2018

Regulation of Bacterial Glycogen Synthesis: Structure-Function Relationship of Adp-Glucose Pyrophosphorylase

Hiral Priyank Patel

LOYOLA UNIVERSITY CHICAGO

REGULATION OF BACTERIAL GLYCOGEN SYNTHESIS: STRUCTURE-FUNCTION RELATIONSHIP OF ADP-GLUCOSE PYROPHOSPHORYLASE

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE

SCHOOL IN CANDIDACY FOR THE

DEGREE OF DOCTOR OF

PHILOSOPHY

PROGRAM IN CHEMISTRY

BY

HIRAL P. PATEL

CHICAGO, IL

AUGUST 2018

ACKNOWLEDGEMENTS

My time in the department of Chemistry and Biochemistry at the Loyola University Chicago has been an incredible opportunity to grow both academically and personally. I would like to thank my advisor Dr. Miguel Ballicora, who gave me the opportunity to join the lab. I am grateful for having the chance to work in his lab, and to learn from someone so experienced and knowledgeable in the field. Dr. Dali Liu is the first one who introduced me to Dr. Ballicora, and recommended me to work as a volunteer. I am grateful to Dr. Dali Liu for giving me the chance to experience my first visit to Loyola University. I would also like to thank my other committee members Dr. Daniel Becker (Chemistry Department) and Dr. Brian Cannon (Physics Department), who served as highly knowledgeable members. I would also like to thank Dr. Stefan Kanzok for allowing us to work in his laboratory for the thermal shift assay. All the members of the chemistry department have been very helpful and I want to thank them for their trust and support.

To My Loving Husband: Priyank
and My Entire Family.

In Ever Living Memory of My Son Kush Patel
Gone but Never Forgotten

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVIATION

ADP-glucose	ADP-Glc
ADP-glucose pyrophosphorylase	ADP-Glc PPase
Fructose 1,6-bisphosphate	FBP
Fructose 6-phosphate	Fru6P
Pyruvate	Pyr
Glucose 1-Phosphate	Glc1P
Isopropyl-β-D-thiogalactopyranoside	IPTG
Phosphate	Pi
Pyrophosphate	PPi
Phosphatase	PPase
3-Phosphoglycerate	3-PGA
Phosphoenolpyruvate	PEP
Glucose 6-Phosphate	Glc6P
Mannose 6-Phosphate	Man6P
6-phosphogluconate	6P-gluconate

ABSTRACT

Starch is an important source for energy, and it has become a significant resource for bio- fuel production. ADP-glucose pyrophosphorylase is the enzyme that controls the synthesis of starch in plants, and glycogen in bacteria. This regulation is mainly driven by allosteric activators (Fru6P, FBP, and Pyruvate) in bacteria. It has been hypothesized that inter-subunit communications are important for the allosteric effect in this enzyme. Here we show that one specific subunit interface and the interaction between amino acids Arg11 and Asp141 are critical for the regulatory signal in the enzyme from *Agrobacterium tumefaciens*. Manipulation of this regulatory signal is critical to obtain species with high polysaccharide content used for biotechnological purposes. We have also found the activator (pyruvate) binding site of the ADP-Glc PPase from *Agrobacterium tumefaciens*. We confirmed that mutation of the residues Lys43 and Gly329 that reside near the pyruvate binding site are important for activation. The produced mutations are either insensitive to pyruvate or are hyper active. According to our analysis, the mutation also affects the binding of the pyruvate molecule. Understanding the control mechanism of this particular enzyme, can give us important insights in to regulating and engineering the enzyme for increased starch production.

CHAPTER ONE

STARCH AND GLYCOGEN SYNTHESIS AND ADP-GLUCOSE PYROPHOSPHORYLASE

Introduction

Starch and glycogen (the storage materials for plants and bacteria, respectively) are both polymers of α-D-glucose. These polymers are linked by α-1,4-linkages and branched by α-1,6-linkages. Starch differs from glycogen in that it consists of glucans that are highly ordered and densely packed [1]. Glycogen is a polymer of glucose attached by α-1,4-linkages with higher numbers of α-1,6-linkage than starch. Starch is composed of amylose and amylopectin. Amylose is made up of fewer α-1,4-linkages while the amylopectin is a more complex structure with higher numbers of α-1,6-linkages (Figure 1) [2]. Starch is an important source for biofuels, which have become a significant resource for many industrial applications. In 1800s, it was found that sugar can be made from starch via hydrolysis [3]. Starch and its derivatives are already widely used in the manufacture of paper, textiles and adhesives. Since starch is biodegradable and renewable in nature, it is being used as an environment-friendly substitute for synthetic additives that can be used in the production of plastics, detergents, pharmaceutical tablets, encapsulation of pesticides, cosmetics and even oil-drilling fluids. This increasing use of starch will require more production of starch in near future. The increased production of starch for several industrial applications generated new challenges for scientists. The characterization and production of starch variants from mutagenesis studies and transgenic technology have been critically important for the synthesis of starch granule. The knowledge gained has enabled genetic manipulation of starch biosynthetic pathway in plants.

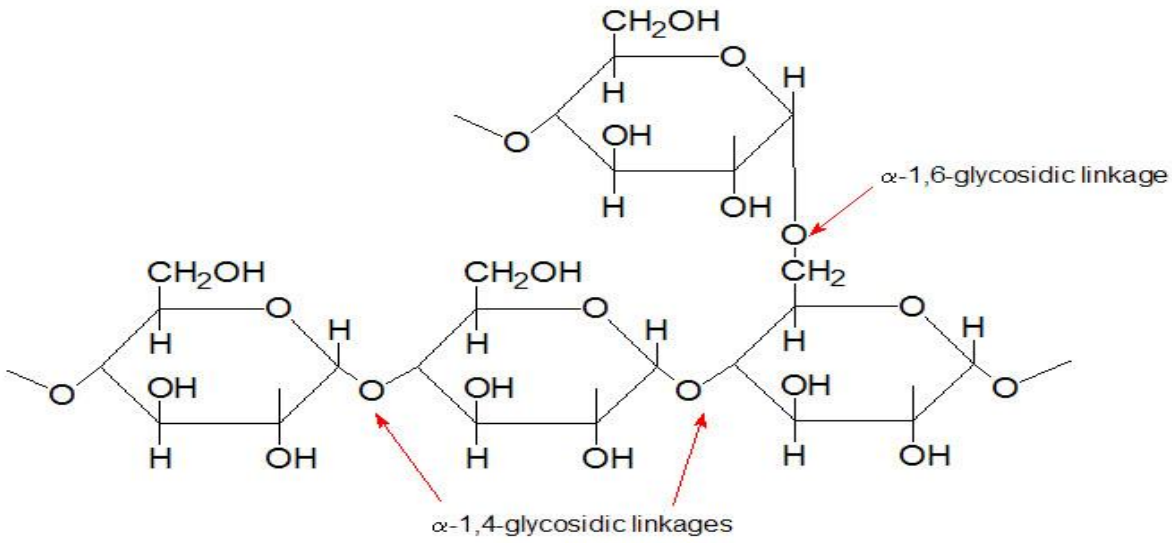


Figure 1. The structure of glycogen and starch. The structure represents the main glucose chain with α-1,4-glycosidic linkages, and the branch with the α-1,6-glycosidic linkage.

Sugar-nucleotides in the biosynthesis reaction

UDP-glucose (UDP-Glc) and ADP-glucose (ADP-Glc) are important sugar nucleotides involved in glycogen synthesis in mammals and starch synthesis in plants [4]. It was first thought that ADP-glucose supported a faster rate of synthesis than UDP-Glc, but ADP-Glc was not isolated at that time [5]. Later, ADP-Glc was isolated from ripening kernels of Zea mays [6]. From that point on, the involvement of ADP-Glc in starch synthesis was further studied in many other species. One of those studies showed that the level of ADP-Glc controls the synthesis of starch in the chloroplasts of spinach [7]. That was a breakthrough because according to previous studies, UDP-Glc was the glycosyl donor for the glycogen synthesis in mammals. Later, it was discovered that the UDP-Glc synthesis deficient *Escherichia coli* strain accumulated normal glycogen level [8]. The further findings of ADP-Glc in many other species ruled out the fact that UDP-Glc was an important precursor of glycogen synthesis and strongly suggested that ADP-Glc was the main glycosyl donor in bacteria and plants [9]. The ADP-glucose pyrophosphorylase (ADP-Glc PPase) is the first identified enzyme to be involved in the ADP-Glc synthesis [10].

Glycogen in bacteria

Glycogen is a major reserve polymer in bacteria; it has been studied in different species including Gram negative, Gram positive, archaebacteria and photosynthetic bacteria [11]. Glycogen is the polymer of glucose with α-1,4-linkages, and few branches with α-1,6-linkages. The bacterial glycogen synthesis is similar as it is in mammals; but the main difference is ADP-Glc PPase catalyzes the donation of glucose instead of the UDP-Glc PPase. The structure of the bacterial glycogen has been less studied than that of the mammalian glycogen, but there is a similarity between them [12]. The accumulation of glycogen by bacteria may give advantages during starvation periods, providing a

stored source of energy and carbon surplus [13]. Bacteria usually accumulate glycogen as a result of an excess source of carbon present at the stationary growth phase) [11]. Some exceptional bacterial species like *Rhodopseudomonas capsulata* and *Streptococcus mitis* synthesize glycogen during exponential growth [11].

Starch in higher plants

Higher plants and unicellular algae produce starch, which is similar in structure to glycogen. Starch is composed of two polyglucans: amylase and amylopectin. Amylose is mainly linear α-1,4-polyglucans, while amylopectin is also branched with α-1,6-linkages. The degree of branching is less in amylopectin than that in the glycogen [1, 14]. Accumulation of starch in the leaves occurs in the chloroplast during photosynthesis in the daylight. Starch degrades to glucose generally used for sucrose synthesis and it serves as a carbon supply for other tissues. Starch is mainly synthesized and stored in the amyloplast of the endosperm of cereals [1]. The accumulation of starch occurs during the development of the tissue, whereas the degradation of starch takes place at the time of seed germination. Plants as well as bacteria, use the ADP-Glc PPase and ADP-Glc for starch synthesis.

The reaction involved in glycogen and starch synthesis

It is generally agreed that reactions catalyzed by ADP-Glc pyrophosphorylase (ADP-Glc PPase or AGPase) (EC 2.7.7.27), starch synthase (EC 2.4.1.21), and starch-branching enzyme (SBE) (EC 2.4.1.18) are the three reactions in starch biosynthetic pathway in both photosynthetic and non-photosynthetic tissues [15]. The reactions catalyzed by those enzymes are illustrated below.

(1) ATP + Glc-1-P $\leftrightharpoons$ ADP-Glc + PPi

(2) ADP-Glc + (α- glucans) $\longrightarrow$ (α- glucans)$_{n+1}$ + ADP

(3) Linear (α- glucans)$_{n+1}$ $\longrightarrow$ Branched (α- glucans)$_{n+1}$ (glycogen or starch)

ADP-Glc PPase catalyzes the first committed reaction in starch and glycogen synthetic pathways. This enzyme catalyzes the conversion of ATP and glucose-1-phosphate (Glc-1-P) to ADP-glucose (ADP-Glc) and inorganic pyrophosphate (PPi) in the presence of Mg^{2+}. This reaction was first described in soybean [10] and it is reversible *in vitro* with an equilibrium constant near 1. The reaction becomes irreversible in the forward direction *in vivo* because of the *Le Chatelier's Principle*. The reason behind it is the inorganic pyrophosphatase, which breaks down PPi to inorganic phosphate. The ADP-Glc is further needed for further production of starch and glycogen synthesis.

The biosynthetic pathway for starch and glycogen synthesis is shown in Figure 2. As described, the α-1,4-linked glucan chains are extended via glycogen synthase, and by starch synthase in bacteria and plants, respectively. Later, the branching enzymes introduce branches with α-1,6-linked glucans to make glycogen and starch (Figure 1).

ADP-glucose pyrophosphorylase

ADP-Glc PPase catalyzes the first and rate-limiting step in starch synthesis in plant or glycogen synthesis in bacteria. This enzyme was first isolated from wheat flour in 1962 by Espada, J., et al [10]. It has been previously shown, that bacteria and plants prefer different sugar nucleotides for glycogen and for starch synthesis, respectively. In agreement to that, for the glycogen synthesis, bacteria prefer ADP-Glc as the donor for glycogen synthesis [16]. The importance of the ADP-Glc PPase was studied further in many species including maize, potato, and Arabidopsis [17-20]. The divalent metal ions were shown to be required in the in vitro reversibility [21].

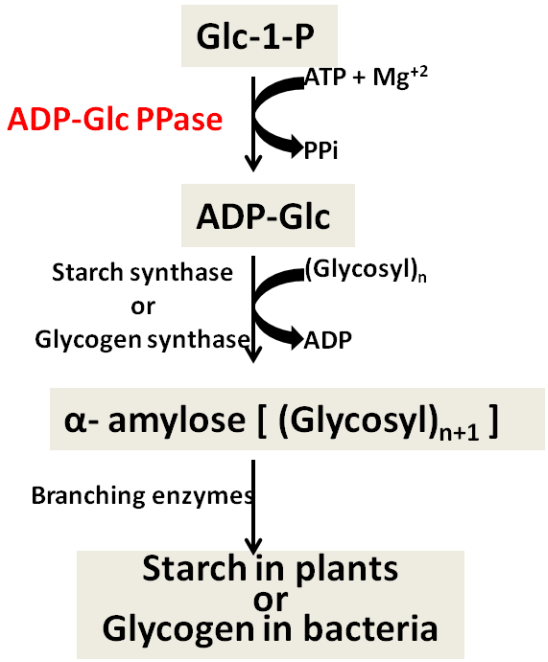


Figure 2. The model of the glycogen and starch biosynthesis pathway. The glycogen synthesis, and starch synthesis pathway are composed of mainly three enzymes. The first enzyme is ADP-Glc PPase (Highlighted), this enzyme catalyzes the first and allosterically regulated step. The second enzyme is glycogen synthase or starch synthase which depends on the glycogen synthesis in bacteria or starch synthesis in plant. The third enzyme is the branching enzyme for adding the branches with α 1,6 linkage.

ADP-Glc PPase is an allosterically regulated by sensing the metabolic intermediates that are present in that organism [11, 22-24]. The previous research has shown the activation by FBP in the *E. coli* ADP-Glc PPase [25]. From previous studies, it has been found that an *E. coli* ADP-Glc PPase mutation can affect the glycogen accumulation. The *E. coli* mutant has accumulated glycogen displayed a faster rate than the wild-type enzyme, with higher affinity toward the activator (FBP) [26]. However, a *E. coli* mutant with the glycogen deficiency had lower glycogen production with lower affinity for the activator [27]. This explains that the ADP-Glc PPase plays an important role in glycogen synthesis. Similarly, there were several studies exploring the importance of the ADP-Glc PPase in the regulation of starch synthesis in plants like maize [28], potato tuber [29, 30], *Arabidopsis thaliana* [19, 20, 31, 32], wheat endosperm [33] as well as unicellular algae [34]. Controlling the ADP-Glc PPase from the potato can control starch production, which shows that the enzyme catalyzes the key regulatory step in starch synthesis [35]. The conserved residues that are involved in triggering the activation of the ADP-Glc PPase from potato tuber have been recently studied [36]. This represents the importance of finding the conserved residues that are involved in triggering the activation of the enzyme, which can later help us understand the regulation of the ADP-Glc PPase.

Glycogen synthase in bacteria or starch synthase in plants

The second step in starch or glycogen synthesis is catalyzed by starch synthase (SS) or glycogen synthase (GS), respectively. The *glg*A gene coding for the glycogen synthase was found, and the whole nucleotide sequence was obtained early in previous studies by Kumar, A. et al. [37]. This enzyme transfers the glucose from the ADP-Glc to a growing α-1,4-linked glucan polymer. ADP-Glc is the glucose donor for the bacterial glycogen synthase or starch synthase in plants. The molecular weight of

the enzyme monomer varies from 48 to 55 kDa [38-40]. In the crystal structure of glycogen synthase from *A. tumefaciens*, glycogen synthase was found to have two Rossmann-fold domains [41]. This structural features are shared with related enzymes, such as glycogen phosphorylase and other glycosyltransferases of the GT-B super family [42]. Glycogen synthase can be classified in two different families; the first family is (GT3), which includes homologous from mammals and yeasts. This enzyme is regulated by phosphorylation, the molecular mass of a monomer is approximately 80 kDa, and uses UDP-Glc as a sugar donor [43]. On the other hand, the other family is (GT5), which includes bacteria and plants with an approximately molecular mass of 50 kDa per monomer. This GT5 family enzyme uses ADP-glucose as a sugar donor, and it is un-regulated [44].

In contrast, to the bacterial glycogen synthase, starch synthase is coded by several genes which were reported earlier [45-47]. There were two kind of starch synthase enzymes were studied; one is bound to starch granule (GBSS, granule-bound starch synthase) and can only be solubilized by α-amylase digestion of that granule. The expression of the wild-type GBSS gene introduced into amylose free (*amf*) potato by *A. tumefaciens* leads to restoration of the amylose in potato [48]. On the other hand, the second type of soluble starch synthase (SSS) can be found in the soluble fraction of the plant extract. Recent studies have shown that several critical residues determined the different activity of the SSS isoform in the rice plants [49]. The conserved residues in the *E. coli* glycogen synthases are close to a conserved motif, which was first recognized as critical for the catalysis of the enzyme [50]. The studies of starch synthase structure-function relationship are still way behind the finding of the other enzymes involved in starch biosynthesis

Branching enzymes

The last step in the biosynthesis of glycogen and starch is catalyzed by branching enzyme (EC

2.4.1.18), which is responsible for creating the branches with the α-1,6-linkages in the polysaccharides [51]. The branching can increase the solubility and decrease the osmotic strength of the polymers. Similar to the previous enzyme study for the GS and SS, there is only one branching enzyme, and only one gene (*glg*B) is present in the bacteria [52-56]. On the other hand, several genes have been found in higher plants [9]. The bacterial glycogen has shorter branches of about 10-13 glucose units, and 10% of the glucosyl bonds are α-1,6-linkages. On the other hand, the amylopectin is formed by longer branches of 20-24 glucose units, with only 5% of α-1,6-linkages [9]. The potato contains two different enzymes SBEI (Starch branching enzyme I), and SBEII (Starch branching enzyme II). Suppression of both enzymes is leads to produce starch with less than 50% of the amylose and no amylopectin [57]. There are several branching enzymes that have been studied for their importance in improving the quality of starch or glycogen. For example, the solubility of starch can be changed by suppressing the branching enzymes [58-60]. The understanding of the branching enzymes is important for starch producing industries to produce starch.

Structure of the ADP-Glc PPase

All ADP-Glc PPases consist of four subunits. Depending on the source, the structure could be a homotetramer (α_4), or a heterotetramer ($\alpha_2\beta_2$). ADP-Glc PPases in bacteria is a homotetramer (α_4) of a ~50 kDa subunits, whereas in the plant ADP-Glc PPase is a heterotetramer ($\alpha_2\beta_2$) [61, 62]. Most of the bacterial ADP-Glc PPases are homotetrameric, except the one from Firmicutes (e.g. *Bacillus subtilis*, and *Geobacillus stearothermophilus, Streptococcus mutans, and Streptomyces coelicolor*), which is heterotetramer (Table 1) [24, 63-68]. The potato tuber enzyme has a heterotetrameric structure consisting of two α or S (small) subunits and two β or L (Large) subunits [69, 70]. It has been reported that both subunits were needed for enzyme activity [71]. In plant enzymes, two different subunits

display different regulatory and kinetic properties. Previously, the small subunit has been mainly known for its catalytic role, while the large subunit is known for its regulatory role. In several different studies, they have expressed variety of ADP-Glc PPase with catalytic small subunits, and different regulatory large subunit [72-80]. In *Arabidopsis thaliana,* the large subunit was expressed with the catalytic small subunit, and found four different types of the heterotetramer with different kinetic and regulatory properties [75, 81]. The small subunit may play a more important role in regulation than in catalysis. For instance, in *Ostreococcus tauri,* the roles are reversed [82]. The variation in subunits functionality is a result of evolution is sub-functionalized. In other words, overlapping roles tend to diverge. In some cases, one of the subunits keeps a more catalytic role, and the other a more regulatory role.

