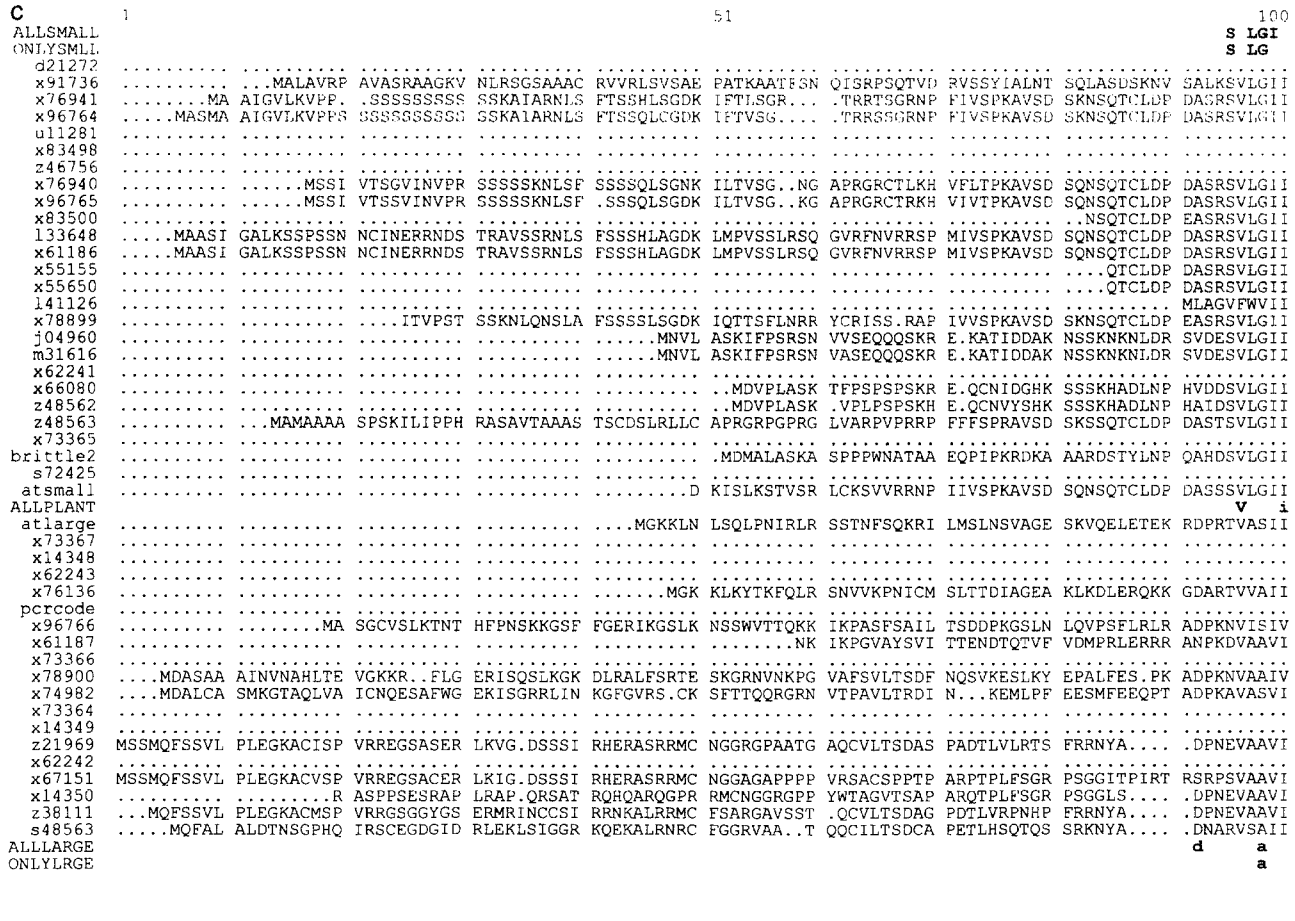
Name: d21272: cDNA, RT PCR, *Oryza sativa*, callus
Name: x91736: cDNA, *Chlamydomonas reinhardti*
Name: x76941: cDNA, *Vicia faba*, var. minor cv. Fribo, cotyledons
Name: x96764: cDNA, Pisum sativum, cv. sugar snap, cotyledons
Name: u11281: cDNA, *Ipomoea batatas*, strain White Star
Name: x83498: cDNA, *Ipomoea batatas*
Name: z46756: cDNA, *Ipomoea batatas*, strain White Star
Name: x76940: cDNA, *Vicia faba*, var. minor cv. Fribo, cotyledons
Name: x96765: cDNA, Pisum sativum, cv. sugar snap, cotyledons
Name: x83500: cDNA, *Spinacia oleracea*
Name: l33648: genomic, *Solanum tuberosum*, cv. Russett Burbank
Name: x61186: cDNA, *Solanum tuberosum*, cv. Russett Burbank, tuber
Name: x55155: cDNA, *Solanum tuberosum*, cv. Desiree {3112 Ebstorf}, tuber
Name: x55650: cDNA, *Solanum tuberosum*, cv. Desiree {3112 Ebstorf}, tuber
Name: l41126: cDNA, *Lycopersicon esculentum*, fruit
Name: x78899: cDNA, *Beta vulgaris*, cv. Zuchtlinie 5S0026, tap root
Name: j04960: cDNA, *Oryza sativa*, strain L.C.V. Biggs M201, endosperm
Name: m31616: cDNA, *Oryza sativa*, strain L.C.V. Biggs M201, leaf
Name: x62241: cDNA, RT PCR, *Hordeum vulgaris*, cv Bomi, endosperm (S39537)
Name: x66080: cDNA, *Triticum aestivum*, cv. Chinese Spring, leaf
Name: z48562: cDNA, *Hordeum vulgaris*, cv Bomi, starchy endosperm
Name: z48563: cDNA, *Hordeum vulgaris*, cv Bomi, leaf
Name: x73365: cDNA, RT PCR, *Arabidopsis thaliana*
Name: brittle2: cDNA, *Zea mays*, endosperm, *brittle-2* locus
Name: s72425: cDNA, *Zea mays*, leaf
Name: atsmall: cDNA, *Arabidopsis thaliana*, above-ground (B. Smith-White, pers. comm.)
Name: atlarge: cDNA, *Arabidopsis thaliana*, above-ground (B. Smith-White, pers. comm.)
Name: x73367: cDNA, RT PCR, *Arabidopsis thaliana*
Name: x14348: cDNA, *Triticum aestivum*, cv. Mardler, leaf

Name: x62243:	cDNA, RT PCR, *Hordeum vulgaris*, cv Bomi, 4 week seedling, (S39540)
Name: x76136:	cDNA, *Solanum tuberosum*, cv. Desiree {3112 Ebstorf}, tuber
Name: pcrcode:	genomic PCR, Spinacia oleracea (B. Smith-White, pers. comm.)
Name: x96766:	cDNA, Pisum sativum, cv. sugar snap, cotyledons
Name: x61187:	cDNA, *Solanum tuberosum*, cv. Russett Burbank, tuber
Name: x73366:	cDNA, RT PCR, *Arabidopsis thaliana*
Name: x78900:	cDNA, *Beta vulgaris*, cv. Zuchtlinie 5S0026, tap root
Name: x74982:	cDNA, *Solanum tuberosum*, cv. Desiree {3112 Ebstorf}, leaf
Name: x73364:	cDNA, RT PCR, *Arabidopsis thaliana*
Name: x14349:	cDNA, *Triticum aestivum*, cv. Mardler, endosperm
Name: z21969:	cDNA, *Triticum aestivum*, cv. Chinese Spring, developing grain
Name: x62242:	cDNA, RT PCR, *Hordeum vulgaris*, cv Bomi, endosperm (S39540)
Name: x67151:	cDNA, *Hordeum vulgaris*, cv Bomi, endosperm
Name: x14350:	cDNA, *Triticum aestivum*, cv. Mardler, endosperm
Name: z38111:	cDNA, *Zea mays*, embryo
Name: s48563:	cDNA, *Zea mays*, endosperm, *shrunken-2* locus

b

Five consensus classes:
ONLYSMLL - residue found only in small subunit class, diagnostic for class membership
ALLSMALL - residue found in all members of small subunit class,
ONLYLRGE - residue found only in large subunit class, diagnostic for class membership
ALLLARGE - residue found in all members of large subunit class,
ALLPLANT - residue found in all plant proteins.
Uppercase - no exceptions, lowercase - one or two exceptions

FIG. 3.