The sequences for plants and for bacteria show a lot of similarities, which allows us to use the bacterial enzyme as a model enzyme for both. Bacterial enzymes have simpler oligomeric structure and easier to be produced recombinantly [79, 83-85].

Table 1. Classification of the ADP-Glc PPases based on the structure, regulatory properties, and the carbon metabolism

Organism	Carbon metabolism	Class	Activators	Inhibitors	Quaternary structure
Accumulating Glycogen Prokaryotes					
Escherichia coli *Salmonella enterica serovar* *Enterobacter aerogenes* *Typhimurium*	The Embden-Meyerhof pathway (glycolysis)	I	FBP, Pyruvate	AMP	Homotetramer (α_4)
Aeromonas formicans *Micrococcus luteus* *Mycobacterium smegmatis*		II	Fru6P	AMP, ADP	
Serratia Marcescens *Enterobacter hafniae* *Clostridium pasteurianum*		III	none	AMP	
Agrobacterium tumefaciens *Arthrobacter viscosus*	Entner-Doudoroff Pathway	IV	Pyruvate, Fru6P	AMP, ADP, Pi	
Chromatium vinosum *Rhodobacter capsulata* *Rhodomicrobium vannielii*	Entner-Doudoroff Pathway	IV	Pyruvate, Fru6P	AMP, ADP	Homotetramer (α_4)
Rhodobacter gelatinosa *Rhodobacter globiformis* *Rhodobacter sphaeroides*	Entner-Doudoroff Pathway & Glycolysis	V	Pyruvate, Fru6P, FBP	AMP, Pi	Homotetramer (α_4)
Rhodocyclus purpureus *Rhodospirillum rubrum* *Rhodospirillum tenue*	TCA cycle Reductive carboxylate cycle	VI	Pyruvate	none	
Bacillus subtilis *Geobacillus stearothermophilus*	TCA cycle during sporulation	VII	None	3-PGA	Heterotetramer (α_2 β_2)
Streptococcus mutans	TCA cycle	VIII	FBP	Pi	Heterotetramer (α_2 β_2)

Table 1 continued

Organism	Carbon metabolism	Class	Activators	Inhibitors	Quaternary structure
Ruminococcus albus	TCA cycle	IX	Pyruvate PEP	FBP	Heterotetramer (α_2 β_2)
Streptomyces coelicolor	TCA cycle	X	Glc6P Man6P	NADPH Pi	Heterotetramer (α_2 β_2)
Rhodococcus jostii	TCA cycle	XI	Glc6P Man6P Fru6P PEP	NADPH 6P-gluconate	Heterotetramer (α_2 β_2)
Cyanobacteria *Synechococcus* PCC 6301 *Synechocystis* PCC 6803 *Anabaena* PCC 7120	Oxygenic photosynthesis Fixing CO_2 through Calvin cycle	XII	3-PGA	Pi	Homotetramer (α_4)
Accumulating starch Eukaryotes					
Green algae *Chlorella fusca* *Chlorella vulgaris* *Chlamydomonas reinhardtii*	Fixing CO_2 through Calvin cycle	XII	3-PGA	Pi	Heterotetramer (α_2 β_2)
Higher Plants **Photosynthetic tissue** Leaves of spinach, wheat *Arabidopsis*, maize, rice	Fixing CO_2 through Calvin cycle	XII	3-PGA	Pi	
Non-photosynthetic tissues Potato tubers Endosperm of maize, barely, and wheat	Heterotrophic cells Metabolizing sucrose imported from photosynthetic tissues	XII XIII	3-PGA None directly, 3-PGA and F6P reverse inhibitor's effect	Pi Pi, ADP FBP	

Allosterical regulation of ADP-Glc PPase

Müller-Röber et al showed that inhibiting the ADP-Glc PPase can reduce starch production [18]. Structure-function insights of the ADP-Glc PPase can be helpful for increasing starch production in the future. The ADP-Glc PPase is allosterically regulated by key intermediates which are involved in carbon assimilation in that particular species [21, 86]. The classification of ADP-Glc PPase is shown in the Table 1, along with their source organisms, structures, and preferred effectors [24, 63-67]. All ADP-Glc PPases are allosterically regulated by the metabolites of the carbon assimilatory pathways [61, 87-89]. Organisms that use the Embden-Meyerhof (glycolysis) pathway, like *Escherichia coli* from class I, use fructose-1,6-bisphosphate (FBP) and pyruvate as activators [90]. Enzymes of class I are inhibited by AMP. The class I enzymes are homotetrameric (approximately 200 kDa molecular mass) and are products of a single gene [61, 62, 91]. The ADP-Glc PPases from organisms of class II, and III assimilate glucose also via glycolysis. These enzymes of class II are activated by FBP and Fru6P. The class III enzymes are insensitive to any known activators (Table 1). Organisms that use the Entner-Doudoroff pathway, like *Agrobacterium tumefaciens* from class IV have response to fructose-6-phosphate (Fru6P), and pyruvate (Pyr) as activators [11, 67]. Overall, the enzymes from the class I to class VI use three main activators, Pyruvate, FBP and Fru6P depending on different physiological conditions, except for the class III enzymes which are insensitive to any activators. Class V enzymes are inhibited by AMP and Pi [63]. The anaerobic bacteria, whose ADP-Glc PPases belongs to class VI, are not able to catabolize glucose but depend on pyruvate, lactate or CO_2 for growth; pyruvate is the main activator for the class VI enzymes [92]. The source organisms of the class VII enzymes utilize the TCA cycle. The structures of class VII enzymes are heterotetramers, which are products of two different genes. ADP-Glc PPases from photosynthetic eukaryotes are activated by

3-phosphoglycerate (3-PGA), and inhibited by inorganic phosphate (Pi) [93, 94]. According to recent research, the *streptococcus mutans* ADP-Glc PPase of class VIII were activated by FBP and inhibited by Pi [95]. Another Firmicutes organism (*Ruminococcus albus*) of class IX shows different regulatory properties. It is activated by phosphoenolpyruvate (PEP) and inhibited by FBP (Dr. Asención, personal communication). The *Streptomyces coelicolor* enzyme from class X is mainly activated by Man-6P and Glc-6P, while it is inhibited by NADPH and Pi [68]. The *Rhodococcus Jostii* enzymes of class XI were mainly activated by Man6P, Glc6P, Fru6P and PEP. The class XI enzymes were inhibited by NADPH and 6P-gluconate [96]. The cyanobacteria and unicellular algae enzymes of class XII were also similarly regulated by 3-PGA, and Pi [97]. The 3PGA is the product of the Calvin Cycle, which utilizes atmospheric CO_2. For the non-photosynthetic tissue, there are two classes of the enzyme, one of them is activated by 3-PGA and inhibited by Pi similar to the Cyanobacteria (and grouped as class XII) [98, 99]. The second type of enzymes shows different regulatory properties, and grouped in class XIII (Table 1). The class XIII enzyme shows poor activation by activators [100, 101]. The enzymes from the potato tuber (class XII) can be regulated by a redox mechanism [102]. The classification of the ADP-Glc PPases from various organisms, indicates a great diversity in allosterism as a result of evolution.

Agrobacterium tumefaciens used for generating transgenic plants

A. tumefaciens has the ability to transfer a particular DNA segment (T-DNA) of the tumor-inducing (Ti) plasmid into the nucleus of infected cells, with the further integration into the host genome to produce crown gall disease [103]. Plant transformation mediated by *A. tumefaciens*, a soil plant pathogenic bacterium, has become the most used method for the introduction of foreign genes into plant cells and the subsequent generation of transgenic plants [104]. The transformation of potato

using *A. tumefaciens* has been well developed, making it easy to study effects of individual genes on starch biosynthesis and starch characteristics. Together with the better understanding of the enzymes involved in starch biosynthesis, the altered genes that encode for these enzymes have been transformed into plants using *A. tumefaciens* to modify starch metabolism [105]. Our research is focused on exploring the structure and function relationship of the enzyme ADP-Glc PPase from the *A. tumefaciens*, to provide insights for future enzyme engineering that can lead to increased starch production.

ADP-Glc PPase from *A. tumefaciens*

The ADP-Glc PPase of *A. tumefaciens* is a homotetramer, which can be represented as a dimer of dimers. The structure has been previously studied by Cupp-Vickery (PDB: 3BRK), in which they compared the structure with the potato tuber crystal structure [83]. In each subunit, the N-terminal domain is mainly involved in catalysis while the C-terminal domain is mainly involved in the regulation. Studies of ADP-Glc PPase have shown that the C-terminal domain is generally important for determining allosteric effector specificity in glycogen synthesis [106]. The activator binding sites have been studied in other species, like potato tuber [70] and *E. coli* [107]. We have recently solved the structure of the ADP-Glc PPase with the pyruvate bound. The homotetramer structure and the single subunit of that enzyme is shown in Figure 3 [108]. The structure guided us to further investigate the key residues in the activator binding site.

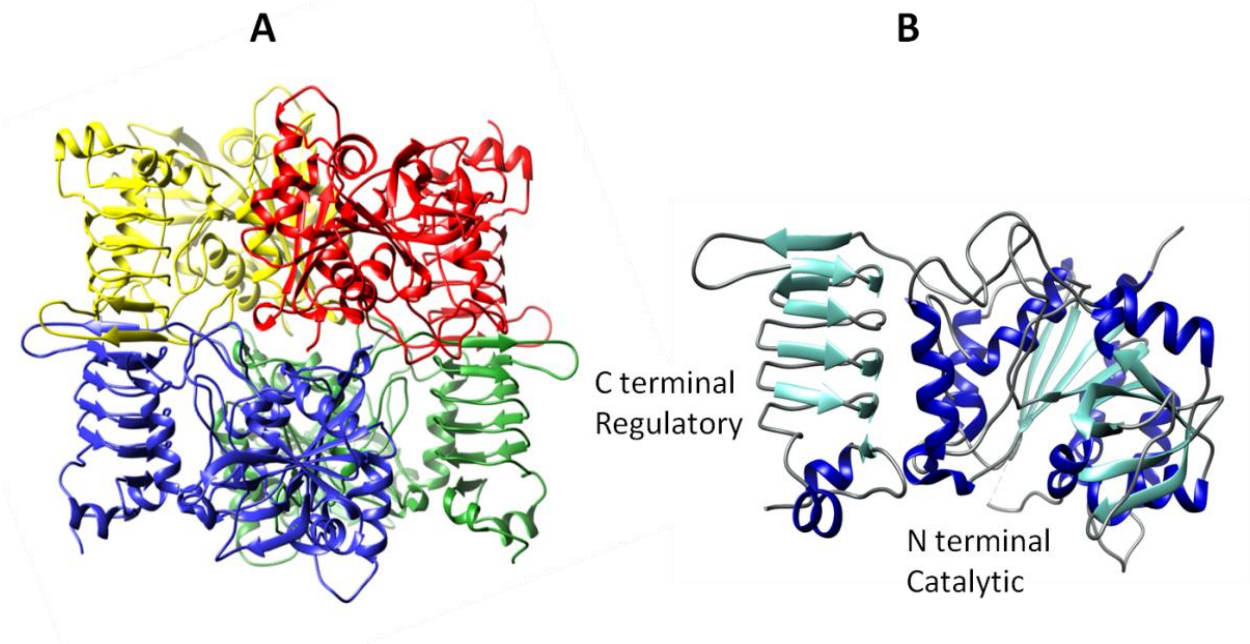


Figure 3. The crystal structure of the ADP-Glc PPase from *A. tumefaciens*. Figure (A) displays the homotetramer structure (α_4) of the ADP-Glc PPase while a single subunit is shown in (B). In the single subunit, the β-sheets are colored in cyan and the α-helixes are colored navy blue.

Our research here provides more details about the residues that are involved in binding the activator and the residues that are involved in the interface-inter subunit interaction.

Residues involved in activator binding and regulation

N-Terminus residues of the *A. tumefaciens* ADP-Glc PPase. *A. tumefaciens* is the bacterium whose ADP-Glc PPase is allosterically activated by Fru6P and pyruvate, whereas it is inhibited by AMP and Pi. According to previous studies, the first 6o amino acids of the N-terminus were important in the allosteric activation. The residues Arg5 and Arg11 are important for pyruvate activation [109]. The Arg11 to alanine mutation showed similar apparent affinity for ATP in the presence of Fru6P to that of the wild-type. On the other hand, mutations in Arg5 and Arg11 positions led to enzyme forms unable to be activated by Fru6P as much as the WT enzyme, however they displayed similar affinities towards activator Fru6P compared to the wild-type enzyme. According to the same study, Arg33 and Arg45 residues were important for the activation by Fru6P. Mutations of those residues showed higher V_{max} compared to the wild-type enzyme, but they were insensitive to further activation. The mutation of Arg22 to alanine showed lowered affinity towards the substrate (ATP) and one of the activators (Fru6P) [109].

C-Terminus residues of the *A. tumefaciens* ADP-Glc PPase. Not all residues that are critical for regulation are in the N-terminus. Some of those are in C-terminal domain but facing the proposed regulatory regions on the N-terminus. For instance, according to structural analysis, the His379 residue of *A. tumefaciens* ADP-Glc PPase was found to be important for the allosteric regulation [85]. In *E. coli*, the homologue residue of His379 is an Arg. The mutation of His379 to Arg increased also, this mutant enzyme could be activated by Fru6P. The similar mutation of His379 to Lys allowed the enzyme to be activated by both activators FBP and Fru6P. Both mutant enzymes had higher affinity

toward FBP [85].

***E. coli* ADP-Glc PPase regions involved in regulation**. The ADP-Glc PPase from *E. coli* is activated by FBP and inhibited by AMP. Recently Asención Diez et al [90] showed that *E. coli* ADP-Glc PPase can be also activated by pyruvate. The Lys39 of the *E. coli* ADP-Glc PPase was found to be important for allosteric regulation [110]. The mutation Lys39 to glutamate, showed decrease in the apparent affinity for the substrate (ATP) in the presence of FBP. This mutant enzyme was not sensitive to inhibition by Pi, and had a lower apparent affinity for the FBP. The random pentapeptide insertion between the identified regions of *E. coli* ADP-Glc PPase (such as Ins3(IDCLN) insertion between Val383 and Ile 384, and Ins65 (SCLNS) insertion between Arg375 and Ser376) showed lower affinity for the FBP. On the other hand, the Ins8(FKHLL) (Insertion between Leu102 and Pro103) showed no activation of the enzyme. The three regions described above are important in the allosteric regulation of the *E. coli* ADP-Glc PPase.

Residues involved in regulation of Cyanobacteria. Mutation of the Lys382 residue in *Anabaena* ADP-Glc PPase changed the apparent affinity towards 3-PGA, an allosteric activator [111]. According to the same reference, the charge and the size of the residue at position 382 manipulates the binding of 3-PGA. This gives us insights into the importance of charge and size of the amino acids that are involved in the activator binding.

Plant ADP-Glc PPase residues in regulation. According to previous studies of potato tuber ADP-Glc PPase, alanine mutations of Lys404 and Lys441 on the small subunit decreased the apparent affinity for the activator, 3-PGA [112]. Like other plant enzymes, the maize enzyme is also activated by 3-PGA (activator for class XII enzyme). The mutation in maize ADP-Glc PPase large subunit R104A and small subunit R107A switched the affinity for the allosteric activator (3-PGA) [113]. In addition,

both the mutant enzymes showed decreased affinity for the 3-PGA and increased binding of Fru6P. This suggested that a single mutation can switch the activator selectivity.

Allosteric mutants (pre-activated). In the previous studies of Meyer, C. R., et al, 1998, mutation of Gly336 residue to Aspartate in *E. coli* ADP-Glc PPase produced a pre-activated enzyme [114]. The introduction of G336D mutated gene from *E. coli* into the plant increased starch production [30]. The mutations of P295 residues (P295D and P295E) exhibited elevated enzyme activity in the absence of activator, and lowest sensitivity to the AMP compared to the wild-type [115]. The double mutant (P295D-G336D) enzyme of *E. coli* has higher activity in the absence of FBP and decreased sensitivity for AMP inhibition compared to the single mutations [115]. At that time, it had been predicted that the substitution at G336 position to a negatively charged group would interact with the positively charged groups that comprise the part of the activator binding site (mimicking the activator binding) [115]. More structural knowledge of the residues involved in the activator binding directly or indirectly can provide important information for the regulation of glycogen synthesis in bacteria.

Inter-subunit interaction in ADP-Glc PPase

Potato tuber ADP-Glc PPase small subunit Inter-subunit interaction. The Cys12 residue of potato tuber small subunit participates in an intermolecular disulfide bridge that, when reduced by DTT, yield higher activation of the ADP-Glc PPase [102]. The interaction between the monomer with the disulfide bridge of Cys12 that contribute the tetrameric form of the enzyme is shown in the Figure 4 [70]. Also in potato tuber enzyme, the intermolecular disulfide bridge located between two Cys12 of from different subunits is essential for the stability at 60°C, after the cleavage of disulfide bond the enzyme is stable up to only 40°C [116]. Furthermore, the residue Cys12 participates in the redox regulation of the potato tuber ADP-Glc PPase [29].

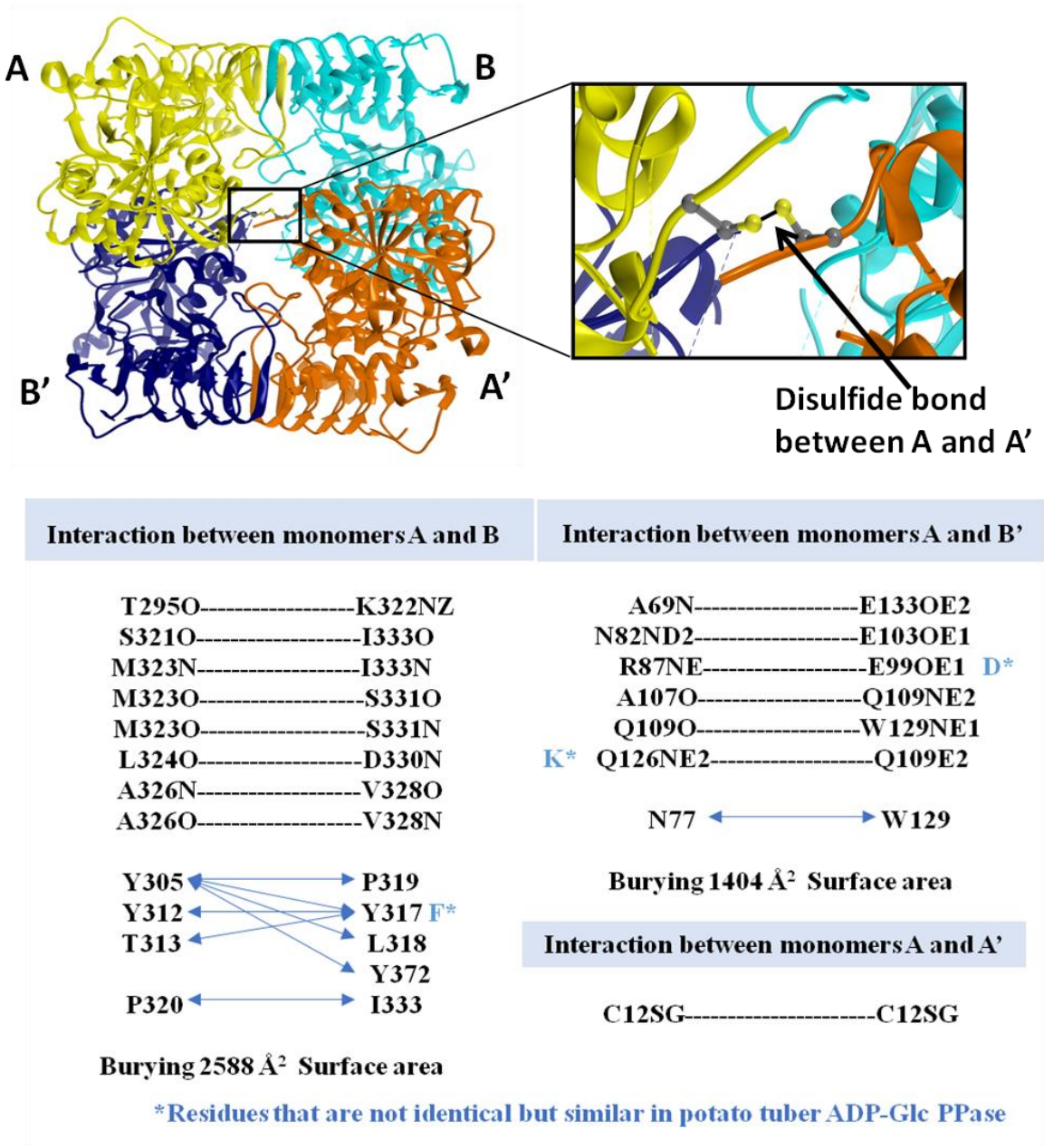


Figure 4. Inter-subunit interactions of the homotetrameric potato tuber crystal structure (PDB: 1YP3). The residues that make the inter-subunit interactions were indicated at the bottom of the figure.

It was shown that adding a cysteine residue at the N-terminus increased the heat-stability of the maize endosperm ADP-Glc PPase [117]. The N-terminal motif (QTCL motif) is conserved in heat stable small subunit of ADP-Glc PPase; introduction of this motif to maize endosperm enzyme increased the heat stability [118]. In author words, the author introduced the disulfide bridge to the maize endosperm ADP-Glc PPase modeling after potato tuber enzyme, resulting in an increase in heat stability of the maize enzyme. Not only did this motif affect the heat stability of the enzyme, but also doubled the k_{cat} for the enzyme with increased affinity for 3 PGA.

Subunit interaction in allosteric regulation. In the previous studies of Greene et al. 1996, randomly mutagenized potato large subunit using hydroxylamine was co-expressed with the small subunit to determine the residues involved in the allosteric effector binding site. As a result, they found Pro52 to leucine mutation showed a 45-fold decrease in the apparent affinity for the 3-PGA compared to wild-type leading to a decrease in the glycogen production [119]. To reverse the down regulatory properties of the potato large subunit mutation (P52L), they introduced a chemically-modified small subunit and co-expressed it with the P55L mutated large subunit [120]. As a result, two mutations in the small subunit (L46F and F112L) co-expressed with P52L mutated large subunit was found to increase the affinity for the 3-PGA. However, the same two small subunit mutations, when co-expressed with the wild-type large subunit, did not show a similar effect. In the same study, the mutations P308L and R350K in the potato small subunit co-expressed with the wild-type large subunit increased the affinity for the 3-PGA. A similar effect was not observed when the mutated small subunits were co-expressed with the P52L mutated large subunit [120]. For comparison, it was necessary to express the mutated small subunits as a homotetrameric form. The expressed homotetramer enzymes with the mutant small subunits were unable to produce glycogen in *E. coli*. The

above results indicate the importance of the interactions between the subunit for the regulatory properties of the ADP-Glc PPase.

Hybrid ADP-Glc PPase. The ADP-Glc PPase subunits from different species (*Arabidopsis thaliana* and potato tuber) can be co-expressed, and produce an active hybrid enzyme with different regulatory properties [121], showing various responses to effectors depending on which large (L) subunit was involved. One hybrid enzyme StuS/APL1 (*Solanum tuberosum* small subunit/*Arabidopsis thaliana* large subunit 1) showed an increased affinity for 3-PGA relative to all other hybrid enzymes, as well as the wild-type *Arabidopsis* ADP-Glc PPase. In the studies of Iglesias, et al, the hybrid enzyme was constructed using cyanobacteria (*Anabaena*) and potato tuber (*Solanum tuberosum*) [122]. The N-terminus and C-terminus domain swapping between Cyanobacteria (homotetramer) and potato tuber (heterotetramer) small subunit created a functional chimeric ADP-Glc PPase [122]. The same study suggested that the small subunits were evolved from the Cyanobacteria ADP-Glc PPase. This research suggests that the understanding of subunit interactions can provide greater insights in designing a hybrid ADP-Glc PPase with specific regulatory properties.

According to other studies, the mutation at the interface can cause changes in the regulation of the maize enzyme, which is also related to the synergy between the large and the small subunits [123, 124]. Previously, the authors have studied different inter-subunit interface interactions, like head-to head, and tail to tail interactions of maize endosperm, which are important for the allosteric properties and affinity for 3-PGA and Pi [125].

Objectives and outline of the thesis

ADP-Glc PPase catalyzes the first and regulatory step in starch and glycogen biosynthesis. The increased use of starch industrial applications generated a new driving force for the needs to increase the production of starch. The potential of genetic engineering centered on ADP-Glc PPase for elevated starch production has been increasingly recognized. Therefore, the main purpose of these studies is to provide important information about the enzyme ADP-Glc PPase from *A. tumefaciens.* Our focus is to understand the mechanism by which the allosteric regulation occurs, by the effect of residues that are in the regulatory sites or those involved in the inter-subunit interface interactions. The layout for the rest thesis is as follows.

Chapter 2: review of the allosterically important residues like Lys 43 (involved in binding pyruvate) and Gly329. The study is followed by the site-directed mutagenesis of K43 residue with three different amino acids, and the kinetic characterization. It also includes solving the crystal structure of the K43A mutant ADP-Glc PPase.

Chapter 3: An analysis of the residues including Arg11 and Asp141, which are involved in the interface inter-subunit interaction, using mutagenesis and kinetic studies. The analytical gel filtration studies provided the insights of altered oligomerization state to some mutant enzymes. The results are discussed thoroughly for implications of the finding.

Chapter 4: Conclusions and future implications.

CHAPTER TWO

A CRITICAL AND SPECIFIC RESIDUE FOR THE PYRUVATE ALLOSTERIC EFFECT IN ADP-GLUCOSE PYROPHOSPHORYLASE FROM *A. TUMEFACIENS*

Abstract

ADP-glucose pyrophosphorylase (ADP-Glc PPase) catalyzes the key regulatory step in bacterial glycogen and plant starch biosynthesis. Based on previous studies the loop region consists of residues from Arg25 to Arg45 are involved in regulation of the enzyme. From this conserved loop, of particular interest is K43. In the crystal structure of the enzyme, K43 faces the interface between the C and N terminal domains and has been proposed to be critical for allosteric regulation. The G336D mutant of the *Escherichia coli* ADP-Glc PPase exists in a pre-activated state [114]. When the *E. coli* G336D enzyme is expressed in potato tuber, it resulted in significantly elevated starch production [30]. We characterized the homologous residue G329D in *A. tumefaciens* ADP-Glc PPase, which we found pre-activated in similar manner [126]. Different mutants of K43 have been characterized, including K43A, K43N, K43R, and K43A/G329D double mutant. The K43A and K43N mutants are insensitive to pyruvate, whereas K43R regained partial activation by pyruvate. Using the thermal shift assay, we found out that, K43A, K43N, G329D, and the double mutant K43A/G329D bind to Fru6P but lost the ability to bind pyruvate. For that reason, we conclude that the most important chemical feature of position 43 to retain regulatory function is the positive charge. Our results show that K43 is an important residue of ADP-Glc PPase from *A. tumefaciens*; it not only is involved in the binding of pyruvate, but also in triggering the activation upon binding.

Introduction

The main regulatory step of starch metabolic pathways is the first committed step catalyzed by ADP-Glc Pyrophosphorylase (ADP-Glc PPase). ADP-Glc PPase is an allosterically regulated enzyme, which catalyzes the conversion of ATP and glucose-1-phsphate (Glc-1-P) to ADP-glucose (ADP-Glc) and inorganic pyrophosphate (PP_i) in the presence of Mg^{2+}; The reaction catalyzed by ADP-Glc PPase is shown below.

ATP + Glc-1-phosphate $\rightleftharpoons$ ADP-Glc + inorganic pyrophosphate

Studies on chimeric ADP-Glc PPase have shown that the C-terminal domain is generally important for determining allosteric effector specificity in glycogen synthesis [106]. In most organisms, this enzyme is allosterically regulated by intermediates of the main carbon assimilatory pathway that exist within host species. ADP-Glc PPase in *Agrobacterium tumefaciens (A. tumefaciens)* is a homotetramer (α_4), with a molecular mass of around 200 kDa, meaning that each subunit is around 50 kDa. This enzyme is activated by Fructose 6-phosphate (Fru6P) as well as Pyruvate, and is inhibited by AMP or ADP.

According to previous studies, Lys39 in the *E. coli* enzyme interacts with the allosteric activator (FBP) [110], and it is very possible that the regulation is determined by a combined arrangement between the N and C terminal domains [61]. Based on the crystal structure we recently solved (PDB: 5W5R) [126], another residue, Lys43, is expected to be important for allosteric activation of the enzyme in *A. tumefaciens*, as a part of the pyruvate binding site (Figure 5). Since, Lys43 is a semi-conserved residue of ADP-Glc PPase among different species. In addition, K43 was found in a close contact with Pyruvate (PDB: 5W5R) [126].

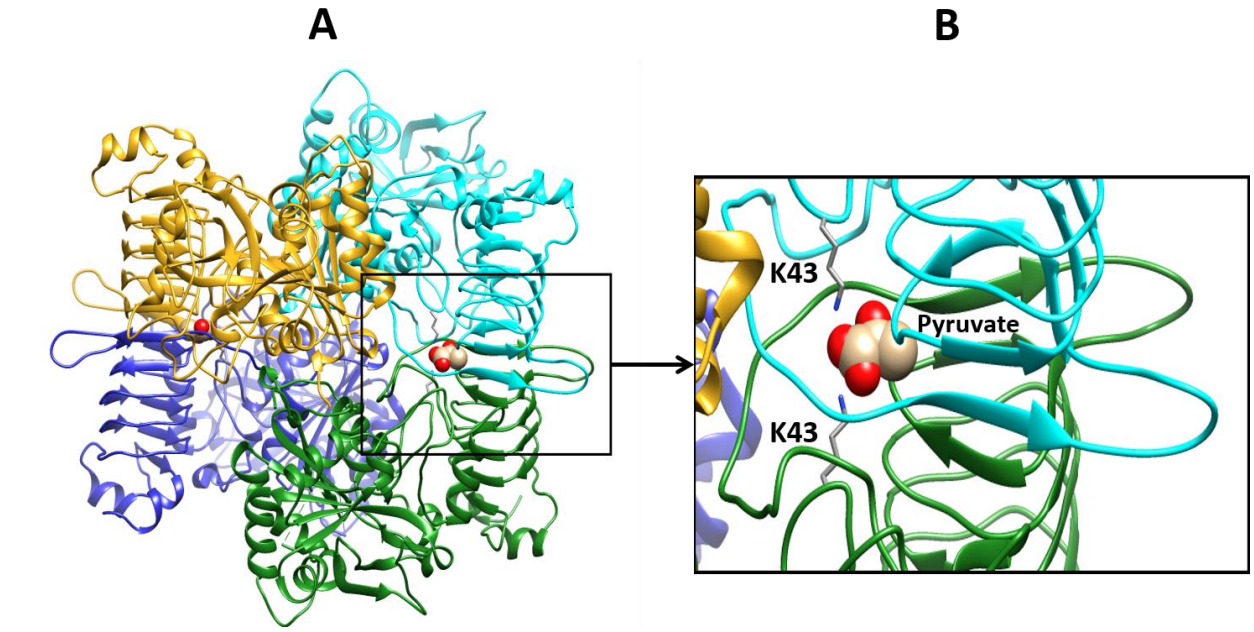


Figure 5. Crystallographic structure of *A. tumefaciens* ADP-Glc PPase with pyruvate-bound (PDB: 5W5R). There are two pyruvate molecules bound to the homotetramer, the pyruvate is shown with carbon colored sandy brown, and oxygens colored red (A). The close-up of the highlighted area represents two Lys43 residues from two neighboring subunits in a close contact with pyruvate (B).

As we know, this enzyme is a homotetramer, but we can discuss it as a dimer of dimers for convenience. In one dimer, two K43 residues from two different subunits are making close contact with the pyruvate. K43 of the pyruvate bound from one subunit is making a hydrogen bond with the carboxylate oxygen, whereas another lysine from the other subunit is forming a hydrogen bond with the ketone oxygen of the pyruvate. The expression of G336D mutant *E. coli* ADP-Glc PPase gene in potato tuber resulted in an increased starch production [30, 114]. We characterized the homologous mutant G329D in *A. tumefaciens* ADP-Glc PPase [126]. Pyruvate was also in the proximity of two G329 residues from two subunits of a dimer through a π- π interaction. Based on the *in silico* model of G329D, it looks like the negatively charged carboxylate of Asp mimics the presence of pyruvate (Figure 6) (the model was built using the UCSF chimera [127]).

In the sequence alignment of the several plants and bacteria, we found arginine, and asparagine at the conserved position 43. We made a series of mutants including K43A, K43N, K43R, and K43A/G329D to probe the significance of K43 in pyruvate activation. We also solved the crystal structure for the K43A mutant along with the kinetic analysis of the mutant. We used the thermal shift assay to investigate the activator binding to the mutants. All this provided important insights in the process of binding and triggering activation by pyruvate.

Results

The structural comparison of wild-type and the mutant K43A ADP-Glc PPases of *A. tumefaciens*.

Previously, we showed that two pyruvate molecules bind to the homotetrameric structure of the P96A mutant of the ADP-Glc PPase of *A. tumefaciens* (PDB: 5W5R). In the structure, K43 residues from two neighboring subunits interact with one pyruvate located in the subunit interface.

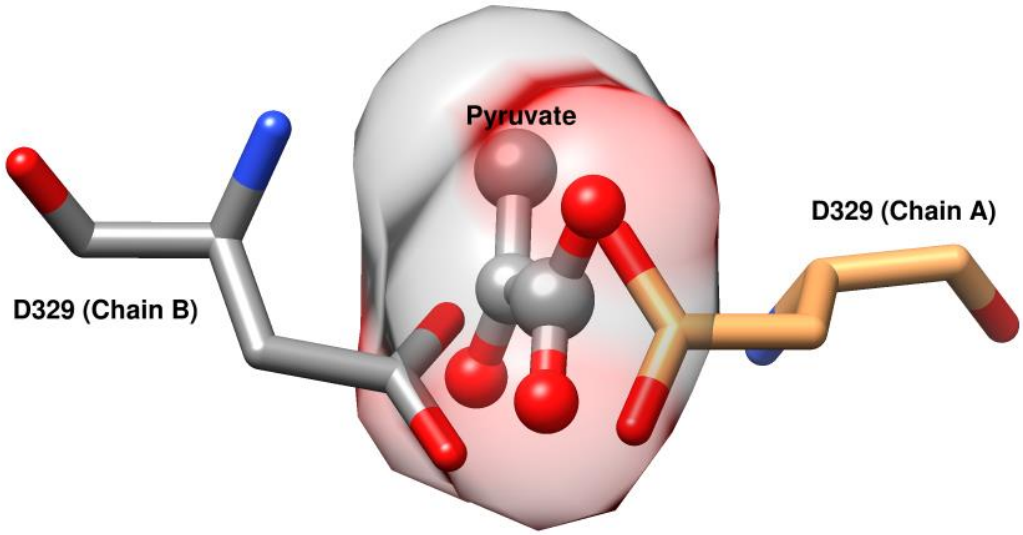


Figure 6. The Aspartate at 329 mimics the oxygens of the pyruvate. The model of the Asp329 in *A. tumefaciens* ADP-Glc PPase with pyruvate bound. The model was produced using the program UCSF chimera.

The K43A mutant enzyme was shown to present a significant effect on regulation [126]. The overall structure of the K43A mutant ADP-Glc PPase is very similar to the wild-type (PDB: 5W6J). When both structures are structurally aligned, the RMSD was only 0.47 Å. Similarly, the structural alignment of the Apo wild-type (PDB: 5W6J) with the ethyl pyruvate-bound structure (PDB: 5W5J) had an RMSD value of only 0.53 Å. The structural alignment of the backbone structure of wild-type with the ethyl pyruvate-bound and K43A mutant gave an RMSD value of only 0.49 Å (Table 2).

Effect of activators (Fru6P, Pyruvate) on the saturation curves of ATP

In the absence of activators, the previously mentioned mutant K43A was found to be partially pre-activated with a V_{max} of 33.2 U/mg compared to wild-type. In the absence of activators, the K43N mutant showed a similar V_{max} and ATP apparent affinity compared to those of wild-type. On the other hand, a mutation that conserved the charge (K43R) showed a 3.7-fold decrease in V_{max} compared to wild-type, but a similar $S_{0.5}$ for ATP (Table 3), also in the absence of the activators.

The presence of pyruvate showed no significant effect on the activation for K43A and K43N mutant ADP-Glc PPase of *A. tumefaciens*. Compared to the data in the absence of activators, the conserved K43R mutation showed an increase in the V_{max} (17.5 U/mg) but a similar $S_{0.5}$ for ATP. K43A and K43N enzyme do retain sensitivity towards Fru6P with a V_{max} of 47.8 U/mg, and 17.0 U/mg, respectively. These two mutants have a similar apparent affinity for ATP compared to wild-type. Similarly, the K43R mutant enzyme increases the V_{max} 4.1-fold compared to the absence of activators, with a similar $S_{0.5}$ as wild-type.

Effect on the activation by Fru6P and Pyruvate (Figure 7)

At a saturating concentration of ATP, Fru6P showed a similar activation for the K43R mutant with a decreased $A_{0.5}$ for Fru6P (0.09 mM) (Table 4).

Table 2: Structural comparison between the wild-type and K43A mutant ADP-Glc PPase.

Comparison[a]	RMSD (Å)[b]
K43A vs. WT	0.47
WT vs. WT_Ethyl Pyruvate[c]	0.53
WT_Ethyl Pyruvate[c] vs. K43A	0.49

[a] The crystal structures of wild-type and K43A ADP-Glc PPase was compared using PDB Viewer.
[b] The RMSD values for the backbone were calculated using the PDB Viewer.
[c] WT_Ethyl Pyruvate represents the wild-type crystal structure of ADP-Glc PPase with Ethyl Pyruvate-bound to it.

Table 3. Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants

Enzyme [a]	Control			+ 1.5mM Pyruvate			+ 1.5mM Fru6P		
	$S_{0.5}$ [b] (ATP)	n_H [b]	*Vmax*	$S_{0.5}$ [b] (ATP)	n_H [b]	*Vmax*	$S_{0.5}$ [b] (ATP)	n_H [b]	*Vmax*
	mM		U/mg	mM		U/mg	mM		U/mg
WT	0.21 ± 0.01	1.9 ± 0.2	11.18 ± 0.30	0.13 ± 0.01	2.2 ± 0.2	85.44 ± 1.91	0.08 ± 0.01	1.7 ± 0.1	123.5 ± 2.4
K43A[b]	0.36 ± 0.04	2.3 ± 0.4	33.23 ± 1.37	0.33 ± 0.06	1.7 ± 0.4	30.49 ± 1.86	0.13 ± 0.02	1.1 ± 0.2	47.79 ± 1.67
K43A/G329D	1.03 ± 0.01	3.5 ± 0.1	9.97 ± 0.07	0.76± 0.02	5.3 ± 0.5	9.71 ± 0.11	0.34 ± 0.04	1.5 ± 0.3	9.00 ± 0.30
K43N	0.16 ± 0.02	2.7 ± 0.7	12.07 ± 0.44	0.18 ± 0.03	1.8 ± 0.5	11.77 ± 0.71	0.04 ± 0.01	1.3 ± 0.6	16.96 ± 1.15
K43R	0.22 ± 0.02	0.8 ± 0.1	3.01 ± 0.07	0.18 ± 0.05	2.1 ± 1.0	17.51 ± 1.37	0.06 ± 0.01	1.1 ± 0.1	30.05 ± 0.50

[a] Assays were performed as described in substrate saturation assay, as stated under Materials and Methods.