C
1
51
100
ALLSMALL
ONLYSMLL
S LGI
S LG
d21272
x91736 ...MALAVRP AVASRAAGKV NLRSGSAAAC RVVRLSVSAE PATKAATFSN QISRPSQTVD RVSSYIALNT SQLASDSKNV SALKSVLGII 87
x76941 .....MA AIGVLKVPP. .SSSSSSSSS SSKAIARNLS FTSSHLSGDK IFTLSGR... .TRRTSGRNP FIVSPKAVSD SKNSQTCLDP DASRSVLGII 86
x96764 .....MASMA AIGVLKVPPS SSSSSSSSSS SSKAIARNLS FTSSQLCGDK IFTVSG.... .TRRSSGRNP FIVSPKAVSD SKNSQTCLDP DASRSVLGII 90
u11281
x83498
z46756
x76940 ......MSSI VTSGVINVPR SSSSSKNLSF SSSSQLSGNK ILTVSG..NG APRGRCTLKH VFLTPKAVSD SQNSQTCLDP DASRSVLGII 82
x96765 ......MSSI VTSSVINVPR SSSSSKNLSF .SSSQLSGDK ILTVSG..KG APRGRCTRKH VIVTPKAVSD SQNSQTCLDP DASRSVLGII 81
x83500 ..NSQTCLDP EASRSVLGII 18
133648 .....MAASI GALKSSPSSN NCINERRNDS TRAVSSRNLS FSSSHLAGDK LMPVSSLRSQ GVRFNVRRSP MIVSPKAVSD SQNSQTCLDP DASRSVLGII 95
x61186 .....MAASI GALKSSPSSN NCINERRNDS TRAVSSRNLS FSSSHLAGDK LMPVSSLRSQ GVRFNVRRSP MIVSPKAVSD SQNSQTCLDP DASRSVLGII 95
x55155 ....QTCLDP DASRSVLGII 16
x55650 ....QTCLDP DASRSVLGII 16
141126 MLAGVFWVII 10
x78899 ....ITVPST SSKNLQNSLA FSSSSLSGDK IQTTSFLNRR YCRISS.RAP IVVSPKAVSD SKNSQTCLDP EASRSVLGII 75
j04960 ......MNVL ASKIFPSRSN VVSEQQQSKR E.KATIDDAK NSSKNKNLDR SVDESVLGII 53
m31616 ......MNVL ASKIFPSRSN VASEQQQSKR E.KATIDDAK NSSKNKNLDR SVDESVLGII 53
x62241
x66080 ..MDVPLASK TFPSPSPSKR E.QCNIDGHK SSSKHADLNP HVDDSVLGII 47
z48562 ..MDVPLASK .VPLPSPSKH E.QCNVYSHK SSSKHADLNP HAIDSVLGII 46
z48563 ...MAMAAAA SPSKILIPPH RASAVTAAAS TSCDSLRLLC APRGRPGPRG LVARPVPRRP FFFSPRAVSD SKSSQTCLDP DASTSVLGII 87
x73365
brittle2 .MDMALASKA SPPPWNATAA EQPIPKRDKA AARDSTYLNP QAHDSVLGII 49
s72425
atsmall .....D KISLKSTVSR LCKSVVRRNP IIVSPKAVSD SQNSQTCLDP DAGSSVLGII 51
ALLPLANT V i
atlarge ....MGKKLN LSQLPNIRLR SSTNFSQKRI LMSLNSVAGE SKVQELETEK RDPRTVASII 56
x73367
x14348
x62243
x76136 ....MGK KLKYTKFQLR SNVVKPNICM SLTTDIAGEA KLKDLERQKK GDARTVVAII 53
pcrcode
x96766 .....MA SGCVSLKTNT HFPNSKKGSF FGERIKGSLK NSSWVTTQKK IKPASFSAIL TSDDPKGSLN LQVPSFLRLR ADPKNVISIV 82
x61187 .....NK IKPGVAYSVI TTENDTQTVF VDMPRLERRR ANPKDVAAVI 42
x73366
x78900 ....MDASAA AINVNAHLTE VGKKR..FLG ERISQSLKGK DLRALFSRTE SKGRNVNKPG VAFSVLTSDF NQSVKESLKY EPALFES.PK ADPKNVAAIV 93
x74982 ....MDALCA SMKGTAQLVA ICNQESAFWG EKISGRRLIN KGFGVRS.CK SFTTQQRGRN VTPAVLTRDI N...KEMLPF EESMFEEQPT ADPKAVASVI 92
x73364
x14349
z21969 MSSMQFSSVL PLEGKACISP VRREGSASER LKVG.DSSSI RHERASRRMC NGGRGPAATG AQCVLTSDAS PADTLVLRTS FRRNYA.... .DPNEVAAVI 95
x62242
x67151 MSSMQFSSVL PLEGKACVSP VRREGSACER LKIG.DSSSI RHERASRRMC NGGAGAPPPP VRSACSPPTP ARPTPLFSGR PSGGITPIRT RSRPSVAAVI 99
x14350 .....R ASPPSESRAP LRAP.QRSAT RQHQARQGPR RMCNGGRGPP YWTAGVTSAP ARQTPLFSGR PSGGLS.... .DPNEVAAVI 75
z38111 ...MQFSSVL PLEGKACMSP VRRGSGGYGS ERMRINCCSI RRNKALRRMC FSARGAVSST .QCVLTSDAG PDTLVRPNHP FRRNYA.... .DPNEVAAVI 91
s48563 .....MQFAL ALDTNSGPHQ IRSCEGDGID RLEKLSIGGR KQEKALRNRC FGGRVAA..T QQCILTSDCA PETLHSQTQS SRKNYA.... .DNARVSAII 88
ALLLARGE d a
ONLYLRGE a