[b] The $S_{0.5}$ and *nH were* calculated using the Hill equation.

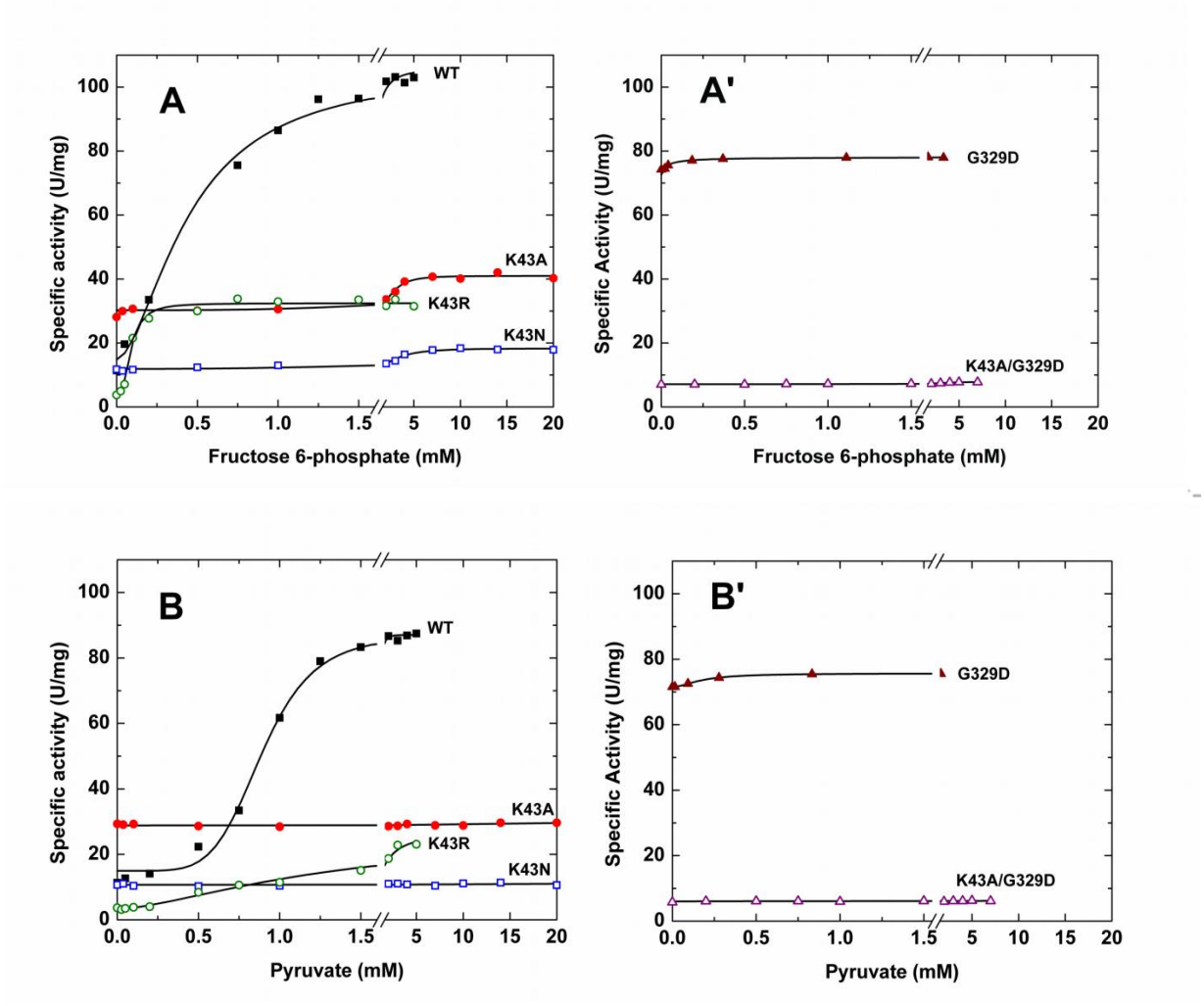


Figure 7. Activator saturation curves of wild type and mutant ADP-Glc PPase from *A. tumefaciens* at saturating concentrations of ATP. The effects of Fru6P (A, and A') and Pyruvate (B, and B') were assayed on the wild type (WT), and K43A, K43N, K43R, G329D and K43A/G329D mutant ADP-Glc PPases. In the figure, A represents the Fru6P saturation curve for wild-type, K43A, K43N, and K43R, while A' represents the G329D, and double mutant K43A/G329D. In the similar manner Figure B represents the pyruvate saturation curve for wild-type (WT), and K43A, K43N, and K43R, while B' represents theG329D, and the double mutant K43A/G329D. The assays were performed as described in Materials and Methods in the presence of 1.5 mM ATP.

Table 4. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.

Enzyme [a]	Fru6P				Pyruvate			
	$A_{0.5}$	n_H	*Vmax*	Activation[b] (V_{max}/V_0) (-fold)	$A_{0.5}$	n_H	*Vmax*	Activation[b] (V_{max}/V_0) (-fold)
	mM		U/mg		mM		U/mg	
WT	0.34 ± 0.07	2.1 ± 0.5	105.2 ± 3.9	9.01	0.45 ± 0.04	2.1 ± 0.4	87.4 ± 2.0	8.53
K43A[c]	2.43 ± 0.23	3.5 ± 1.1	40.45 ± 0.51	1.34	N/A[c]	N/A[c]	28.84 ± 1.21	0.99
K43A/G329D	3.11 ± 0.70	2.3 ± 0.9	7.90 ± 0.18	1.22	N/A[c]	N/A[c]	6.37 ± 1.06	1.10
K43N	3.95 ± 1.52	2.6 ± 1.0	21.28 ± 4.64	1.88	N/A[c]	N/A[c]	12.24 ± 1.22	1.02
K43R	0.09 ± 0.01	2.5 ± 0.6	32.41 ± 0.61	10.88	1.33 ± 0.20	1.5 ± 0.3	26.59 ± 2.16	8.16

[a] Assays were performed as described in activator saturation assay, as stated under Materials and Methods.
[b] Activation fold is calculated by dividing the maximum velocity (V_{max}) by the velocity in the absence of activator (V_0) (V_{max}/V_0).
[c] No significant activation was observed to calculate activation parameters.

On the other hand, K43A, and K43N showed 1.3- and 1.9-fold activation with 7.1- and 11.1-fold increases in the $A_{0.5}$ value for Fru6P, respectively. The charge conservative mutation K43R displays similar activation fold for pyruvate and a 3-fold increase in the $A_{0.5}$ value. In this case the V_{max} was 26.6 U/mg compared to that of the wild-type V_{max} of 87.4 U/mg. On the other hand, K43A, and K43N mutant enzymes were completely insensitive to pyruvate, with an activation fold of 0.99, and 1.02 (and a V_{max} of 28.8 U/mg, and 12.24 U/mg), respectively. However, the $A_{0.5}$ for pyruvate cannot be calculated to the data for K43A, and K43N of *A. tumefaciens* ADP-Glc PPase since there is no significant activation.

Binding analysis by Thermal shift assay

The binding of ligands can increase the stability of an enzyme, shifting melting temperature. We analyzed the capability of the *A. tumefaciens* ADP-Glc PPase and mutants to bind activators by a thermal shift assay. According to our analysis, the conservative mutation K43R behaves like the wild-type (Figure 8, and Table 5). The mutant enzymes K43A and K43N showed an increase in melting temperature of 8.9°C, and 8.1°C in the presence of Fru6P, respectively. On the other hand, in the presence of pyruvate these mutants do not show any change in the melting temperature. The K43A mutant displayed two peaks in the thermal stability assay. It was not confirmed at which temperature the K43A mutant enzyme completely melts. To confirm the final melting temperature for the K43A mutant we did heat stability assays. According to that, the K43A mutant melting temperature is around 50°C in the absence of activator (Figure 9).

Positive charge is crucial for binding pyruvate, and signaling.

The G329D mutant enzyme was pre-activated (high activity in the absence of activators), but it is insensitive to further activation by pyruvate (Figure 7) [126].

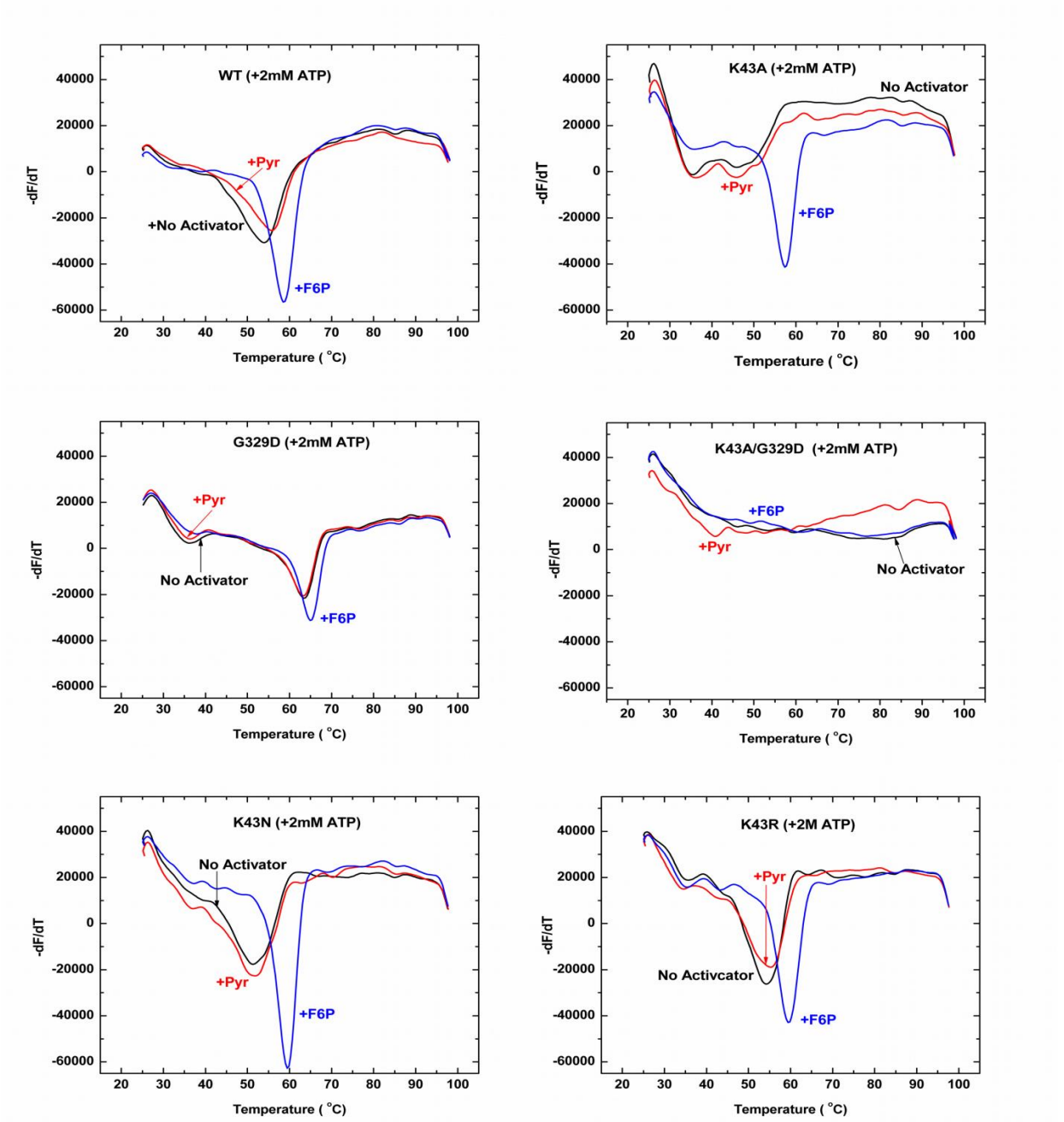


Figure 8: Thermal shift analysis of wild-type and mutant ADP-Glc PPase. The effect of temperature on the stability of the enzyme was assayed in the presence and absence of the activator for the wild-type (WT), and mutants K43A, K43N, K43R, G329D, and K43A/G329D ADP-Glc PPases. The thermal shift assays were performed as described in the materials and methods. The black line represents the absence of the activators, red line represents the presence of pyruvate, and the blue line represents the presence of the Fru6P

Table 5. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.

Enzyme [a]	Substrate (ATP) (0mM)			Substrate (ATP) (2mM)		
	T_m (°C) [b]			T_m (°C) [b]		
	No Activator	+Pyr	+Fru6P	No Activator	+Pyr	+Fru6P
WT	56.0 ± 0.2	58.3 ± 0.9	57.8 ± 0.4	54.0 ± 0.2	55.8 ± 0.7	58.7 ± 0.2
K43A	49.6 ± 0.5	50.1 ± 1.5	52.2 ± 0.2	48.6 ± 1.3	46.0 ± 0.9	57.5 ± 0.1
G329D	63.4 ± 0.1	63.3 ± 0.1	63.8 ± 0.1	63.3 ± 0.1	63.3 ± 0.1	64.9 ± 0.1
K43A/G329D	N/A[c]	N/A[c]	N/A[c]	N/A[c]	N/A[c]	N/A[c]
K43N	54.1 ± 0.5	54.5 ± 0.4	55.9 ± 0.1	51.5 ± 0.3	51.7 ± 0.8	59.6 ± 0.1
K43R	53.7 ± 0.5	54.8 ± 0.1	56.6 ± 0.2	54.5 ± 0.3	55.1 ± 0.7	59.5 ± 0.1

[a] Thermal shift assays were performed for wild-type and mutant ADP-Glc PPases as described in Materials and Methods.
[b] The melting temperature were determined using the Step-One software.
[c] The melting temperature for the double mutant K43A/G329D was not applicable due to lower stability of the enzyme.

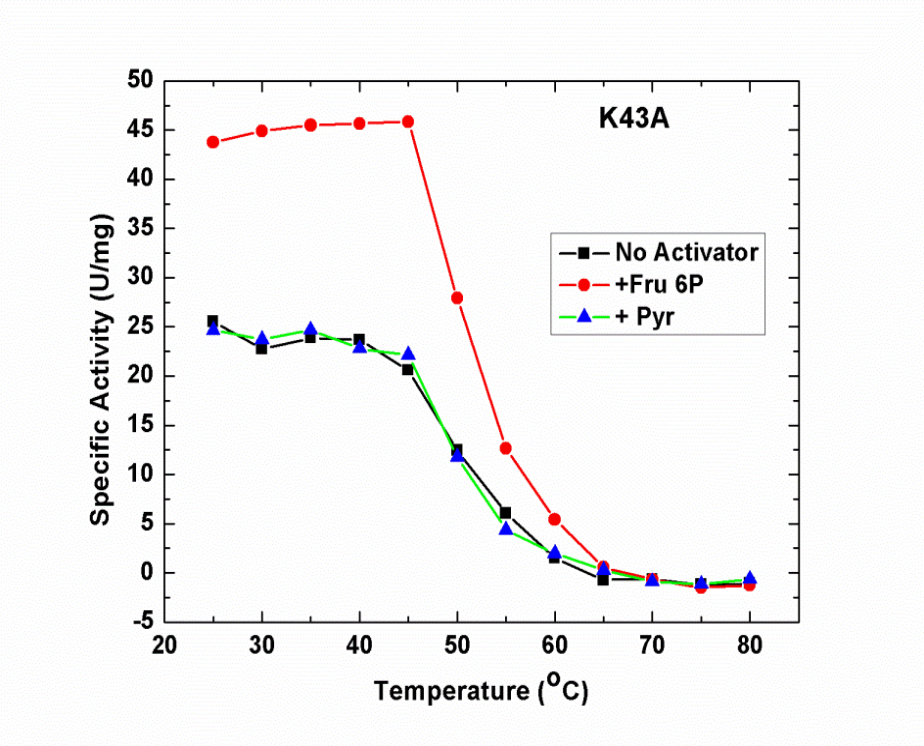


Figure 9: Heat stability assay for K43A mutant ADP-Glc PPase. The effect of heat on the structure stability was analyzed using the heat stability assay in the absence and the presence of the activators (Fru6P, and pyruvate). The (■) represents the absence of the activators, the (●) represents the data in the presence of 1.5 mM Fru6P, and the (▲) represents the presence of 1.5 mM pyruvate.

The G329D mutation mimics the presence of the pyruvate, by placing a carboxylate group at the position where the oxygens of pyruvate bind (Figure 5) [126]. The presence of aspartate at 329 positions seems to hinder the binding of pyruvate, since it is not expected to see any change in the thermal shift assay. In the presence of Fru6P, the melting temperature does increases up to 1.6°C for the G329D mutant enzyme. On the other hand, for G329D mutant the presence of pyruvate does not cause any change in the melting temperature of 63.3°C (Figure 8) (Table 5). Overall, D329 mimics the presence of pyruvate, but inhibits its binding. On the other hand, the K43A mutant enzyme does not bind pyruvate (Figure 8), probably due to lack of the positive charge at that position.

We hypothesize that K43 is involved not only in the binding of the activator but also in the signal that is transmitted to the active site for the activation. A way to test this hypothesis would be to "force" pyruvate into the regulatory site of the K43A mutant and see if the enzyme is activated. In theory, this is not feasible since K43 is critical for binding. However, we can mimic the presence of pyruvate with G329D. For this reason, we made the double mutant K43A/G329D to probe the importance of the positive charge at position 43 in *A. tumefaciens* ADP Glc PPase for activation signaling.

The double mutant is completely insensitive to pyruvate, while in the presence of Fru6P, it shows an increase in affinity for ATP without changing the V_{max} (Table 4). At a saturating concentration of ATP, the double mutant enzyme gets slightly activated by Fru6P with 1.2 activation fold, while there was no activation by pyruvate. The thermal shift assay for the double mutant could not detect a clear melting temperature, however from the kinetic parameters we can conclude that the double mutant binds to Fru6P.

Discussion

The ADP-Glc PPase is an allosterically regulated enzyme, which is involved in starch synthesis in plants, and glycogen synthesis in bacteria [61, 106]. Understanding enzyme regulation can be helpful for increasing starch, and glycogen production. Recently, the pyruvate-bound structure of ADP-Glc PPase from *A. tumefaciens* (mutant P96A) was solved (PDB: 5W5R). The crystal structure with activators bound (pyruvate and ethyl pyruvate) do not show any conformational differences compared to the structure without any activator. One of the reasons behind this could be the binding of sulfate to the crystal structure near Arg45 keeping the enzyme in an inactive form [126]. According to previous studies, the activator Fru6P competes with the sulfate [85]. We analyzed pyruvate activation in the presence of sulfate. According our analysis, at higher concentrations of sulfate, pyruvate does not activate at the previously known maximum velocity (Figure 10) [126]. Particularly, it shows a non-competitive effect, since pyruvate cannot displace sulfate. It was also confirmed from the structure that both ligands can co-exist together. According to previous studies, ethyl pyruvate and methyl pyruvate can activate ADP-Glc PPase from *A. tumefaciens* [126]. On the other hand, glyoxylate (similar structure to pyruvate, only lacking the methyl group) does not activate ADP-Glc PPase. Furthermore, glyoxylate, even at higher concentration, does not inhibit the enzyme in the presence or the absence of the pyruvate (Figure 11). This indicates that glyoxylate cannot bind to the enzyme. Similarly, alanine (similar structure like pyruvate, but tetrahedral in shape) cannot bind to the enzyme (Figure 11). This indicates the specificity of the pyruvate binding site in *A. tumefaciens* ADP-Glc PPase.

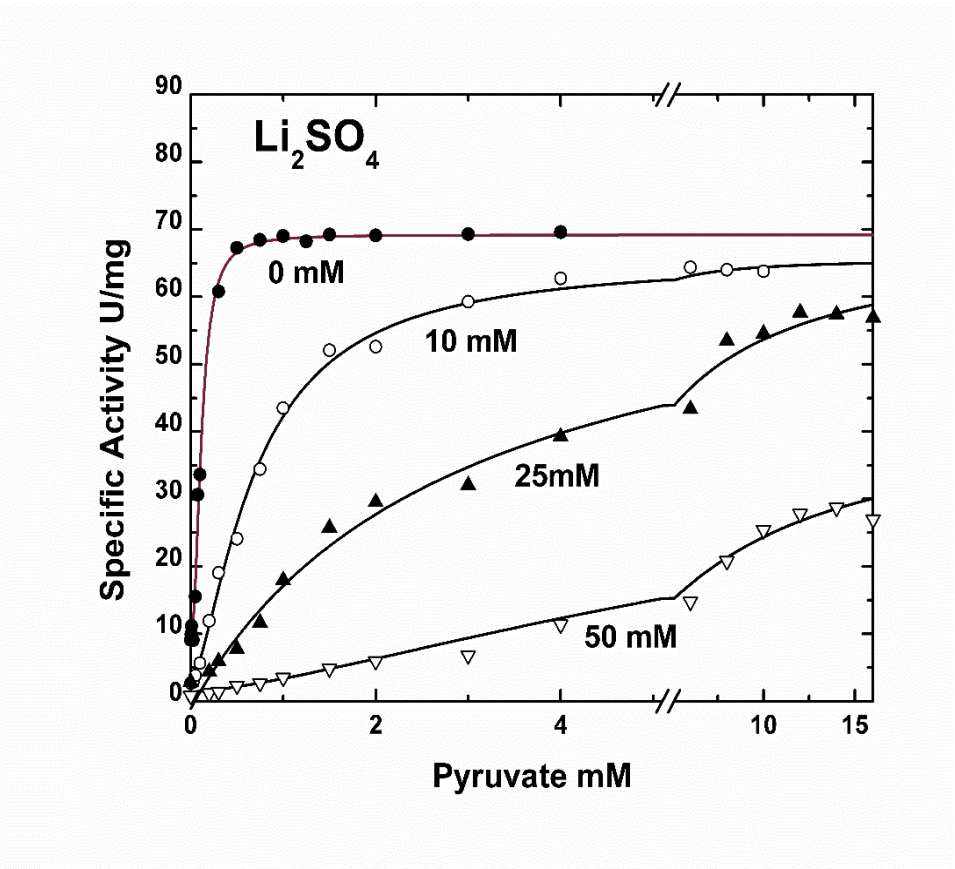


Figure 10. Effect of Li_2SO_4 on the activation by pyruvate of the *A. tumefaciens* ADP-Glc PPase. Saturation curves for pyruvate were obtained as described in Materials and Methods. In each curve, a constant amount of Li_2SO_4 was added as indicated.

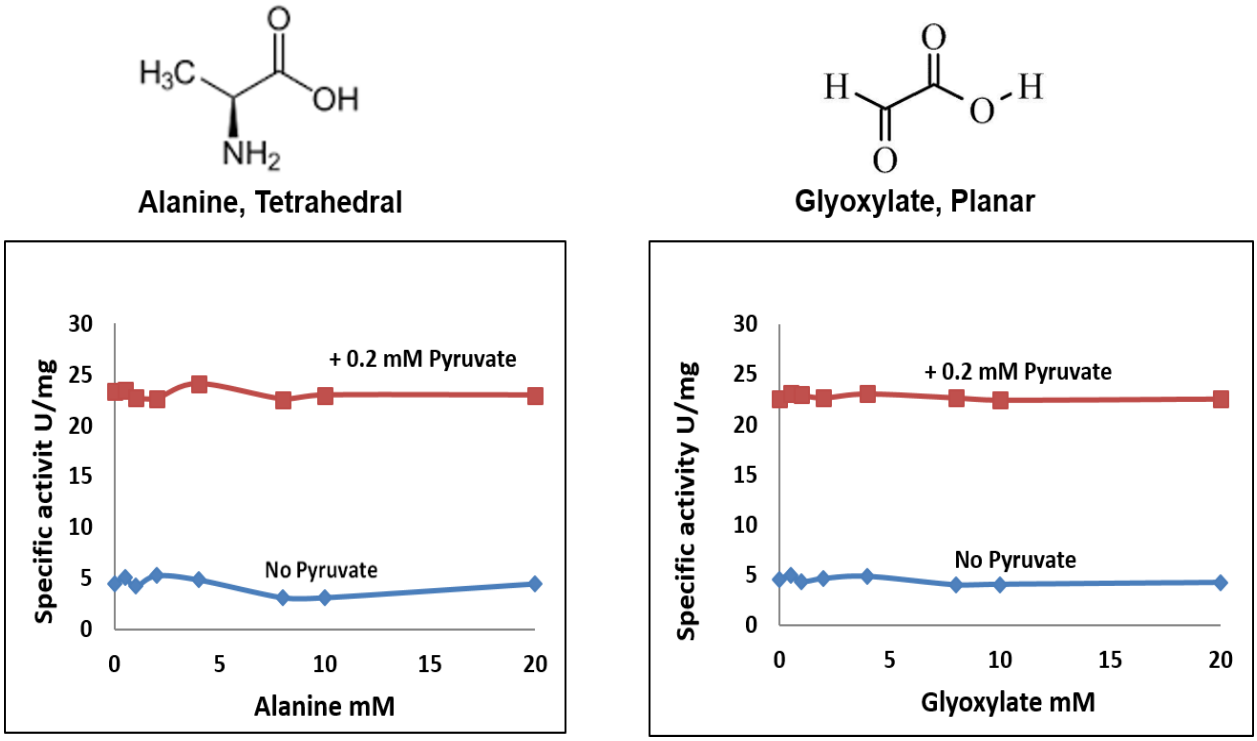


Figure 11. Kinetic characterization of ADP-Glc PPase activation by pyruvate analogs. The structure of the pyruvate analogs alanine and glyoxylate are shown. Alanine and glyoxylate do not activate ADP-Glc PPase in the absence of any activator (Blue line). In the presence of 0.2 mM Pyruvate, alanine and glyoxylate do not compete with the Pyruvate. Since, there is no inhibition in the activity of the ADP-Glc PPase (Red line).

Positive charge is crucial for binding pyruvate

In that structure, the K43 residues from subunits A, and B form hydrogen bonds with the pyruvate oxygens [126]. The K43 is critical for binding the pyruvate, as was demonstrated with the K43 mutations (Figure 7). As we know, the mutant K43A was completely insensitive to pyruvate, and was partially activated with Fru6P [126]. To find whether this mutation disturbed the enzyme, we crystallized the K43A mutant ADP-Glc PPase and solved the structure. The comparison of the K43A crystal structure with the recently-solved pyruvate, and ethyl pyruvate-bound structures showed no noticeable differences (Figure 12).

According to our results, the binding of the ligand ethyl pyruvate did not cause much difference in the overall homotetrameric structure. Similarly, the K43A mutation did not cause a disturbance in the homotetrameric structure of the ADP-Glc PPase (Table 2). To explore the importance of the side chain in position 43, we characterized several mutations. We retained the activation by binding the pyruvate with the K43R mutation that preserved the charge at 43 positions (Figure 7). On the other hand, the K43A, and K43N mutant could not bind to pyruvate and lost activation. Consistent with those results, we concluded that the positive charge at position 43 is a critical factor for the activation by pyruvate.

The Aspartate mutation of G329 makes the enzyme pre-activated

The G336D of the *E. coli* ADP-Glc PPase exists in a pre-activated state. When G336D is expressed in potato tuber, it resulted in significantly more starch production [30, 114]. In the conserved mutation G329D of ADP-Glc PPase from *A. tumefaciens*, the modeling of the aspartate side chain indicates that it mimics the presence of pyruvate (Figure 6). Moreover, the K43A mutant was unable to bind to pyruvate based on the experimental results. Conversely, with the double mutant K43A/G329D we were able to mimic the presence of pyruvate with the Aspartate at 329 position (Figure 13).

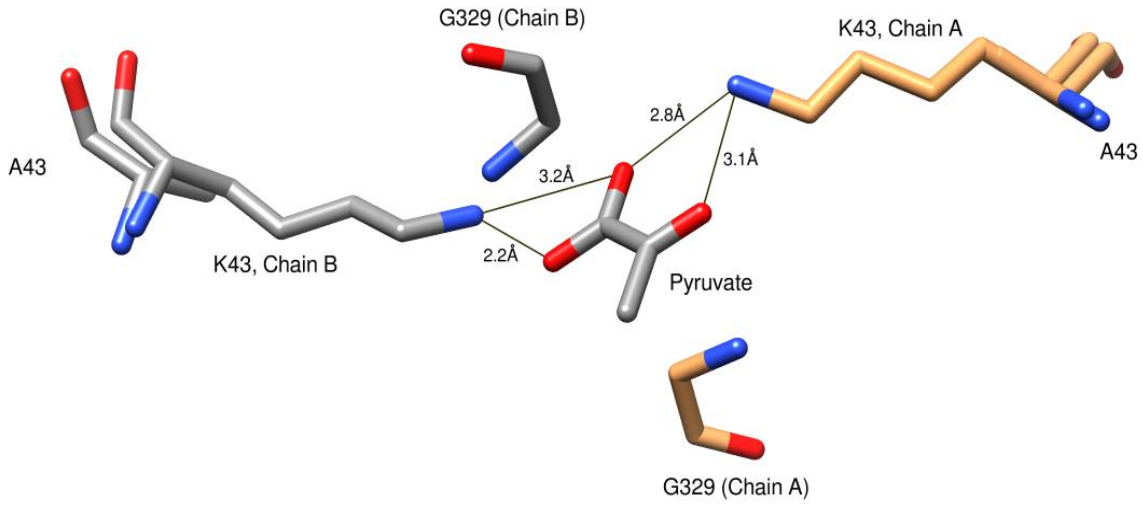


Figure 12: The alignment of K43A mutant with the wild type. The figure represents the hydrogen bonds of the K43 with the pyruvate in the wild type and the A43 aligned to K43.

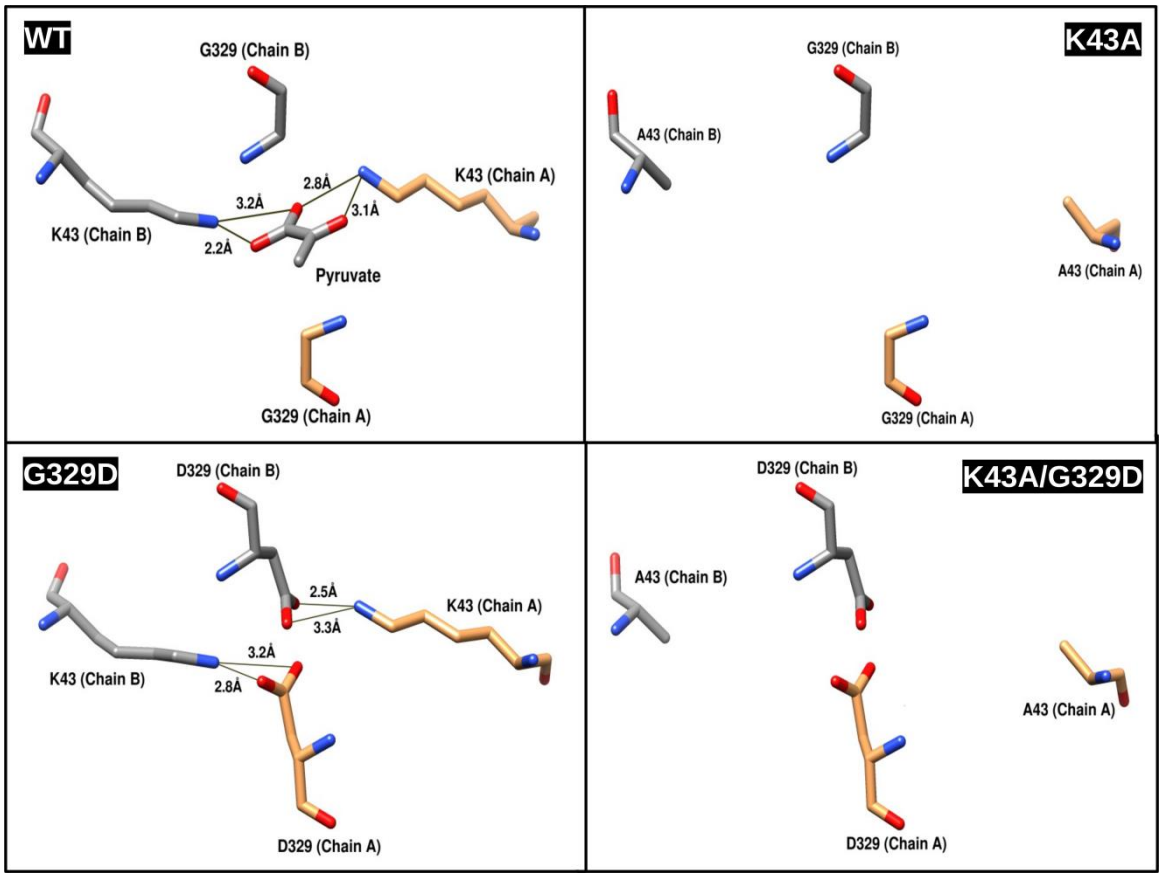


Figure 13. The model for ADP-Glc PPase from *A. tumefaciens* wild-type, and mutants K43A, G329D, and K43A/G329D. The models above represent the residues at pyruvate binding site for wild –type, and mutant enzymes.

Despite the presence of pyruvate is mimicked by the Asp at position 329, the enzyme is not activated. For that reason, we suggest that K43 is involved not only in binding to pyruvate, but also in triggering the activation after binding of the activator.

Materials and Methods

Site-Directed Mutagenesis

Quick Change Lightning Multi Site-Directed Mutagenesis Kits were used from Agilent Technologies for the site directed mutagenesis. To obtain the mutations K43A, K43N, K43R, and G329D, the pet28c vector containing the *A. tumefaciens* ADP-Glc PPase was used as a template. The primers that we used are K43A forward 5'-CGCGGTTTATTTTGGCGGCGCGGCGCGC-3'; K43R forward: 5'-CGCGGTTTATTTTGGCGGCAAGGCGCGC-3'; K43N forward: 5'-CGCGGTTTATTTTGGC GGCAACGCGCGC-3'; and G329D forward 5'-CGTCGGTCGTCTCGGATGACTGCATCATTTC-3'. The oligonucleotides for mutations were synthesized by Integrated DNA Technologies (IDT). The mutations were verified with genetic sequencing performed by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility in Chicago, Illinois.

Expression and Purification

The wild-type and the mutants were expressed in *Escherichia coli* BL21 DE3 cells. The transformations were performed using the wild type, and each different mutant genes. The transformation mixes were plated on selection plates treated with X-gal for blue-white colony selection. White colonies were screened for the correct ligation of the mutated gene into the pet28c vector using colony PCR. The Bl21 cells with correct plasmids were grown to an OD_{600} between 1.1-1.3 at 37°C and cooled on ice. The culture was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 25°C for 16 hours with shaking at 250 rpm. The cells were harvested by centrifugation and

sonicated in buffer C (50 mM HEPES (pH 7.5), 10% glycerol, 200 mM NaCl, and 10 mM Imidazole). Crude extracts were loaded onto a pre-equilibrated 5 ml His-Trap FF column (Ni^{2+}Sepharosecolumn) and eluted with a linear gradient 0-50% of buffer E (50 mM HEPES pH 7.5, 10% glycerol, 200 mM NaCl, and 750 mM imidazole). Active fractions were pooled, and concentrated after SDS-PAGE and enzyme assay, which was stored in the buffer E at -80°C in aliquots and used for kinetic characterization. The concentrated proteins were used for further purification using equilibrated gel filtration column (Superdex^{TM}200, 10/300 GL) with buffer X (50 mM HEPES (pH7.5), 5% glycerol, and 200 mM NaCl). Fractions containing enzyme were pooled, concentrated, and supplemented with 5% (v/v) glycerol. The proteins were stored at -80°C until use, which remained stable and fully active for at least three months.

Enzyme assay

MG-AM assay: Malachite Green-Ammonium Molybdate (MG-AM) solution was prepared by using 1.5% MG solution, and 34 mM Ammonium Molybdate. Tween-20 is used to maintain the color. The total reaction volume was 50 µl, from which 30 µl was premix, 10 µl of enzyme, and 10 µl of the variant (which in case of the activator saturation curve were activators or in the substrate saturation curve were substrate) (Figure 14)

Substrate saturation assay. We used the various concentrations of a substrate (ATP) for this assay. The premix used for control reaction contains 50 mM HEPES (pH-7.5), 14mM $MgCl_2$, 1.5mM Glc 1-Phosphate, 0.005 U/µL PPase, and 0.2 mg/ml BSA. The reaction starts with 10µl of the enzyme dilution. We added the MG-AM solution, after incubation of 10 minutes to bind with PPi formed in the reaction; and we measured the data using a spectrophotometer at 595nm absorbance. We did the same reaction with the presence of 1.5 mM activators (Fru6P, and Pyruvate) in the premixes.

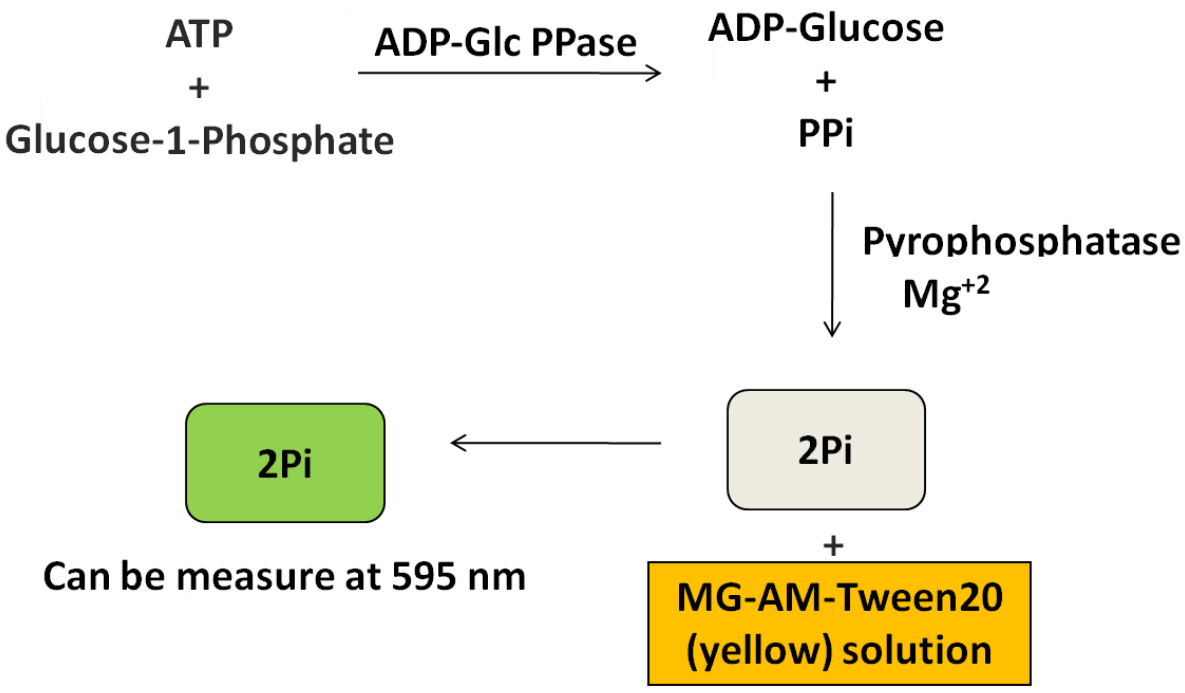


Figure 14. The colorimetric malachite green enzymatic assay. The Malachite Green Ammonium Molybdate (MG-AM) solution is yellow in color, which can bind to Pi to give the green color. It can be measured at 595 nm, and give us the relative number of the ADP-Glucose produced in the reaction.