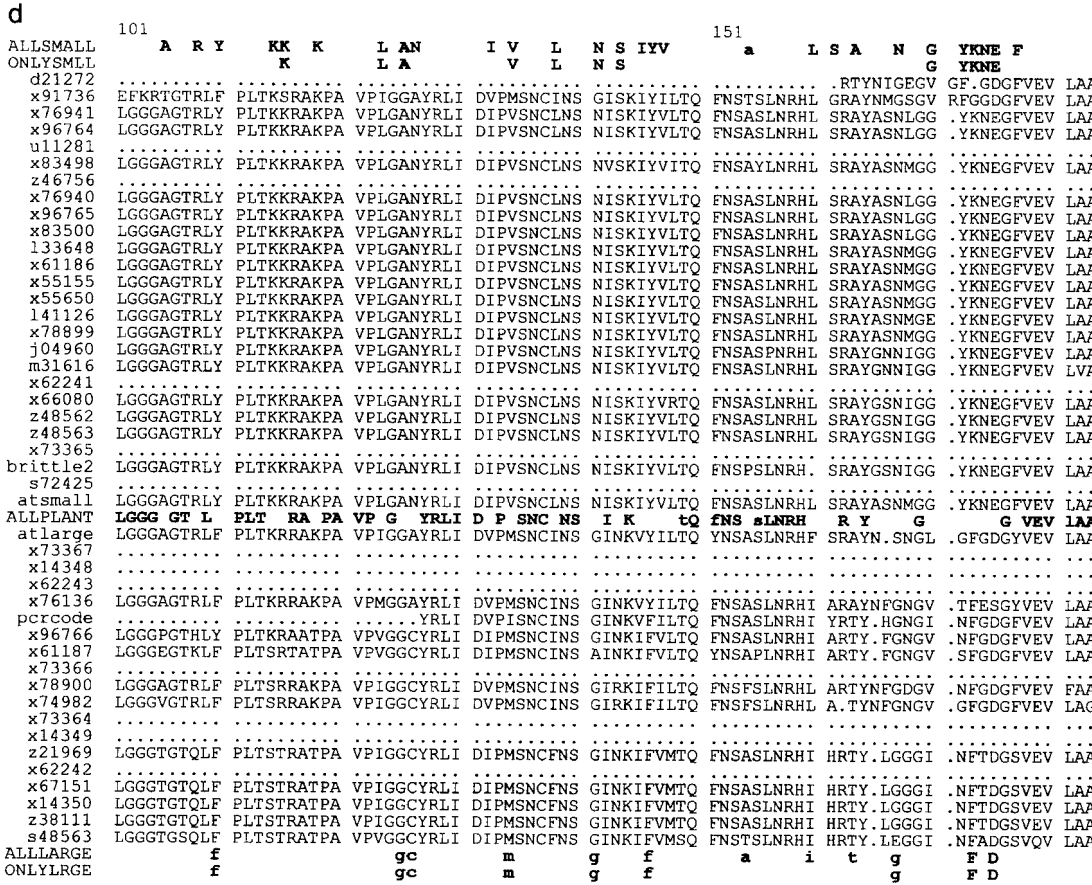
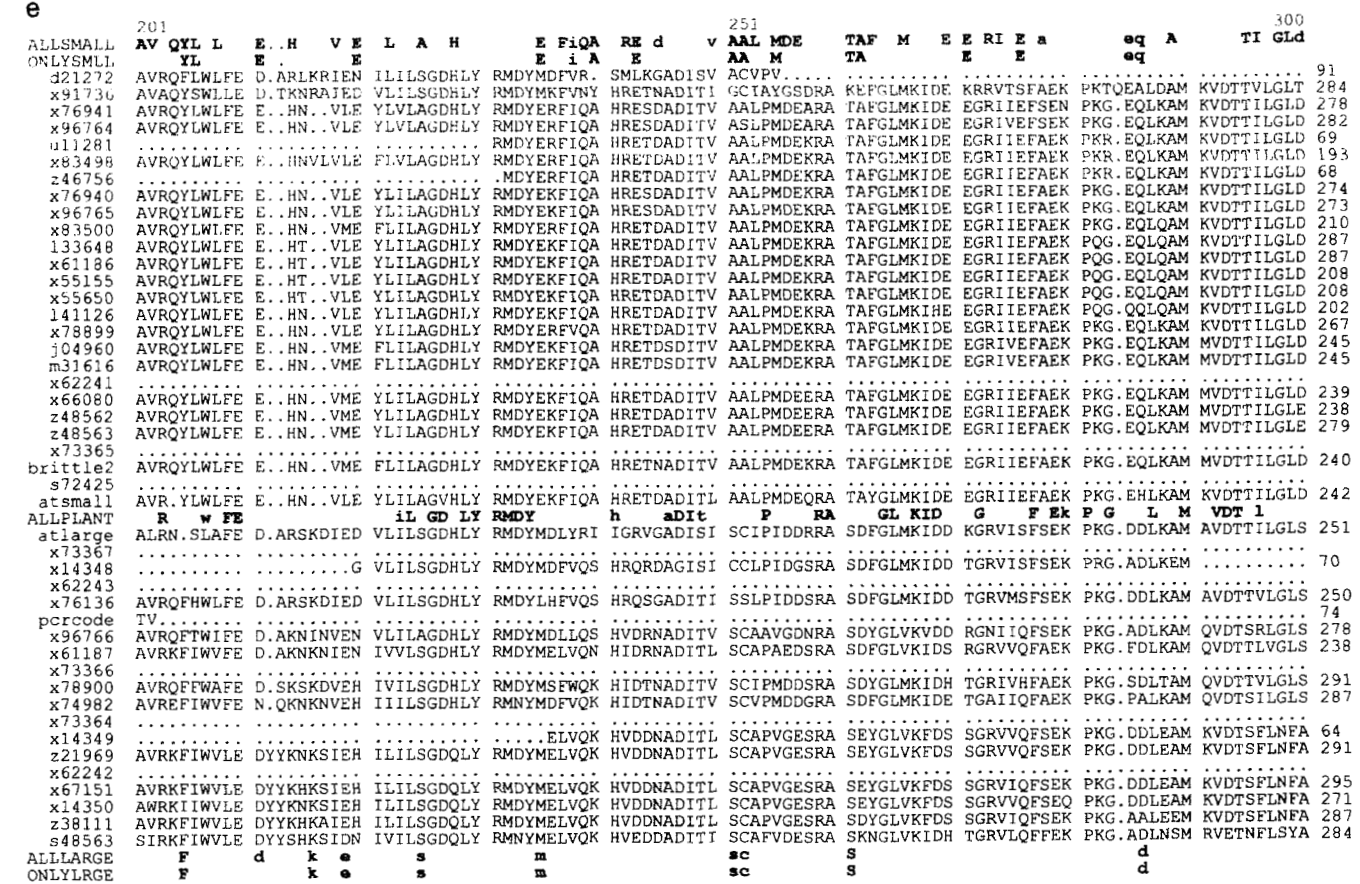


FIG. 3. *Continued*

e



f

```
           301                                                351                                              400
  ALLSMALL D. RAKe         I  i S  VML LLR   FPgAN      VI G t   G r Q  L d G        e A YN  LGIT KKP PD S   RS   I   P
  ONLYSMLL    R                  I  i S    M       g      V  G     G r      d          e A       I     K   D     R
    d21272 .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    x91736 P.EEAAEKPY IASMG.IYVF KKSVLLQLLN DSYAKANDFG GEIIPSAAK. DHNVVAYPFY GYWEDIGTIK SFFEENLKSC HP..ATFEFY DPQSPIYTSP 379
    x76941 D.DRAKEMPY IASMG.IYVV SKHVMLDLLR DKFPGANDFG SEVIPGATEL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 376
    x96764 D.ERAKEMPY IASMG.IYVV SKHVMLDLLR DKFPGANDFG SEVIPGATEL GLRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 380
    u11281 D.QRAKELPF IASMG.IYVI SKNVMLNLLR EKFPGANDFG SEVIPGATSI GMRVQAYLFD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPISTTP 167
    x83498 D.QRAKELPF IASMG.IYVI SKNVMLNLLR EKFPGANDFG SEVIPGATSI GMRVQAYLFD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPISTTP 291
    z46756 D.QRAKELPF IASMG.IYVI SKNVMLNLLR EKFPGANDFG SEVIPGATSI GMRVQAYLFD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPISTTP 166
    x76940 D.ERAKEMPF IASMG.IYVI SKNVMLDLLR DKFPGANDFG SEVIPGATSI GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 372
    x96765 D.ERAKEMPF IASMG.IYVI SKNVMLDLLR DKFPGANDFG SEVIPGATSV GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 371
    x83500 D.ERAKEMPY IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSI GLTVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 208
    133648 D.KRAKEMPF IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 385
    x61186 D.KRAKEMPF IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 385
    x55155 D.KRAKEMPF IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 306
    x55650 D.KRAKEMPF IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 306
    141126 D.KRAKEMPF IASMG.IYVI SKDVMLNLLR DKFPGANDFG SEVIPGATSL GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 300
    x78899 D.ERAKEMPF IASMG.IYVI SKDVMLNLLR EQFPGANDFG SEVIPGATSI GLRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSSPIYTQP 365
    j04960 D.VRAKEMPY IASMG.IYVI SKNVMLQLLR EQFPGANDFG SEVIPGATNI GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 343
    m31616 D.VRAKEMPY IASMG.IYVI SKNVMLQLLR EQFPGANDFG SEVIPGATNI GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 343
    x62241 ......KYPY IAGMG.IYVI SKHVMLQLLR EQFPGANDFG SEVIPGATST GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPIPDFSFY DRSAPIYTQP 93
    x66080 D.ARAKEMPY IASMG.IYVI SKHVMLQLLR EQFPGANDFG SEVIPGATST GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPIPDFSFY DRSAPIYTQP 337
    z48562 D.ARAKEMPY IASMG.IYVI SKHVMLQLLR EQFPGANDFG SEVIPGATST GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPIPDFSFY DRSAPIYTQP 336
    z48563 D.ARAKEMPY IASMG.IYVI SKHVMLQLLR EQFPGANDFG SEVIPGATST GMRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPIPDFSFY DRSAPIYTQP 377
    x73365 .......YPY IAGMG.IYVV SRDVMLDLLR NQFPGANDFG SEVIPGAPFL GLRVQAYLYD GYWEDIGTIE AFYNANLGIT KKPVPDFSFY DRSAPIYTQP 92
  brittle2 D.VRAKEMPY IASMG.IYVF SKDVMLQLLR EQFPEANDFG SEVIPGATSI GKRVQAYL.H GYWEDIGTIA AFYNANLGIT KKPIPDFSFY DRFAPIYTQP 337
    s72425 .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   atsmall D.QRAKEMPF IA........ .......... .......... .......... .......... .......... .......... .......... .......... 353
  ALLPLANT         Py IASMG Yv        l LL        DFG SE  P a       V AY     YWEDiGTi   F  An       P   F FY D   P  T
   atlarge K.EEAEKKPY IASMG.VYVF KKEILLNLLR WRFPTANDFG SEIIPFSAK. EFYVNAYLFN DYWEDIGTIR SFFEANLALT EHP.GAFSFY DAAKPIYTSR 347
    x73367 .......YPY IAGMG.VYVF KKEILLNLLR WRFPTANDFG SEIILL.AK. EFYVNAYLSN DYWEHIGTIR SFFEANLALT EHP.GAFSFY DAAKPIYTSR 89
    x14348 ..EEAEKKPY IASMG.VYIF KKEILLNLLR WRFPTANDFG SEIIPAAAR. EINVKAYLFN DYWEDIGTIK SFFEANLALA EQP.SKFSFY DASKPMYTSR 165
    x62243 ......KYPY IAGMG.VYIF KKEILLNLLR WRFPTANDFG SEIIPAAAR. EINVKAYLFN DYWEDIGTIK SFFEANLALA EQP.SKFSFY DASKPMYTSR 91
    x76136 P.EEAKEKPY IASIGKVYVF KKDILLNLLR WRFPTANDFG SEIIPASTK. EFCVKAYLFN DYWEDIGTIR SFFRANLALT EHP.PRFSFY DATKPIYTSR 347
   pcrcode .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    x96766 P.QDALKSPY IASMG.VYVF KKDVLLKLLK WRYPTSNDFG SEIIPSAIR. EHNVQAYFFG DYWEDIGTIK SFYDANLALT EES.PKFEFY DPKTPIFTSP 374
    x61187 P.QDAKKSPY IASMG.VYVF KTDVLLKLLK WSYPTSNDFG SEIIPAAID. DYNVQAYIFK DYWEDIGTIK SFYNASLALT QEF.PEFQFY DPKTPFYTSP 334
    x73366 .......YPY IAGMG.VYVF RKEGLLKLLR SSYPTSNDFG SEIIRARRK. LHNVQAFLFN DYWEDIGTIG SFFDANLALT EQP.PKFQFY DQKTPFFTSP 90
    x78900 DLE.AMSNPY IASMG.VYVF RTDVLMELLN RKYPSSNDFG SEIIPSAVG. ESNVQAYLFN DYWEDIGTIK SFFDSNLALT QQP.PKFEFY DPKTPFYTSA 387
    x74982 EQE.ASNFPY IASMG.VYVF KTDVLLNLLK SAYPSCNDFG SEIIPSAVK. DHNVQAYLFN DYWEDIGTVK SFFDANLALT KQP.PKFDFN DPKTPFYTSA 383
    x73364 .......YPY IAGMG.VYCF KTEALLKLLT WRYPSSNDFG SEIIPAAIK. DHNVQGYIYR DYWEDIGTIK SFYEASIALV EEH.PKFEFY DQNTPFYTSP 90
    x14349 .IDDPAKYPY IASMG.VYVF KRDVLLNLLK SRYAELHDFG SEILPRALH. DHNVQAYVFT DYWEDIGTIR SFFDANMSLC EQP.PKFEFY DPKTPFFTSP 160
    z21969 .IDDPAKYPY IASMG.VYVF KRDVLLNLLK SRYAELHDFG SEILPRALH. DHNVQAYVFT DYWEDIGTIR SFFDANMALC EQP.PKFEFY DPKTPFFTSP 387
    x62242 ......KYPY IASMG.VYVF KRDVLLNLLK SRYAELHDFG SEILPRALH. DHNVQAYVFT DYWEDIGTIR SFFDANMALC EQP.PKFEFY DPKTPFFTSP 91
    x67151 .IDDPAKYPY IASMG.VYVF KRDVLLNLLK SRYAELHDFG SEILPRALH. DHNVQAYVFT DYWEDIGTIR SFFDANMALC EQP.PKFEFY DPKTPFFTSP 391
    x14350 .IDDPAKYPY IASMG.VYVF KRDVLLNLLK SRYAELHDFG SEILPRALH. DHNVQAYVFT DYWEDIGTIR SFFDANRALC EQP.PKFEFY DPKTPFFTSP 367
    z38111 TCTLPAEYPY IASMG.VYVF KRDVLLDLLK SRYAELHDFG SEILPKALH. EHNVQAYVFT DYWEDIGTIR SFFDANMALC EQP.PKFEFY DPKTPFFTSP 384
    s48563 .IDDAQKYPY LASMG.IYVF KKDALLDLLK SKYTQLHDFG SEILPRAVL. DHSVQACIFT GYWEDVGTIK SFFDANLALT EQP.SKFDFY DPKTPFFTAP 380
  ALLLARGE          Y        v F K   L                  I      .            d          S F    aL         .
  ONLYLRGE                   v   K   L                  I      .            d          S F    aL         .
```