Activator saturation assay. The premix used for this reaction contains 50 mM HEPES (pH-7.5), 7 mM $MgCl_2$, 1.5 mM Glc 1-Phosphate, 1.5 mM ATP, 0.005 U/µl PPase, and 0.2 mg/ml BSA. The reaction starts with 10µl of the enzyme dilution. After incubation of 10 minutes the MG-AM solution is added to bind with PPi formed in the reaction, the data are observed with a spectrophotometer at 595nm absorbance.

Kinetics

$S_{0.5}$and $A_{0.5}$ values indicate the concentration of substrate or activator needed to give 50% of the maximal velocity. The Hill coefficient (*nH*) was calculated from fitting data in Origin™ 6.0 to the non-linear least square formula, by using the Hill equation $V= V_0 + [(Vmax*S^n) / (K_m+ S^n)]$ for activator saturation plots, and for substrate saturation plots the Hill equation $V= (V_{max}*S^n) / (K_m+ S^n)$ was used. All values were the average of duplicate reactions with errors of 10% or more. We used these plots to calculate the substrate concentration ($S_{0.5}$), and activator concentration ($A_{0.5}$), which gives 50% of the maximum velocity (V_{max}) and the Hill number (*nH*). Catalytic efficiency (k_{cat}/K_m) (*Ce*) was calculated with Origin™ 6.0, and using the Hill equation $V=Ce*V_{max}*S^n/ (K_m+ S^n)$.

Note: For all enzyme assays, standard PPi curves were performed, which are used to calculate the PPi formed in the reaction.

Thermal shift assay

All the thermal shift assays (melt curve), were done using the Step One Real-Time PCR System and Step One software was used to design the assay. The assay was done with 20µl of final reaction volume, with 20 mM HEPES (pH 7.5), SYPRO Orange Dye (4X), 0.02 mM purified protein, 1 mM Fru6P, 1 mM Pyruvate, and 2 mM ATP. Depending on the requirements of the particular assay, we added or neglected the activators (Fru6P, Pyr) and/or substrate (ATP). We did the entire assay with triplicates, and compared with the wild type (WT) data. No protein control also done for all with/ without activators, and substrate, to avoid any false results. The continuous melt curve has been done with the ramp increment of 1%, starting from 25.0°C (2.00 min) to the end temperature of 99.0°C (2.00 min).

Heat stability assay

An aliquot of the enzyme was incubated at different temperatures (from 25°C to 80°C) with buffer A (50 mM HEPES (pH 7.5), 0.2 mg/ml BSA) for 10 minutes. Right after the incubation, 10 µl of heat treated enzyme was used to assay the activity at 37°C for 10 minutes. The reaction was assayed with duplicates, as described in the method of MG-AM assay.

Crystallization and data collection

The K43A mutant enzyme was crystallized after screening and optimization using the hanging drop method. The reservoir solution contained 50 mM HEPES (pH 7.5), and 2M lithium sulfate, and the hanging drops were prepared by mixing 1.5 µl of 9.8 mg/ml mutant ADP-Glc PPase and 1.5 µl of the reservoir solution. The crystals were grown for 1-2 week at 20°C. The large sized crystals were soaked in cryo-condition with 25% glycerol, and the reservoir solution before freezing it with liquid nitrogen. The data was collected at SBC-19-ID Beamline by collaborators (Dr. Dali Liu and Dr. Romila Mascarenhas.

CHAPTER THREE

INTER-SUBUNIT SURFACE INTERACTION AND THE REGULATION OF THE ADP-GLUCOSE PYROPHOSPHORYLASE FROM *AGROBACTERIUM TUMEFACIENS*

Abstract

ADP-glucose pyrophosphorylase (ADP-Glc PPase) is a key regulatory enzyme involved in starch and glycogen synthesis in plants and bacteria, respectively. The enzyme from *A. tumefaciens* is a homotetramer allosterically regulated by fructose 6-phosphate (Fru6P) and pyruvate (Pyr). It has been hypothesized that inter-subunit communications are important for the allosteric effect in this enzyme. However, no specific interactions have been identified to be part of the regulator signal. Three pairs of distinct subunit-subunit interfaces are present in this enzyme. Here we focus on an interface features the interaction between R11 and D141 of one subunit and residues D141 and R11, respectively, of the neighbor subunit. Previously, it was shown that a mutation at R11 position caused disruption of the activation of the enzyme. For that reason, our hypothesis was that the interaction between R11 and D141 is critical for the allosteric effect. To prove our hypothesis, we introduced several mutations in those two sites (D141A, D141E, D141N, D141R, R11D and R11K). According to our results, changes in charge were the ones that affected the regulation the most (Table 6). To prove that the interaction is important rather than the presence of specific residues, the mutant R11D was partially rescued with the double mutant R11D/D141R. This double mutant could not restore the effect of the activators on V_{max}, but it did rescue the Fru6P and Pyr effect on the affinity for the substrates. All these results indicate the critical functional role of the D141 and R11 residues in this subunit interface and the relay of the

allosteric signal.

Introduction

The ADP-Glc Pyrophosphorylase (ADP-Glc PPase, EC 2.7.7.27) catalyzes a key regulatory step in glycogen biosynthesis in bacteria, and starch biosynthesis in higher plants. The accumulation of glycogen by bacteria may give advantages during starvation periods, providing a stored source of energy and carbon surplus [13]. ADP-Glc PPases in bacteria function as homotetramers (α_4) of a ~50 kDa subunit, whereas in the plant this enzyme functions as a heterotetramer (α_2 β_2) [61, 62]. ADP-Glc PPase catalyzes the conversion of ATP and glucose-1-phosphate (Glc-1-P) to ADP-glucose (ADP-Glc) and inorganic pyrophosphate (PPi) in the presence of Mg^{2+}. This reaction was first described in soybean [10]. The ADP-Glc PPase is allosterically regulated by key intermediates and is involved in carbon assimilation in the host organism. The study of *A. tumefaciens* ADP-Glc PPase shows that this enzyme is activated by fructose 6-phosphate (Fru6P) and pyruvate (Pyr), but inhibited by phosphate (Pi).

The similarities between the plant and bacterial enzymes have allowed the use of the bacterial forms, which have simpler oligomeric structure and more convenient in producing recombinant proteins, as models for the plant system [83]. The plant ADP-Glc PPase consists of two small and two large subunits, and both subunits are needed for enzyme activity [71]. The interaction between the two subunits has been studied before showing its importance in the allosteric regulation of the ADP-Glc PPase. The synergistic interaction between small and the large subunits provides the regulatory and kinetic properties of the ADP-Glc PPase in higher plants [123]. The areas of the large subunit that participate in tail-to-tail and head-to-head interactions with the small subunit are important for the allosteric properties of ADP-Glc PPase of maize endosperm [125].

The previous studies have shown that the N-terminal Arginines 5 and 11 are clearly involved in activation by pyruvate in the case of *A. tumefaciens* ADP-Glc PPase [109]. Based on the published structure of the *A. tumefaciens* ADP-Glc PPase (PDB: 5W6J) R11 is located at the interface, and is also involved in the inter-subunit interaction [108]. R11 makes a salt bridge with D141 of a neighboring subunit. In the same interface, D141 of the first subunit makes one additional salt bridge with R11 of the second subunit (Figure 15). Therefore, in the homotetramer structure of the ADP-Glc PPase of *A. tumefaciens*, four R11 and four D141 make a total of four salt bridges.

In the potato tuber ADP-Glc PPase, the intermolecular disulfide bridge located between Cys12 of the two small subunits is critical for the enzyme's stability at 60°C [116]. After cleavage of the disulfide bond the enzyme is stable only up to 40°C [116]. This highlights that the stability of the enzyme is affected by residues that interact in the inter-subunit interface. In previous computational studies, it was found that conserved residues in the potato tuber are part of the subunit-subunit interacting regions [128]. In addition, it was found that relatively similar ADP-Glc PPase subunits from different species, such as potato tuber and *A. thaliana,* can interact and produce active hybrid forms with varying regulatory properties [121]. It was observed that inter-subunit interactions may also play a role in the allosteric regulation of the enzyme. Removal of the C12 disulfide bond either by mutation or by reduction results in an enzyme that is nearly constitutively active [29], which indicates that the inter-subunit interaction between the α subunits is an important part of the allosteric mechanism [70]. Here we are reporting kinetic properties of different mutants D141A, D141E, D141N, D141R, R11D, R11K and a double mutant R11D/D141R of *A. tumefaciens* ADP-Glc PPase, which provides important information about contribution of inter-subunit surface interaction to allosterism. This research suggests that a better understanding of the subunit interactions can provide great insights in

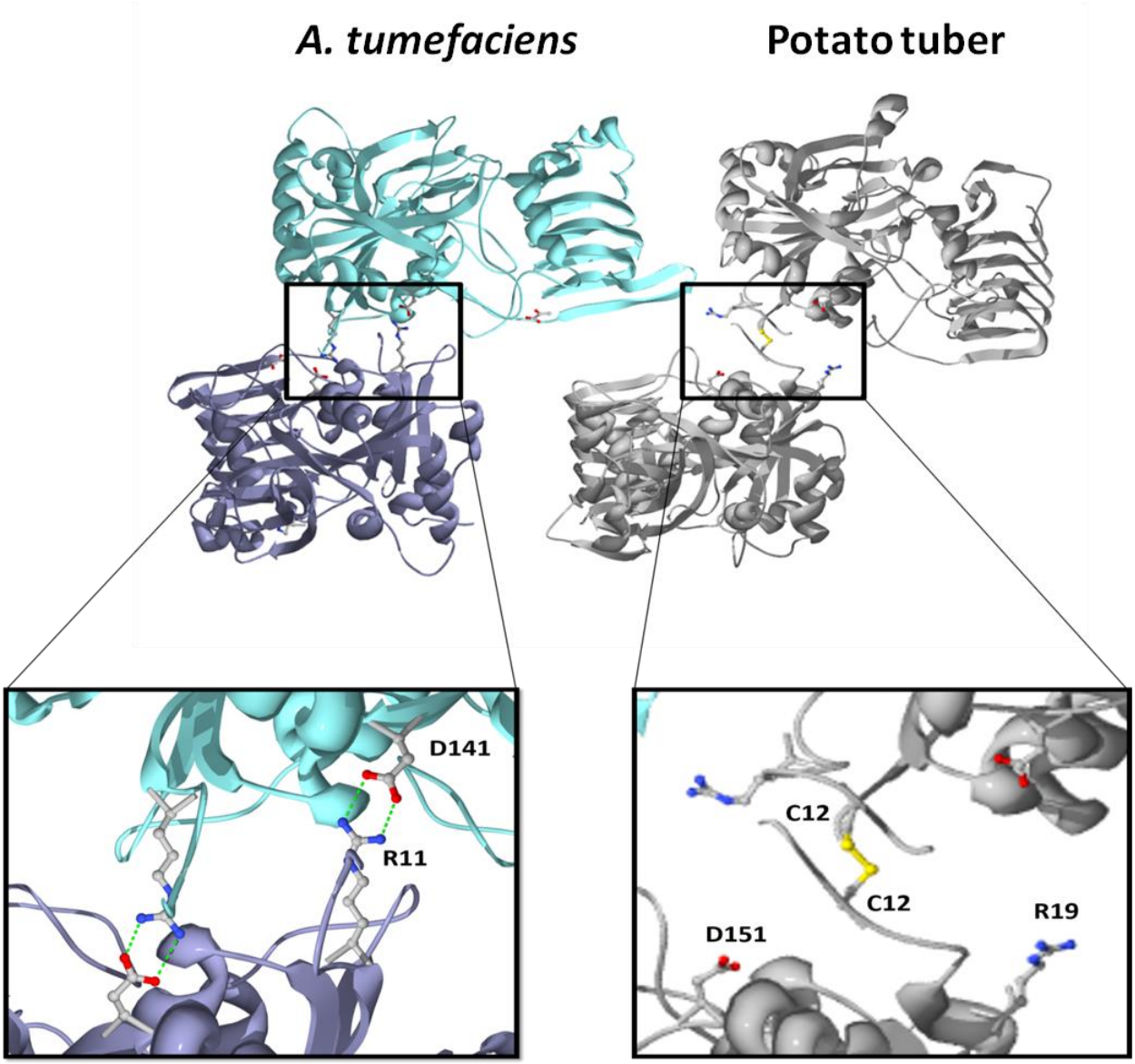


Figure 15. The interface Inter-subunit interaction in *Agrobacterium tumefaciens* and Potato tuber. In the *A. tumefaciens* dimer structure four hydrogen bond between Arg11 and Asp141 are shown. While in the potato tuber dimer the disulfide bond between Cys12 is shown with conserved Arg19 and Asp151 residue

designing hybrid ADP-Glc PPase forms with desired regulatory properties.

Results

According to previous studies, R11A mutated ADP-Glc PPase enzyme of *A. tumefaciens* displayed desensitization to pyruvate, partial activation by Fru6P, and increased sensitivity to phosphate inhibition. It was concluded that R11 is involved in pyruvate activation, and postulated that perhaps it was acting in part by providing an anionic binding site for the carboxyl group of the ligand [109]. However, we observed in the crystal structure of the *A. tumefaciens* enzyme (PDB: 5W6J) that R11 makes salt bridge with D141 of a neighboring subunit. In the same interface, D141 of the first subunit makes another two hydrogen bonds with R11 of the second subunit. Therefore, in the homotetramer structure of ADP-Glc PPase of *A. tumefaciens*, four R11 and four D141 are making a total of four salt bridge. (Figure 15)

Sequence analysis

According to an alignment of representative ADP-Glc PPases from diverse species, D141 is highly conserved. Out of 63 plant sequences, 58 sequences have Aspartate at homologous positions to D141 of ADP-Glc PPase of *A. tumefaciens*. Another 5 plant sequences have Asparagine instead of Aspartate. From 103 bacterial sequences, 93 sequences have Aspartate, 9 sequences have Asparagine, and only 1 sequence have Glutamate. On the other hand, at homologous positions to R11, Arginine is less conserved compared to Aspartate at 141 positions. In plants, 16 sequences have Arginine, 23 sequences have Lysine, and 24 sequences have other various amino acids. Similarly, in bacteria, 35 sequences have Arginine, 29 sequences have Lysine, and 39 sequences have other various amino acids (Appendix A).

Gel filtration analysis

We sought to analyze the roles of D141 by site directed mutagenesis. We obtained several mutants: D141A, D141E, D141N, D141R, R11D, R11K, and double mutant R11D/D141R of ADP-Glc PPase of *A. tumefaciens*. Considering that the mutation is at the interface, the quaternary structure might be disrupted. For that reason, we analyzed the molecular mass for each mutant. According to an analytical gel filtration study, all the mutants ADP-Glc PPase were homotetramers like the wild-type (Figure 16). The only exception was the D141R mutant that appeared to be a homodimer (Figure 16). The retention volume (ml) for the wild-type was 13.33 ml, whereas for the D141R mutant it was 14.56 ml.

Effect of activators (Fru6P, Pyruvate) on the saturation curves of ATP

The effect of the presence of the activators (Fru6P, Pyruvate) on the kinetic parameters for the substrate ATP was analyzed (Table 6). With these experiments we probed whether the synergy between regulators and substrates has been altered by changing the side chains in positions 11 and 141. In the absence of activators, the non-conservative mutations D141N and D141R almost lost their activation with a drastic decrease in the V_{max} (186- and 559-fold, respectively). The R11D mutant had a lower apparent affinity for the ATP (5.3-fold higher $S_{0.5}$ compared to wild-type). The R11A and D141A mutants had a 2.2- and 7-fold higher $S_{0.5}$value for ATP. On the other hand, in the absence of activators the conservative mutation R11K was slightly more activated (3.6-fold higher in V_{max}) than the wild-type with a 4.8-fold decrease in $S_{0.5}$ value for ATP. In absence of activators, the D141E mutant had similar kinetic characteristics to the wild-type.

In the presence of activator Fru6P, the non-conserved mutations D141A, D141N, and D141R had no significant effect on the kinetic parameters for the substrate ATP. The R11D mutant was the one

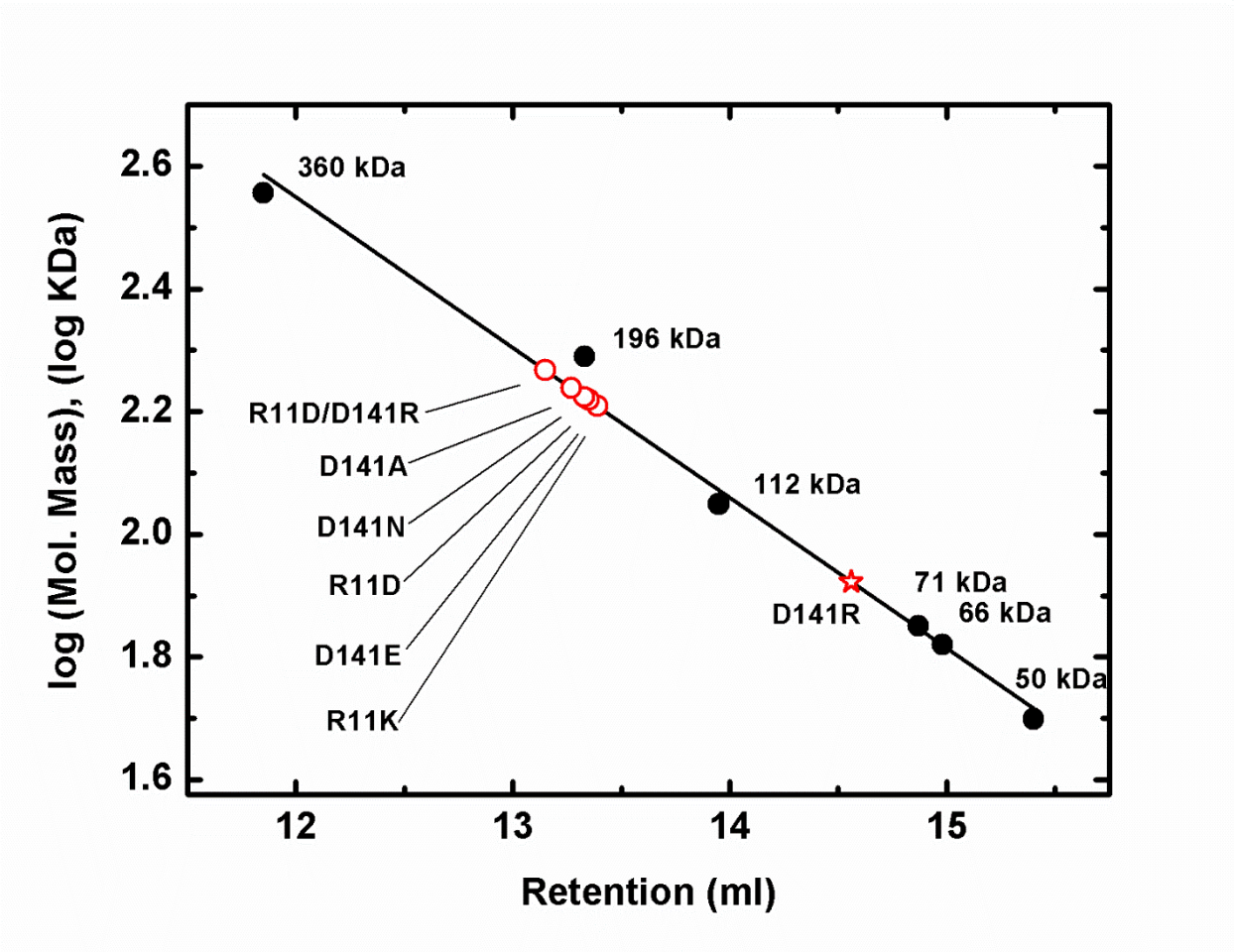


Figure 16. Analytical gel filtration chromatography of ADP-Glc PPases wild-type (WT), and mutant (D141A, D141E, D141N, D141R, R11D, R11K, and R11D/D141R). The gel filtration analysis has been performed using the Superdex^{TM}200, 10/300 GL column. The wild-type ADP-Glc PPase is homotetramer, with a molecular weight of ~ 196 kDa. The markers (●) for this analysis were described in materials, and methods. The (☆) represents the D141R enzyme, and the (○) represents R11D, R11K, D141A, D141E, D141N, and R11D/D141R mutant enzymes.

Table 6. Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants

Enzyme [a]	Control			+ 1.5mM Fru6P			+ 1.5mMPyruvate		
	$S_{0.5}$ (ATP)	n_H	*Vmax*	$S_{0.5}$ (ATP)	n_H	*Vmax*	$S_{0.5}$ (ATP)	n_H	*Vmax*
	mM		U/mg	mM		U/mg	mM		U/mg
WT	0.21 ± 0.01	1.9 ± 0.2	11.18 ± 0.30	0.08 ± 0.01	1.7 ± 0.1	123.48 ± 2.36	0.13 ± 0.01	2.2 ± 0.2	85.44 ± 1.91
D141A	1.49 ± 0.06	2.0 ± 0.2	3.82 ± 0.15	1.39 ± 0.09	2.1 ± 0.3	3.52 ± 0.11	0.66 ± 0.03	2.9 ± 0.3	8.28 ± 0.11
D141E	0.24 ± 0.02	1.8 ± 0.2	12.58 ± 0.23	0.09 ± 0.01	1.4 ± 0.1	27.15 ± 0.40	0.10 ± 0.03	2.2 ± 1.0	21.76 ± 0.59
D141N	3.99 ± 0.70	3.1 ± 1.1	0.06 ± 0.01	3.01 ± 0.39	3.0 ± 0.9	0.05 ± 0.01	3.19 ± 0.21	4.1 ± 0.9	0.05 ± 0.01
D141R	0.56 ± 0.10	1.0 ± 0.2	0.02 ± 0.01	0.70 ± 0.24	0.8 ± 0.2	0.03 ± 0.003	0.32 ± 0.02	1.1 ± 0.1	0.02 ± 0.01
R11A[b]	0.47 ± 0.10	2.4 ± 0.5	4.10 ± 0.50	0.05 ± 0.01	1.5 ± 0.5	8.3 ± 0.5	0.29 ± 0.02	2.1 ± 0.4	5.0 ± 0.3
R11D	1.11 ± 0.25	1.0 ± 0.1	14.98 ± 1.33	3.39 ± 1.81	0.7 ± 0.1	20.74 ± 4.01	1.13 ± 0.34	0.8 ± 0.1	16.72 ± 1.61
R11K	1.01 ± 0.05	1.4 ± 0.1	40.86 ± 0.98	0.06 ± 0.03	1.2 ± 0.5	47.11 ± 1.93	0.23 ± 0.01	2.8 ± 0.5	36.21 ± 0.64
R11D/D141R	0.44 ± 0.03	1.7 ± 0.2	12.11 ± 0.31	0.06 ± 0.01	1.4 ± 0.7	10.29 ± 0.13	0.18 ± 0.01	2.4 ± 0.4	13.85 ± 0.19

[a] Assays were performed as described in Assay A (substrate saturation), as stated under Materials and Methods.

[b] The results for R11A mutant ADP-Glc PPase for *A. tumefaciens* are from literature [109].

with the lowest apparent affinity for ATP in the presence of Fru6P. While, according to previous studies, the R11A enzyme in the presence of Fru6P had a similar $S_{0.5}$ for ATP compared to the wild-type [109], but, the V_{max} (8.3 U/mg) was lower than the wild-type (123.5 U/mg). In contrast, in the presence of Fru6P, mutations that conserved the charge (R11K and D141E) displayed a similar apparent affinity for ATP but only 38%, and 22% of wild-type activity, respectively.

The presence of pyruvate shows no significant effect on the ATP kinetic parameters of the R11D, D141R and D141N mutants. The R11A was slightly activated by pyruvate with a relatively lower $S_{0.5}$ for ATP, compared to the data in the absence of activators. The D141A, which was insensitive to Fru6P, was fairly activated in the presence of pyruvate with a higher affinity for ATP. On the other hand, the conservative mutants R11K, and D141E restored 42%, and 25%, respectively, of the wild-type activity with an identical apparent affinity for ATP compared to wild-type.

Catalytic efficiency

In the absence of activators, the conserved mutations R11K and D141E showed similar catalytic efficiency as wild-type (Table 7). In the presence of Fru6P the R11K and the D141E showed 2- and 6-fold decrease, while in the presence of pyruvate they showed 4- and 3-fold decrease in the catalytic efficiency. The mutant enzymes R11D and D141A showed 5- and 17-fold decrease in the catalytic efficiency compared to wild-type in the absence of activators. There is not much difference in the catalytic efficiency in the presence of activators for the R11D and D141A mutant enzymes. In the absence of activator, the R11A mutant displayed a 6-fold decrease compared to the wild-type. R11A mutant enzyme activity increased up to 19-fold in the presence of Fru6P, there was little effect of pyruvate compared to the data in the absence of activator. On the other hand, the D141N and D141R mutant enzymes completely lost their activation in the presence of the activators.

Table 7. Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants.

	Control	+ 1.5mM Fru 6P	+1.5mM Pyruvate
Enzyme [a]	$k_{cat} / S_{0.5}$ [b] (ATP)	$k_{cat} / S_{0.5}$ [b] (ATP)	$k_{cat} / S_{0.5}$ [b] (ATP)
	$s^{-1}\ mM^{-1}$	$s^{-1}\ mM^{-1}$	$s^{-1}\ mM^{-1}$
WT	44.5 ± 1.7	1292 ± 55	549 ± 25
D141A	2.57 ± 0.26	2.27 ± 0.12	10.55 ± 0.38
D141E	43.2 ± 2.8	231 ± 11	185.6 ± 7.8
D141N	0.020 ± 0.001	0.020 ± 0.001	0.020 ± 0.001
D141R	0.040 ± 0.003	0.040 ± 0.001	0.070 ± 0.001
R11A[c]	7.27	138	14.4
R11D	9.6 ± 1.2	4.04 ± 0.52	12.4 ± 1.0
R11K	33.3 ± 1.5	661.3 ± 9.9	141.6 ± 7.7
R11D/D141R	22.45 ± 0.52	135.1 ± 9.2	64.7 ± 3.3

[a] Assays were performed as described in Assay A (substrate saturation), as stated under Materials and Methods.

[b] $k_{cat} / S_{0.5}$ (ATP) was calculated as described in Materials and Methods.

[c] The results for R11A mutant ADP-Glc PPase for *A. tumefaciens* are from the literature [109].

Activation at saturating concentrations of substrate (ATP)

To determine the contribution of the inter-subunit interaction between R11 and D141 on the enzyme regulation, we analyzed the activation by Fru6P and Pyr of different mutants. At saturating concentrations of the substrate ATP, the altered ADP-Glc PPases had significant changes when compared to the wild-type enzyme. The R11D, D141R and D141N exhibits no apparent activation by Fru6P at a saturating concentration of ATP, whereas the D141E, R11K and R11A were activated slightly by Fru6P with a 11-, 6.8- and 11.3- fold decrease in $A_{0.5}$ value. On the other hand, D141A activates at 3.3-fold with a decrease in apparent affinity (2.5-fold increase in $A_{0.5}$ value compared to the wild-type enzyme) (Table 8)

The R11A, R11D, D141R, and D141N mutants were insensitive to pyruvate activation at saturating concentration of ATP (Figure 17). R11K and D141E were only slightly activated by pyruvate (1.86- and 1.47-fold, respectively) with an 11- and 7.5-fold decrease in $A_{0.5}$ value. On the other hand, D141A was activated 17.8-fold by pyruvate with a similar apparent affinity relative to the wild-type. However, the absolute V_{max} of D141A was only 7.3 U/mg compared to 87.4 U/mg of the wild-type enzyme in the presence of pyruvate. In fact, the maximum activities (V_{max}) reached by all the mutants in the presence of saturating concentrations of ATP and activator were lower than the wild-type. Only D141E and R11K, which are conservative mutations, reached V_{max} of 31.9, and 47.73 U/mg respectively, for Fru6P. Similarly, for pyruvate the V_{max} reached up to 21.73, and 36.9 U/mg, respectively for D141E, and R11K.

Activation at sub-saturating concentrations of substrate (ATP)

At a sub-saturating concentration of substrate, the effect of activators (Fru6P, Pyr) on the altered ADP-Glc PPases were remarkable (Figure 18). The R11D and D141R single mutant enzymes show

Table 8. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.

Enzyme [a]	Fru6P				Pyruvate			
	$A_{0.5}$	n_H	*Vmax*	Activation[b] (V_{max}/V_0) (-fold)	$A_{0.5}$	n_H	*Vmax*	Activation[b] (V_{max}/V_0) (-fold)
	mM		U/mg		mM		U/mg	
WT	0.34 ± 0.07	2.1 ± 0.5	105.2 ± 3.9	9.01	0.45 ± 0.04	2.1 ± 0.4	87.4 ± 2.0	8.53
D141A	0.84 ± 0.07	2.7 ± 0.6	1.13 ± 0.04	3.32	0.30 ± 0.02	2.8 ± 0.5	7.29 ± 0.14	17.78
D141E	0.03 ± 0.01	1.9 ± 0.6	31.86 ± 0.17	1.81	0.06 ± 0.03	1.2 ± 0.5	21.73 ± 0.23	1.47
D141N	2.97 ± 0.15	3.9 ± 0.6	0.02 ± 0.01	1.20	N/A[d]	N/A[d]	0.02 ± 0.01	1.00
D141R	N/A[d]	N/A[d]	0.02 ± 0.01	0.96	N/A[d]	N/A[d]	1.01 ± 0.17	0.95
R11A[c]	0.03 ± 0.008	1.3 ± 0.4	7.2 ± 0.2	1.8	N/A[d]	N/A[d]	5.0 ± 0.5	1
R11D	N/A[d]	N/A[d]	9.96 ± 0.16	0.98	N/A[d]	N/A[d]	9.96 ± 0.55	0.98
R11K	0.05 ± 0.02	1.0 ± 0.3	47.73 ± 0.94	2.91	0.04 ± 0.02	0.6 ± 0.3	36.9 ± 2.3	1.86
R11D/D141R	1.32 ± 0.88	1.5 ± 0.3	13.75 ± 0.96	1.25	N/A	N/A	10.49 ± 0.87	0.99

[a] Assays were performed as described in Assay B (activator saturation), as stated under Materials and Methods.
[b] Activation fold is calculated by dividing the maximum velocity (V_{max}) by the velocity in the absence of activator (V_0) (V_{max}/V_0).
[c] The results for R11A mutant ADP-Glc PPase for *A. tumefaciens* are from literature [109].
[d] No significant activation was observed to calculate activation parameters.

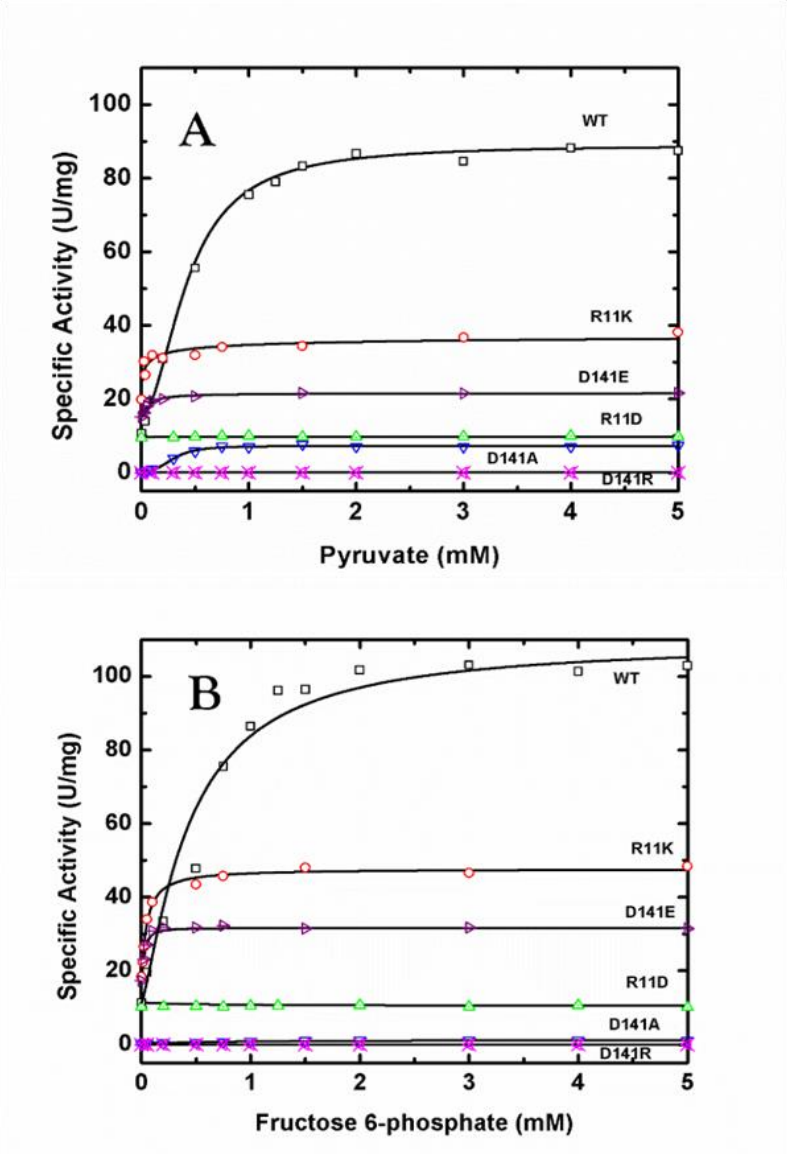


Figure 17. Activator saturation curves of wild-type and mutant ADP-Glc PPase from *A. tumefaciens* at saturating concentrations of ATP. The effect of Pyruvate (A) and Fru6P (B) was assayed on the wild-type (WT), and D141A, D141E, D141R, R11D and R11K mutant ADP-Glc PPases. The assays were performed as described in Materials and Methods (Assay B) in the presence of 1.5mM ATP.

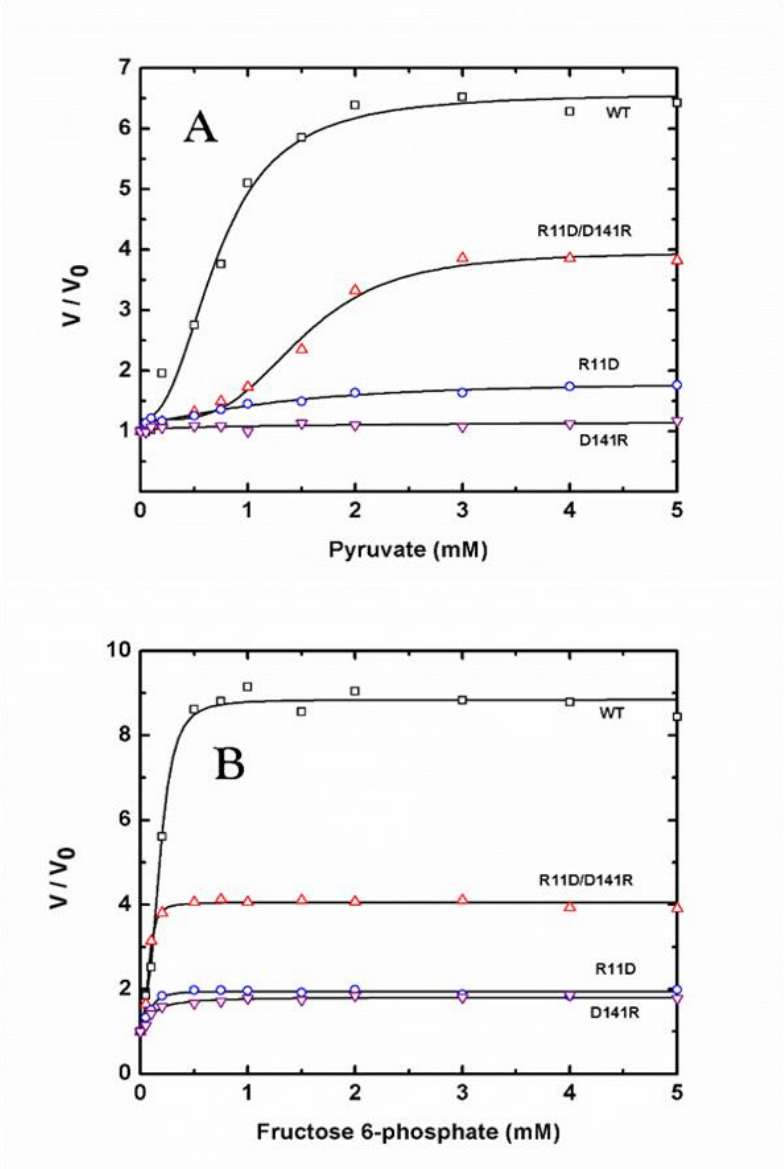


Figure 18. Relative activation curves of wild-type and mutant ADP-Glc PPase from *A. tumefaciens* at sub-saturating concentrations of ATP. The effect of Pyruvate (A) and Fru6P (B) was assayed on the wild-type (WT), andR11D/D141R, D141R, and R11D mutant ADP-Glc PPases. The assays were performed as described in Materials and Methods (Assay B) in the presence of 0.2 mM ATP.V_0 is the velocity assayed in the absence of activator. In the Pyruvate saturation curve, V_0 for WT, R11D, D141R, R11D/D141R were 8.20 U/mg, 2.23 U/mg, 0.006 U/mg, and 1.97 U/mg respectively. Correspondingly, in the Fru6P saturation curve, $V_0$for WT, R11D, D141R, R11D/D141R were 8.71 U/mg, 2.35 U/mg, 0.007 U/mg, and 2.17 U/mg respectively.

only 21%, and 24% activation, respectively, by Fru6P compared to wild-type. However, the activation by pyruvate for R11D and D141R were 18%, and 27%, respectively.

Activation rescue by a secondary mutation (R11D/D141R)

The analysis of the single mutant enzyme R11D and D141R showed a drastic effect on the activation, whereas the double mutant R11D/D141R rescued the activation by both the activators (Fru6P and Pyr). The single mutants R11D and D141R showed decrease in apparent affinity for the substrate ATP in both the absence and presence of the activators (Fru6P, Pyr). In other words, those mutations made the activators unable to increase the affinity of ATP for the enzyme. Compared to the wild-type, the ATP $S_{0.5}$ of R11D increased 42.4- and 8.7-fold in the presence of Fru6P and Pyr, respectively. Similarly, for D141R, the ATP $S_{0.5}$ in the presence of Fru6P and Pyr increased up to 8.7- and 2.5- fold, respectively (Figure 19).

In the absence of any activator, the double mutant R11D/D141R had an identical V_{max} compared to wild-type, and the ATP $S_{0.5}$ increased by only ~2-fold, whereas, in presence of activators (Fru6P, Pyr), the ATP $S_{0.5}$ was similar compared to wild-type (Figure 19, and Figure 20). The ability of D141R to "rescue" the regulatory deficiencies of R11D and vice versa is evident at low concentrations of substrate since the effect is on the apparent affinity.