FIG. 3. *Continued*

g

```
                  431                                                 451                                                 500
ALLSMALL      SK LD AD TDSV GE   VIKN KI    VV L  CI SEGAIIED L LM   Y     ad   L  k        I     h   A
ONLYSMLL         LD AD      GE   VIK          L  C  E AIIE  L L             ad   L  k        I     h   A
  d21272  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  x91736  ALLPPATVRN CKVIDAIIAQ GSFVSDCTIN NAVIGIRSII GQNCTIQDAL VMGADYYESD DQRATLL..K KGGVPVGIGA NSVITNAIID KNARVGKNVK 477
  x76941  RYLPPSKMLD ADITDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYY.ET DADRRFLAAK GG.VPIGIGK NSHIRRAIID KNARIGDDVK 474
  x96764  RYLPPSKMLD ADITDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYY.ET DADRRFLAAK GG.VPIGIGK NSHIKRAIID KNARIGDDVK 478
  u11281  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRRLLAAK GS.VPIGIGR NSHIKRAIIH NIARIGNDVK 265
  x83498  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRRLLAAK GS.VPIGIGR NSHIKRAIIH NIARIGNDVK 389
  z46756  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRRLLAAK GS.VPIGIGR NSHIKRAIIH NIARIGNDVK 264
  x76940  RYLPPSKMLD ADITDSVIGE GCVIKNCKIF HSVVGLRSCI SEGAIIEDTL LMGADYY.ET EADKRFLAAK GS.VPIGIGK NSHIKRAIVD KNARIGENVK 470
  x96765  RYLPPSKMLD ADITDSVIGE GCVIKNCKIF HSVVGLRSCI SEGAIIEDTL LMGADYY.ET EADKRFLAAK GS.VPIGIGK NSHIKRAIVD KNARIGENVK 469
  x83500  RYLPPSKMLD ADITDSVIGE GCVIKNCKIH HSVIGLRSCI SEGAIIEDTL LMGADYY.ET DADRKLLAAK GS.VVLGIGQ NSHIKRAIID KNARIGDNVK 306
  l33648  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRKLLAAK GS.VPIGIGK NCHIKRAIID KNARIGDNVK 483
  x61186  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRKLLAAK GS.VPIGIGK NCHIKRAIID KNARIGDNVK 483
  x55155  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRKLLAAK GS.VPIGIGK NCHIKRAIID KNARIGDNVK 404
  x55650  RYLPSSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DADRKLLAAK GS.VPIGIGK NCHIKRAIID KNARIGDNVK 404
  l41126  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET DAERKLLAAK GSVVPIGIGK NCLYKRAIID KNARIGDNVK 399
  x78899  RYLPPSKMLD ADITDSVIGE GCVIKNCKIH HSVIGLRSCI SEGAIIEDTL LMGADYY.ET DADRKFLAAK GS.VPIGIG. .......... .NARIGDDVK 452
  j04960  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET EADKKLLGEK GG.IPIGIGK NCHIRRAIID KNARIGDNVK 441
  m31616  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET EADKKLLGEK GG.IPIGIGK NCHIRRAIID KNARIGDNVK 441
  x62241  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDAL LMGADYY.ET EADKKLLAEK GG.IPIGIGK NSHIKRAIID KNARI..... 186
  x66080  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYY.ET EADKKLLAEK GG.IPIGIGK NSHIKRAIID KNARIGDNVM 435
  z48562  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYY.ET EADKKLLAEK GG.IPIGIGK NSHIKRAIID KNARIGDNVM 436
  z48563  RHLPPSKVLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYY.ET EADKKLLAEK GG.IPIGIGK NSHIKRAIID KNARIGDNVM 475
  x73365  RYLPPSKMLD ADVTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDSL LMGADYY.ET ATEKSLLSAK GS.VPIGIGK NSHIKRAIID INARIG.... 185
brittle2  RHLPPSKVLD ADVTDSVI.E GCVIKNCKIN HSVVGLRSCI SEGAIIEDSL LMGADYY.ET EADKKLLAEK GG.IPIGIGK NSCIRRAIID KNARIGDNVK 434
  s72425  .......... ..VTDSVIGE GCVIKNCKIH HSVVGLRSCI SEGAIIEDTL LMGADYYAET EADKKLLAEN GG.IPIGIGK NSHIRKAIID KNARIGDNVK 87
 atsmall  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ALLPLANT  r LPP        i   Gc   C   hs  G rS          GAD Y          G  PiG G  N      iid  Narig  v
 atlarge  RNLPPSKIDN SKLIDSIISH GSFLTNCLIE HSIVGIRSRV GSNVQLKDTV MLGADYYKTE AEVAALLAE. .GNVPIGIGE NTKIQECIID KNARVGKNVI 445
  x73367  RNLPPSKIDN SKLIDSIISH GSFLTNCLIE HSIVGIRSRV GSNVQLKDTV MLGADYYQTE AEVAALLAE. .GNVPIGIGE NTKIQEC.ID KNARIG.... 182
  x14348  RNLPPSMISG SKITDSIISH GCFLDKCRVE HSVVGIRSRI GSNVHLKDTV MLGADFYETD MERGDQLAE. .GKVPIGIGE NTSIQNCIID KNARIGKNVT 263
  x62243  RNLPPSMISG SKITDSIISH GCFLDKCRVE HSVVGIRSRI GSNVHLKDTV MLGADFYETD AERGDQLAE. .GKVPIGIGE NTSIQNCIID MN..... 181
  x76136  RNLPPSAIDN SKIVDSIVSH GIFLTNCFVE HSVVGIRSRI GTNVHLKDTV MLGADYYETD AEIRSQLAE. .GKVPLGIGE NTRIKDCIID KNARIGKNVV 445
 pcrcode  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  x96766  GFLPPTKIDN SRVVDAIISH GCFLRDCTIQ HSIVGERSRL DYGVELQDTV MMGADYYQTE SEIASLLAE. .GKVPIGIGR NTKIKNCIID KNAKIGKEVV 472
  x61187  RFLPPTKIDN CKIKDAIISH GCFLRDCSVE HSIVGERSRL DCGVELKDTF MMGADYYQTE SEIASLLAE. .GKVPIGIGE NTKIRKCIID KNAKIGKNVS 432
  x73366  RFLPPTKVDK CRILDSIVSH GCFLRECSVQ HSIVGIRSRL ESGVELQDTM MMGADFYQTE AEIASLLAE. .GKVPVGVGQ NTRIKNCIID INARIG.... 184
  x78900  RFLPPTKVDR CKIVDSIVSH GCFLQESSIQ HSIVGVRSRL ESGVEFQDTM MMGADYYQTE SEIASLLAE. .GKVPVGVGQ NTKIKNCIID KNAKIGKDVV 485
  x74982  RFLPPTKVDK SRIVDAIISH GCFLRECNIQ HSIVGVRSRL DYGVEFKDTM MMGADYYQTE CEIASLLAE. .GKVPIGVGP NTKIQNCIID KNAKIGKDVV 481
  x73364  RFLPPTKTEK CRIVNSVISH GCFLGECSIQ RSIIGERSRL DYGVELQDTL MLGADSYQTE SE.SRLLAE. .GNVPIGIGR DTKIRKCIID KNARIG.... 186
  x14349  RYLPPTKSDK CRIKEAIILH GCFLRECKIE HSIIGVPSRL NSGSELKNAM MMGADSYETE DEISRLMSE. .GKVPIGVGE NTKISNCIID MNARIGRDVV 258
  z21969  RYLPPTKSDK CRIKEAIISH GCFLRECKIE HSIIGVRSRL NSGSELKNAM MMGADSYETE DEISRLMSE. .GKVPIGVGE NTKISNCIID MNARIGRDVV 485
  x62242  RYLPPTKSDK CRIKEAIISH GCFLRECKIE HSIIGVRSRL NSGSELKNAM MMGADSYETE DEISRLMSE. .GKVPIGVGE NTKISNSYYD MNARI..... 183
  x67151  RYLPPTKSDK CRIKEAIISH GCFLRECKIE HSIIGVRSRL NSGSELKNAM MMGADSYETE DEISRLMSE. .GKVPIGVGE NTKISNCIID MNARIGRDVV 489
  x14350  RYLPPTKSDK CRIKEAIILH GCFLRECKIE HTAF...SRL NSGSELKNAM MMGADSYETE DEMSRLMSE. .GKVPIGVGE NTKISNCIID MNARIGRDVV 462
  z38111  RYLPPTKSDK CRIKDAIISH GCFLRECAIE HSIVGVPSRL NSGCELKNTM MMGADLYETE DEISRLLAE. .GKVPIGVGE NTKISNCIID MNCQGWKERL 482
  s48563  RCLPPTQLDK CKMKYAFISD GCLLRECNIE HSVIGVCSRV SSGCELKDSV MMGADIYETE EEASKLLLA. .GKVPIGIGR NTKIRNCIID MNARIGKNVV 478
ALLLARGE                  sh  fL           R     l    M      T  E     e. .  V       T I  c
ONLYLRGE                  sh  fL           R     l    M         E     e. .          T    c
```

h

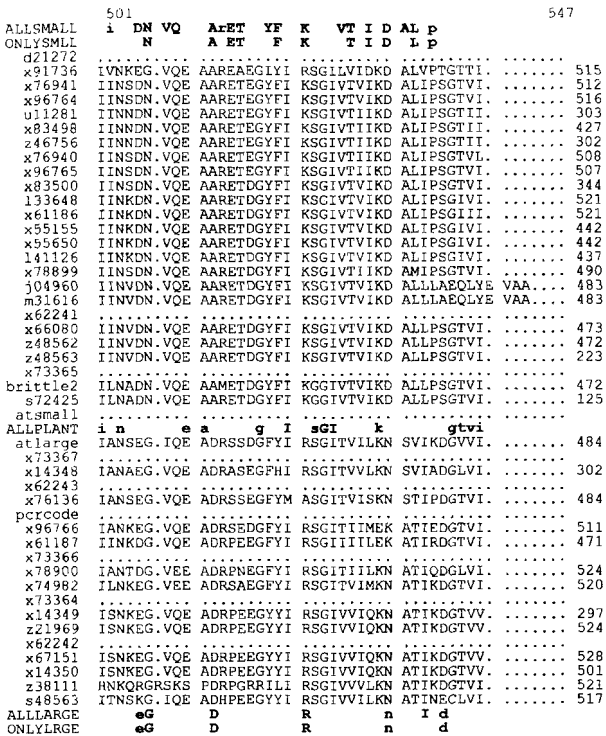


FIG. 3. Alignment of the primary structures of ADPGlc PPase proteins from various plants. The sequences were obtained from either GenBank or EMBL nucleotide sequence library, except for *Zea mays brittle* 2 (obtained from L. C. Hannah) and *Arabidopsis thaliana* large and small subunits (B. S. White and J. Preiss, unpublished results, 1998). Alignment of the sequences was done essentially as described by Smith-White and Preiss (1992). The small subunit sequences are shown in the top and the large subunit in the bottom. Also shown are five diagnostic classes that have been formulated for residues in the sequence, which are residues in sequences found only in the small subunit, residues found in all members of the small subunit class sequences, residues found in sequences present only in the large subunit, residues found in all members of the large subunit class, and residues found in all plant ADPGlc PPases.

similarity in sequence between the small subunits isolated from different plants and tissues supports this view. In the case of the large subunit, in which amino acid sequences have less similarity to what is observed for the small subunits, it is possible that the different large subunits lend different regulatory properties for the heterotetrameric ADPGlc PPases of different species and/or tissues. Thus, because the different sequences of the large subunit reflect their occurrences in different plant tissues (e.g., leaf, stem, guard cells, tuber, endosperm, root) (Smith-White and Preiss, 1992), it is possible that these sequence differences render the isolated enzyme from different tissues to have different allosteric properties.

VIII. HYDROPHOBIC CLUSTER ANALYSIS

Preparation of single crystals followed by X-ray diffraction analysis can picture accurately the structure of a protein at a high, atomic resolution. Computing advances have accelerated the process of converting a diffraction pattern into a molecular model. However, crystallization is far from a routine procedure, as the conditions required by a particular protein can only be found by screening a multitude of media (now available commercially) known to favor crystallization. Obtaining a crystal is a hit-or-miss business with no theory; the proteins whose structures have been revealed so far have not been chosen for their interest but because of their propensity to crystallize. Sometimes good crystals can be grown, but they do not diffract because they have very large unit cell dimensions or they decay rapidly in the X-ray beam. Some proteins are easier to crystallize than others, and the ADPGlc PPase from *E. coli* and *Anabaena* are among the more difficult ones. One factor affecting crystallization is the high degree of hydration of the molecule, and so far only small crystals have been obtained, and these were unstable under X-ray diffraction (Mulichak *et al.*, 1988).

Until good crystals have been obtained, there are other avenues for obtaining information about the structure of proteins that are difficult to crystallize. As observed by Kendrew when he solved the structure of the myoglobin, the main driving force for folding water-soluble globular protein molecules is to pack hydrophobic side chains into the interior of the molecule, thus creating a hydrophobic core and a hydrophilic surface. The main chain in the interior is arranged in secondary structures to neutralize its polar atoms through hydrogen bonds. There are two main types of secondary structure: alpha- (α-) helices and beta- (β-) sheets. Protein structures are built up by a combination of secondary structural elements, α-helices, and β-strands. These form the core regions—the interior of the molecule—

and are connected by loop regions at the surface. Schematic diagrams where these structures are highlighted are useful; in addition, α-helices and β-strands that are adjacent in the amino acid sequence are usually adjacent in the tertiary structure.

Hydrophobic cluster analysis (HCA) is a technique that displays the clusters of hydrophobic amino acids present within the primary sequence. It has been used to align amino acid sequences, to predict secondary structures, and to help find similar structures in proteins with a low homology (Lemesle-Varloot *et al.,* 1990). The amino acids are plotted as an α-helix and the representation is duplicated to avoid cutting off clusters with the "wrapping" that occurs when one turn of the helix is completed. In the original method, hydrophobic amino acids are highlighted and encircled to signal the presence of a hydrophobic cluster. This technique was applied to the ADPGlc PPase from *E. coli* (Ballicora *et al.,* 1996) and, since the technique is most useful when homologous proteins are compared, the sequence of the enzyme from a cyanobacteria (*Anabaena*) was also analyzed. The modification by Rost and Sander (1993) of the original technique facilitates the identification of clusters and, in the case of ADPGlc PPases from *E. coli* and *Anabaena,* it stresses the similarities between the two proteins. Proline and glycine are known "breakers" of helices and sheets, and a cluster is not drawn when one of these amino acids is included in it. From the hydrophobic analysis using the profile neural network (PHD) program, it is clear that the ADPGlc PPases from *E. coli* and *Anabaena* are identical in the position of many clusters, and in some others the differences are small. There are some insertions and deletions in the sequence, but they do not alter the general pattern of the clusters because, in these insertions, the analysis shows no buried amino acids. This suggests that the small insertions seen among ADPGlc PPases are not part of the "core" of the protein. Analysis of higher plant ADPGlc PPases show a similar pattern of clusters. For example, even though the homology in amino acid sequence is lower between the enzyme from *E. coli* and the small and large subunits of the potato tuber enzyme, all the clusters present in the bacterial enzyme are also present in both subunits of the plant ADPGlc PPase. This indicates that the ADPGlc PPases from different sources share a common folding pattern, despite a different quaternary structure (heterotetramer in plants, homotetramer in bacteria) and a different specificity for the activator.

If the ADPGlc PPases from different sources have similar three-dimensional structures, the structure of one should help predict the secondary structure of another. The sequence of enzymes from *E. coli* and *Anabaena,* and also from the two subunits of the potato tuber enzyme, were analyzed using the PHD program. One general structure that fits all of

these proteins was predicted (Fig. 4). The ADPGlc PPase is an α/β protein, but some parts of it are mainly beta, such as the C-terminal and the domain denoted as 3. To verify whether the model is valid, it was tested against the biochemical data available, including the results of partial proteolysis that is, trypsin treatment of the *Anabaena* and the *E. coli* enzymes (Y. Y. Charng and J. Preiss, unpublished results, 1992) and proteinase K (M. Wu and J. Preiss, unpublished results, 1997) digestion of the *E. coli* enzyme. The peptides obtained by protease treatment were analyzed (Fig. 5). Exposed loops would be more sensitive to proteolysis, and the protease studies, which actually cut in sites predicted as loops by the model, confirm the structure proposed. The only exception is the α-helix predicted near the

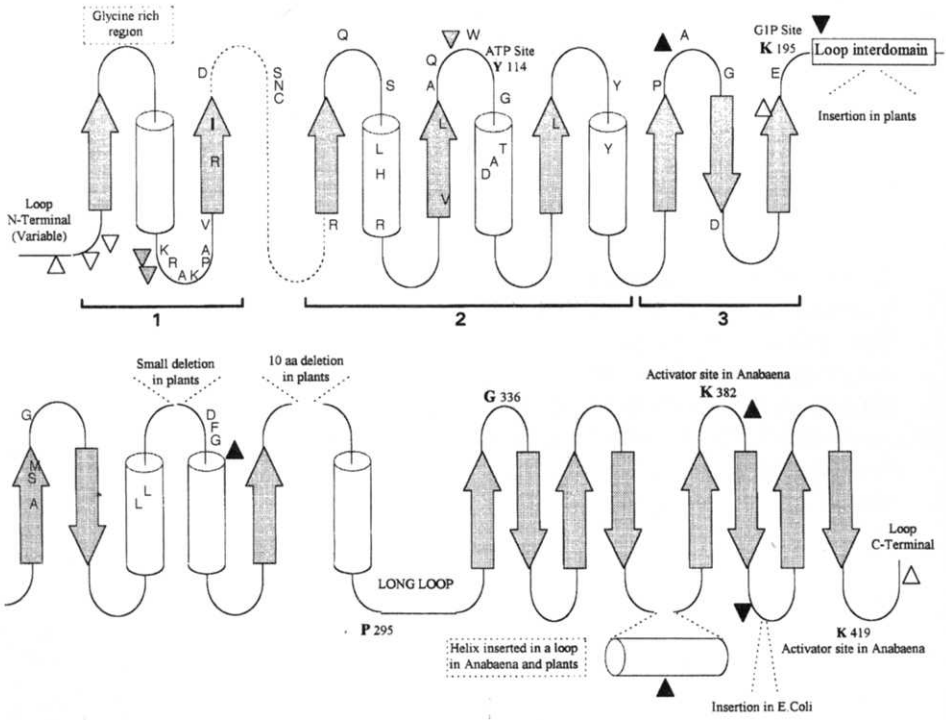


FIG. 4. Profile neural network (PHD) prediction of the secondary structure of the ADPGlc PPase. The structure shown was obtained by applying the program to the sequences of ADPGlc PPases from *E. coli* and *Anabaena*, and the two subunits of the potato tuber ADPGlc PPase. Section 1 contains the Fru-1,6-BP activator site KRAKPAV in a loop as well as R67. Section 2 has the putative ATP binding site, Y114, in a loop area between a β-strand and an α-helix starting at GTAD. The Glc-1-P binding site is also seen in a loop among a series of predicted β-strands. The topology between regions 1 and 2 cannot be ascertained (dotted line).

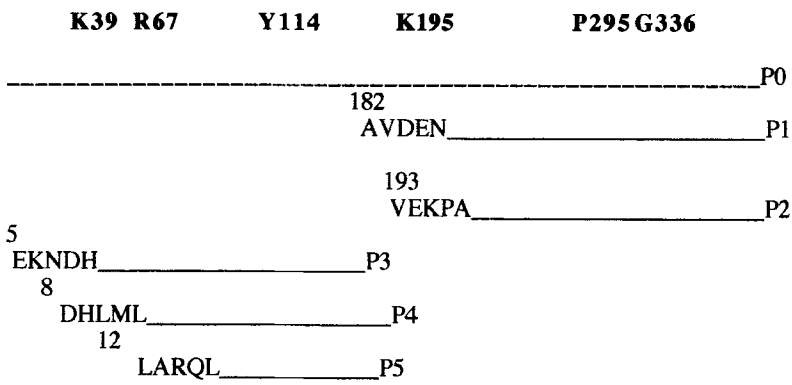


FIG. 5. Controlled digestion of the ADPGlc PPase from *E. coli* by protease K. The full segment P0 represents the sequence of the native enzyme; the amino acid residues known to be important in enzyme function, binding of the substrates, or of the allosteric ligands are shown. Protease K cleaves first the more accessible peptidic linkages 181 to 182 and 192 to 193, inactivating the enzyme and originating the peptides P1 and P2. Cleavage also occurs at the *N*-terminal side of the polypeptide at the 4-5, 8-9, and 11-12 bonds, giving rise to peptides P3, P4, and P5. When the enzyme is incubated with the protease in the presence of ADPGlc, fructose-1,6-bis-P and Mg^{++}, the internal peptide bonds are protected, and only degradation at the *N*-terminal is observed.

C-terminal of the *Anabaena* enzyme. Since this is an insertion (20 aa) that is absent in *E. coli,* and it is not predicted as buried in *Anabaena,* most likely this helix is not part of the core and is part of a loop. It is also worth noting that most of the conserved amino acids known to have roles in the binding of substrates (*E. coli* Y114, K195) and activators (*E. coli* K39, *Anabaena* K382, K419) are located in loops or are very close to loops. The residues P295 and G336 that are involved in areas important for the regulation of the *E. coli* enzyme (Preiss and Romeo, 1989, 1994; Preiss, 1996) are also in loops.

A common supersecondary structure ("motif") seen in nucleotide binding proteins in general (Rossman *et al.,* 1974) is also present in this model—that is, the glycine loop in the domain 1, which would bind the phosphates of the ATP, and the region 2, with three β-sheets and α-helices compatible with a Rossmann fold. It is likely that regions 1, 2, and 3 form a catalytic domain, composed of a typical α/β structure where the substrates bind on the top of the model as depicted in Fig. 4. The prediction of the secondary structure of the ADPGlc PPase in region 1 + 2, is identical to the accepted structure of the oncogenic protein H-Ras (p21), which is used as one of the folding models for nucleotide phosphate binding GTP (Tong *et al.,* 1991).

In region 2, the loops on the N side of the β-sheets (C end of the helices) have no amino acids conserved in all the sequences of the ADPGlc PPases

known. This is compatible with the idea that the binding of ATP is located in the other side of the α/β structure. For topological reasons, these loops would not be accessible to the substrate and, as a consequence, evolutionary pressure to conserve the amino acids in these loops is lower than in the loops located at the C end of the β-sheets.

IX. TRANSCRIPTION

Reeves *et al.* (1986) determined the levels of the wheat gliadin and ADPGlc PPase polypeptides, and of their respective mRNAs, using gliadin cDNAs and antibody to the spinach leaf ADPGlc PPase during wheat endosperm development. The mRNA contents for these proteins accumulated coordinately during endosperm development. Gliadin mRNA could be detected at 1% of the maximum level as early as 3 days after flowering (DAF). The mRNA levels for both the gliadins and ADPGlc PPase reached a maximum at about 14 DAF. Thereafter, the mRNA for the ADPGlc PPase decreased whereas the gliadin mRNAs decreased only after 18 DAF. The pyrophosphorylase enzyme increased to a maximum together with its mRNA until 14 to 18 DAF, and then decreased. In contrast, there seemed to be a delay in the expression of the gliadin proteins, and the maximum level was not reached until 31 DAF. Thus, there may be additional levels of control at the translational level since the gliadin proteins were not observed until several days after the appearance of the mRNA. Even though the mRNA levels of the ADPGlc PPase and gliadins appear to be regulated in the same manner, at the translational or posttranslational levels there may be different regulation modes for the two protein families.

The developmental pattern of the ADPGlc PPase gene was determined by Northern and dot blot hybridization analyses (Anderson *et al.*, 1991). The gene is transcribed at the highest level during early development, about 5 to 7 DAF, attaining a level of about 0.2% of the total mRNA, which declines during the later periods of seed development. This pattern of transcription is consistent with the rate of starch accumulation, which is at its highest 7 to 9 DAF (Perez *et al.*, 1975).

The developmental expression of the gene encoding the potato 50-kDa subunit was studied, and the pattern of accumulation of the corresponding mRNA closely followed ADPGlc PPase activity. Thus, the gene appears to be regulated at the transcriptional level for the wheat. It is evident that the regulation of starch synthesis during development in wheat, rice seeds, and potato tuber is similar. There is a close correlation in the activity of ADPGlc PPase and the starch synthetic rate—results that are consistent with the view that gene expression regulates the rate of starch accumulation.

Thus, regulation of ADPGlc PPase, at both the transcriptional level and by allosteric control of the enzyme, modulates the rate of ADPglucose synthesis and starch synthesis.

Northern blot analysis of mRNA isolated from potato leaf, stolon, and tuber against cDNA for the small subunit, indicated that the ADPGlc PPase gene is expressed in the tuber and leaf, but not in the stolon tissue. Thus, regulation at the transcriptional level during tuber development occurs in a tissue-specific manner (Anderson *et al.,* 1990). The size of the mRNA transcripts are 1.8 kb, both in leaf and tuber, on the basis of Northern blot hybridization, suggesting that the same gene may be expressed in both tissues. Conversely, in rice (Krishnan *et al.,* 1986) and wheat (Olive *et al.,* 1989), different genes seem to be encoding the same type of subunit expressed in different parts of the plant (leaf or endosperm). Further analysis is needed to determine if the same or different genes are expressed in different tissues, and whether the situation is different for the large and small subunits and for different species.

In this decade, several authors have reported evidence of changes in ADPGlc PPase and other starch and carbohydrate enzyme expressions brought about by increased availability of sugar (for review, see Koch, 1996). The sugar-inducible enzymes ("feast genes") can be phosphorylase (St. Pierre and Brisson, 1995), ADPGlc PPase (Müller-Röber *et al.,* 1990; Krapp and Stitt, 1995), granule-bound starch synthase, branching enzyme (Kobmann *et al.,* 1991), sucrose synthase (Sus 1) (Müller-Röber *et al.,* 1990; Karrer and Rodriguez, 1992; Koch *et al.,* 1995), invertase (Kobmann *et al.,* 1991), and sucrose-P synthase (Hesse *et al.,* 1995). Examples include those found to be repressible ("famine genes" induced by sugar starvation or depletion), the α-amylase (Karrer and Rodriguez, 1992), and another sucrose synthase isozyme, *sh* 1 (Koch *et al.,* 1992).

X. GENOMIC DNA

Treatment of the rice genomic DNA with *Eco*R1, *Bam*H 1, and *Hin*d III produced two or three bands of DNA fragments ranging from 3 to 5 kb, which hybridized in Southern blots with the rice ADPGlc PPase cDNA. Based on the cDNA copy standards run on the same gel, it was concluded that there are about three gene copies per haploid genome, and the ADPGlc PPase genes are organized in a small family that could be divided into at least two groups on the basis of the restriction fragments obtained (Krishnan *et al.,* 1986).

Using the cDNA clone for the small subunit of the rice ADPGlc PPase (Krishnan *et al.,* 1986) as a probe, the genomic DNA corresponding to the

small subunit of ADPGlc PPase was isolated (Anderson *et al.*, 1991) and its structure was determined by nucleotide sequencing. A comparison of the genomic nucleotide sequence with the isolated cDNA sequence revealed a complex gene structure with 10 exons and 9 introns in a size of about 6 kb. The exon sizes are in a range of 99 to 293 base pairs (bp) and the intron sizes range from 84 to 1435 bp. The first intron was the largest, with 1435 bp. The intron splice sites, with the exception of intron 2, contain GT/AG borders and are similar to the splice site consensus sequences (Mount, 1982; Brown, 1986). The intron-2 site did not follow the GT/AG rule (Breathnach and Chambon, 1981) but did show some similarity to the splice site consensus sequences. Thus it is possible, as suggested by Aebi *et al.* (1987), that the overall splice site sequence rather than the particular bases are necessary for correct splicing. The transcription start point is 30 bp downstream of the TATA box and the polyadenylation site was 188 bp downstream of the stop codon. The TATA or Hogness box is a nearly universal sequence, about 25 bp upstream from the transcription start site, reading TATAAAT, and is probably a site of binding for transcription factors.

In Southern blot analysis, only nuclear DNA hybridized with the tuber cDNA corresponding to the potato tuber ADPGlc PPase small subunit, indicating that the gene encoding the enzyme is localized in the nucleus. It is estimated that there are one to two gene copies per haploid genome, and digestion of the potato nuclear DNA with the restriction enzymes *Eco*R1 and *Hin*d III yielded two or three hybridizable fragments totaling 3.6 (*Eco*R1) or 6.7 kb (*Hin*d III) in size. The structure of a genomic clone encoding the analogous rice endosperm small subunit-specific gene has been determined and is almost 6.5 kb in size (Anderson *et al.*, 1991). This rice endosperm gene is interrupted by 9 introns, indicating a structure that is more complex than that of most plant genes. The estimated length of the potato tuber small subunit PPase gene suggests that it may have a complex exon/intron structure, which is a complexity also observed in two other genes involved in starch metabolism—sucrose synthase (Werr *et al.*, 1985) and the granule-bound starch synthase (from maize endosperm; Klösgen *et al.*, 1986), which have 16 and 14 exons, respectively. The multiple introns present in the ADPGlc PPase, and other genes coding for enzymes of starch metabolism, may have a role in gene expression as in the case of the alcohol dehydrogenase gene (Callis *et al.*, 1987).