In the presence of saturating concentration of ATP, the Fru6P $A_{0.5}$ value for the R11D/D141R lowered by 3.9-fold, and it showed sensitivity to pyruvate. At a sub-saturating concentration of ATP, the wild-type enzyme is activated 8.4-fold by Fru6P (Figure 18). R11D and D141R are activated only 1.78- and 1.98-fold, respectively. The double mutant had the ability to partially recover it to 3.9-fold activation. A similar rescue effect was observed for pyruvate. The wild-type enzyme, R11D, D141R, and the double mutants were activated 6.4-, 1.17-, 1.76- and 3.82-fold, respectively (Figure 18)

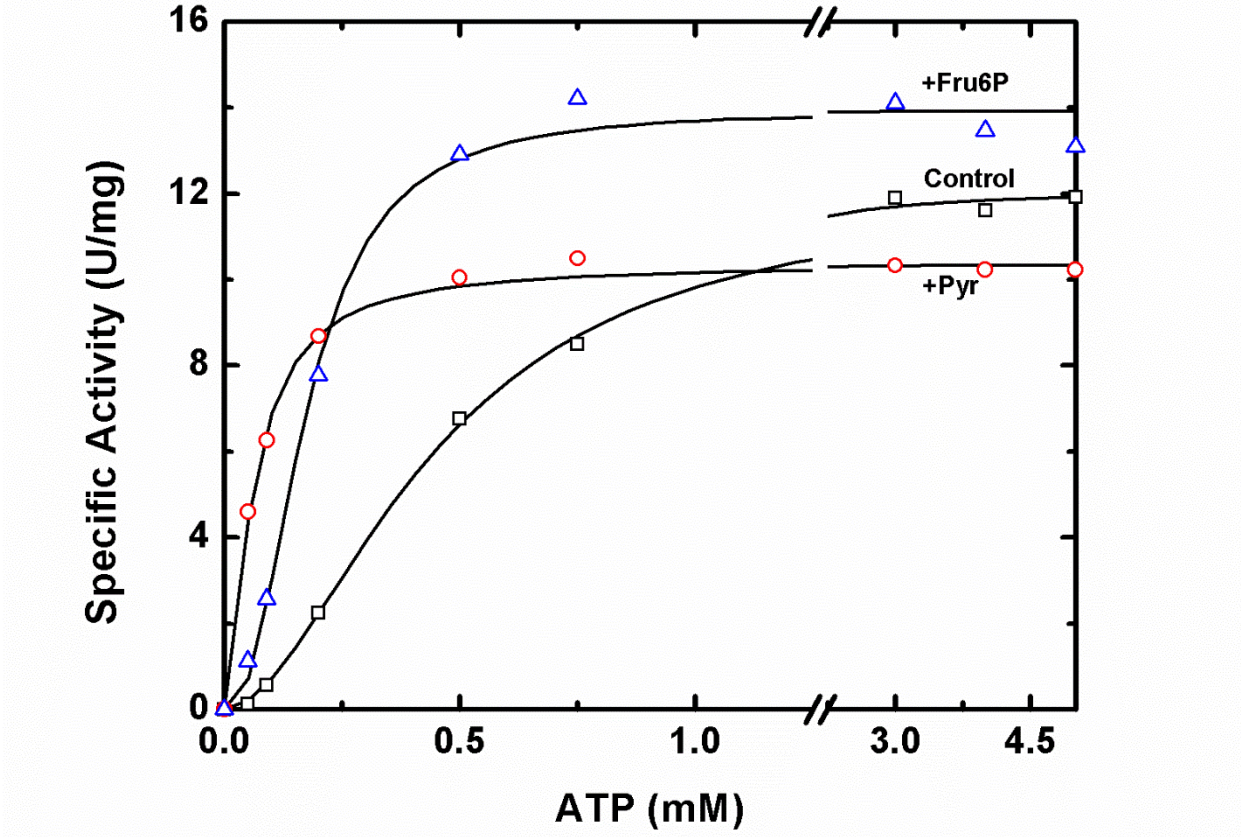


Figure 19. Substrate saturation curve of *A. tumefaciens* ADP-Glc PPase double mutant R11D/D141R. The ATP saturation curve for the R11D/D141R double mutant ADP-Glc PPase was conducted in the absence of activator (control), in the presence of 1.5 mM Pyruvate, and in the presence of 1.5 mM Fru6P. The substrate saturation assays were performed as described in the Materials and Methods (Assay A)

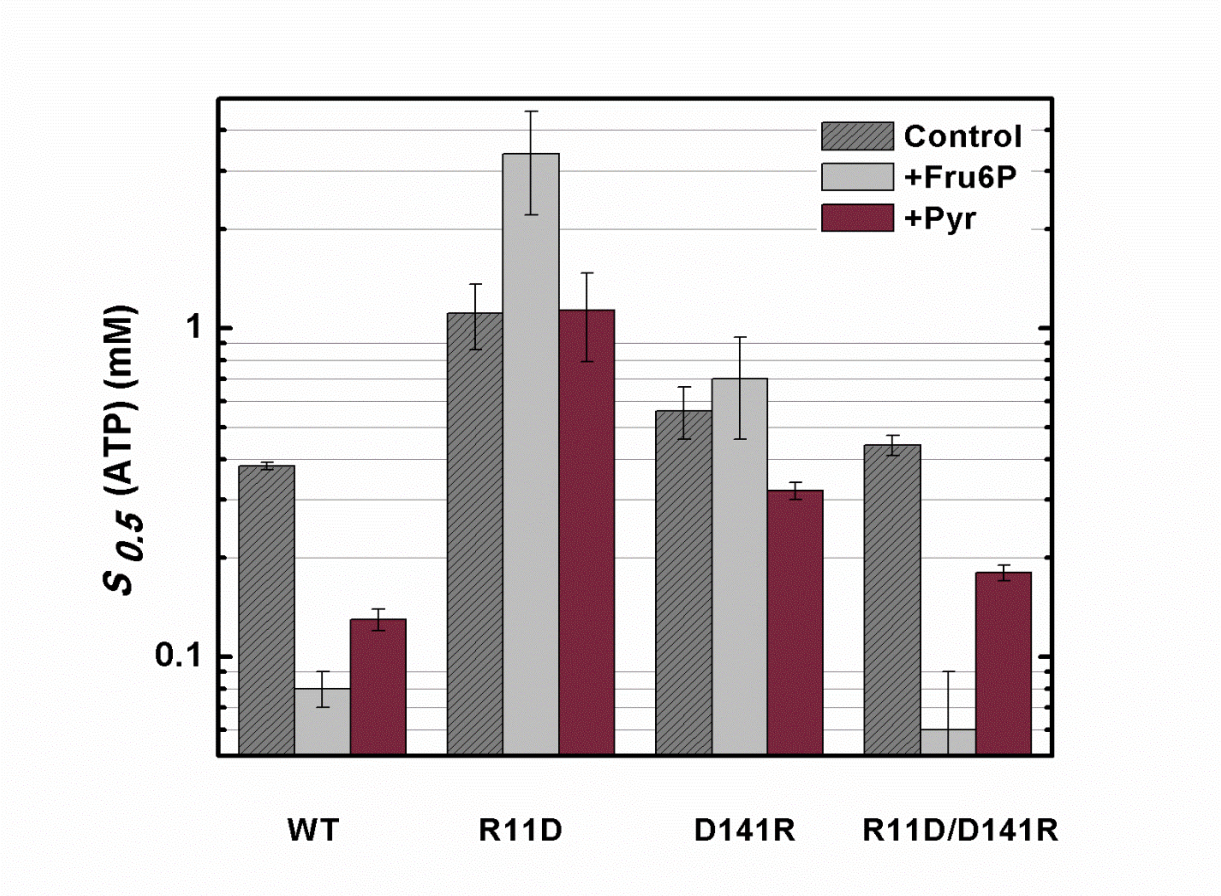


Figure 20. Effect of Fru6P and Pyr on the apparent affinities for ATP in the wild-type (WT), and mutant (R11D, D141R, R11D/D141R), *A. tumefaciens* ADP-Glc PPases. The apparent affinity ($S_{0.5}$) for ATP was calculated as described in Materials, and Methods. Assays have been performed as stated under Materials and Methods (Assay A). The control shows the data without any activators present, whereas the other +Fru6P, and +Pyr were assayed in the presence of 1.5 mM of Fructose6-Phosphate, and Pyruvate, respectively.

Discussion

Structural comparison between plant, and bacterial ADP-Glc PPase

The bacterial (homotetramer) and plant (heterotetramer) ADP-Glc PPases catalyzes the same allosterically regulated step in glycogen biosynthesis, and starch biosynthesis, respectively. The similarity between bacterial and plant enzyme allows us to study the less complex oligomeric structure of bacteria to give us insights for the plant [83]. The study of the specific amino acids involved in interface inter-subunit interaction provided important insights into structural stability and regulation of the enzyme.

The ADP-Glc PPase of *A. tumefaciens* functions as a tetramer, which can be represented as a dimer of dimers. For convenience, the dimer subunit interactions are divided into three groups: NN-1 (α2- α3 or α1- α4), NN-2 (α2- α4 or α1- α3), and CC (α2- α1 or α3- α4) (Figure 21) [85]. The NN refers to N-terminal to N-terminal interaction which has two possible interaction (NN-1 and NN-2). Similarly, CC refers to C-terminal and C-terminal interaction which has only one possible interaction.

Our studies are mainly about the NN-1 dimer interface interaction of *A. tumefaciens* ADP-Glc PPase, which is similar to the dimer of the potato tuber enzyme which forms disulfide bonds between C12 residues of A, and A' (of α subunits, the β subunits do not have disulfide bond, though it has no conserved C12 residue) [70]. Removal of the disulfide bond between C12 residues either by mutation or by reduction results in an enzyme that is nearly constitutively active [29]. In the previous studies of potato tuber ADP-Glc PPase researchers have seen that inter-subunit motion may also play a role in the allosteric regulation of the enzyme, and the crystal structure study of the potato tuber enzyme indicates that the inter-subunit interaction between the α subunits is an important part of the allosteric mechanism [70]. Overall the disulfide bond between α subunits of potato tuber keeps the two catalytic

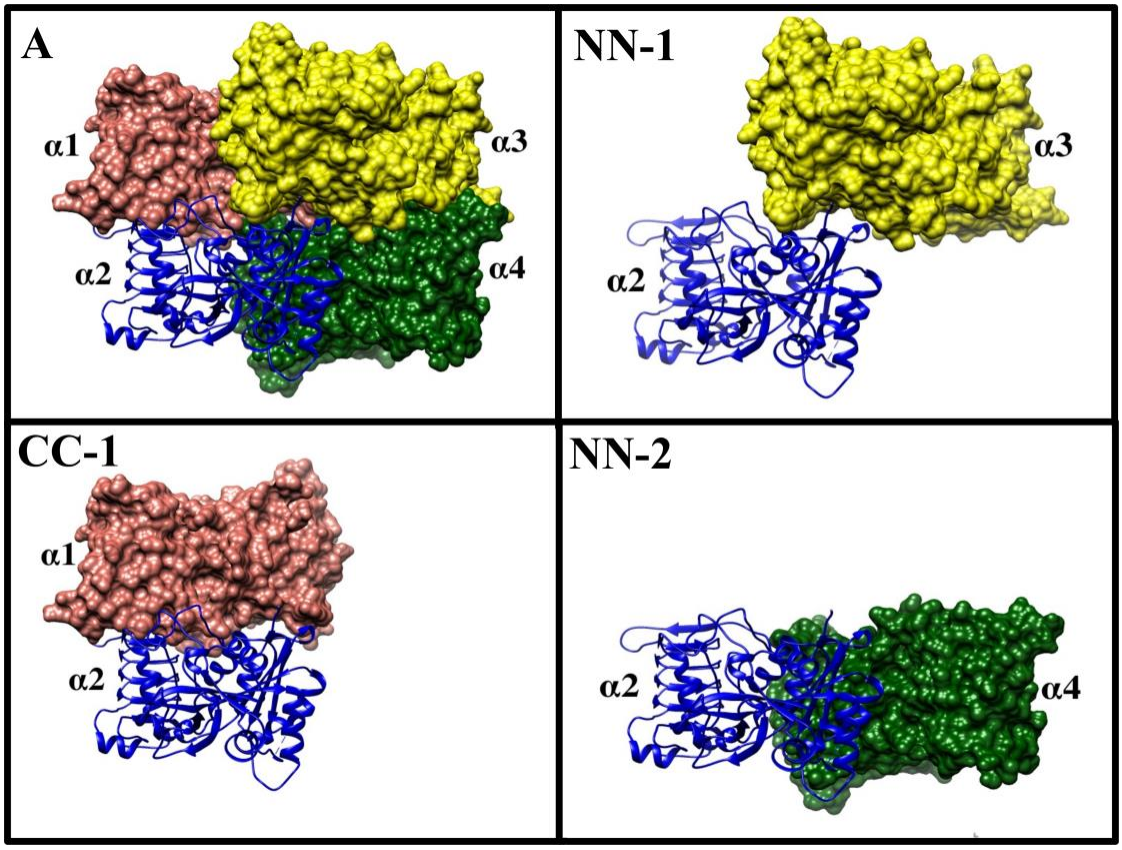


Figure 21. The homotetramer and the dimer interface interaction in ADP-Glc PPase *A. tumefaciens*. The homotetramer with labeled subunits was shown in (A), the C terminal to C terminal subunit interaction was shown in (CC-1), the N-terminal to N-terminal has two different interaction which are shown in (NN-1) and (NN-2).

dimers in their inactive form. The NN-1 dimer of *A. tumefaciens* tetramer does not have a disulfide bond at the interface, but it rather had salt bridge between R11 and D141.

According to our analysis, in the crystal structure of the *A. tumefaciens* ADP-Glc PPase [108], within one dimer α2- α3 of NN-1 group, R11 of α2 makes a salt bridge with D141 of a neighboring subunit α3. In the same interface, D141 of the α2 subunit makes another salt bridge with R11 of the α3 subunit (Figure 15). Similarly, the other dimer α1-α4 of NN-1 group makes another two salt bridges between R11, and D141. Therefore, in the homotetramer structure of the ADP-Glc PPase of *A. tumefaciens*, four R11 and four D141 were making a total of eight hydrogen bonds. The conserved D151 and R19 of the potato tuber small subunit are found to be in the similar arrangement as of the D141 and R11 of *A. tumefaciens* ADP-Glc PPase. This analysis leads us to the hypothesis that may be potato tuber ADP-Glc PPase enzyme has a similar mechanism of activation.

The dimer interactions NN-2, and CC-1 from *A. tumefaciens* is identical to the head-to-head and the tail-to-tail interaction, respectively, between the large subunit (SH2) and the small subunit (BT2) from maize endosperm [125]. Overall, the head-to-head and tail-to-tail interactions of maize endosperm are very important for the allosteric properties, and specifically affinity for 3-PGA and Pi [125]. The mutation (H333Y) in the large subunit (SH2) increased the stability of the maize endosperm enzyme through the enhanced subunit interaction with the small subunit (BT2) [117, 129].

In the CC dimer interface of potato tuber small subunits, the Y312, and Y317 were making pi-pi interactions. The Y315 residue from maize has been studied before, which is a homologous residue in potato. The Sdevo355 (Y315C, K295E) mutant of small subunit along with wild-type as well as the mutant large subunit gives the up-regulation in maize. According to their studies, the synergy between large and small subunits is important for the regulation of the enzyme [123, 124]. The previous studies

give major insights on the importance of the subunit interactions in the regulation as well as in the stability of the ADP-Glc PPase structure.

The effect of mutations at interface NN-1

The mutation at the interface of R11 and D141 residues affects the allosteric regulation as well as the structural stability of the ADP-Glc PPase. The single mutation (H333Y) increased the stability of the maize endosperm enzyme through the subunit interaction, which was established previously. According to our molecular weight-based analysis the D141R mutant enzyme was a homodimer, in contrast with other homotetramer mutant enzymes (Figure 16). In the mutation D141R, we can say that the two positive charges facing each other destroyed the interface interaction with R11, and resulted in a homodimer. Overall, our results support the statement that the structural stability of the enzyme depends on the interface interaction. The residues R11 and D141 are involved in the inter-subunit surface interactions, which can hold the enzyme in its proper position to become activated. We can hypothesize that after the cleavage of the disulfide bond between C12 in potato tuber ADP-Glc PPase, the conserved residues R19 and D151 are coming together to stabilize the structure. The conserved mutation R11K, and D141E restored the activation by Fru6P and pyruvate which shows the charges are important.

At the saturating concentration of ATP, Fru6P and pyruvate does not activate the mutant enzymes R11D and D141R compared to the wild-type enzyme. On the other hand, at the sub-saturating concentration of ATP, the R11D and D141R mutants were slightly activated by both activators (Fru6P, and pyruvate) (Figure 18). It gave us the idea of activation by activators at a really low concentration of substrate ATP. Though, we cannot avoid the fact that the mutant R11D and D141R enzymes still have a lower activation fold (1.78- and 1.98-fold, respectively) than the wild-type (8.4-fold) for Fru6P.

Similarly, for pyruvate, the R11D and D141R mutants have lower activation fold (1.17- and 1.76-fold, respectively) relative to the wild-type (6.4-fold). Very gratifying was our effort to rescue the activation by performing the double mutation R11D/D141R. At sub-saturating concentrations, the double mutant R11D/D141R restores the partial activation with the Fru6P and pyruvate with 3.9- and 3.8-fold activation, respectively, compared to wild-type (8.4- and 6.4-fold, respectively). These results indicate that single mutations R11D and D141R in the enzyme dramatically lowered the effect of the activators. In contrast, the double mutant enzyme R11D/ D141R restores the analogous effects of activators compared to wild-type. Overall, we can say that it is not necessary to have Arg at the11 position or Asp at the 141 position, but the charges are important to make the interactions. If the R11 to D141 interaction is not present, it also affects the affinity for the activators.

Our results clearly show the involvement of the R11 and D141 in the inter-subunit surface interactions. The disturbance in this interaction affects the stability of the structure, and it also affects the allosteric regulation. These data support the explanation from previous studies, which states that the extreme N-terminal R11 is quite clearly involved in activation by pyruvate in the case of *A. tumefaciens* ADP-Glc PPase [109]. Our studies provide more detailed information about the inter-subunit surface interaction related to the stability and the allosteric regulation of the ADP-Glc PPase of *A. tumefaciens*. This study also provides insights into the manipulation of the enzyme for increased starch production.

Materials and methods

Site-Directed Mutagenesis

The PCR for site-directed mutagenesis was performed using a Q5 Site-Directed Mutagenesis Kit (New England Biolab). The pet28c vector containing the *A. tumefaciens* ADP-Glc PPase was used as a template. The oligonucleotides for mutations were synthesized by Integrated DNA Technologies

(IDT). The pairs of primers that were used for generating the mutations D141A, D141E, D141N, D141R, R11D and R11K in *A. tumefaciens* ADP-Glc PPase were as follows: For D141A, forward, 5'-CAT ATT TAC AAA ATG GCC TAC GAA TAC -3' and reverse 5'-GTC GCC GGC CAG AAT GAC CAT-3' ; For D141E, forward, 5'- CAT ATT TAC AAA ATG GAG TAC GAA TAC ATG CTG C -3' and reverse 5'-GTC GCC GGC CAG AAT GAC CAT-3' ; For D141N, forward, 5'- CAT ATT TAC AAA ATG AAC TAC GAA TAC ATG CTG C -3' and reverse 5'-GTC GCC GGC CAG AAT GAC CAT-3' ; For D141R, forward, 5'- CAT ATT TAC AAA ATG CGC TAC GAA TAC ATG CTG -3' and reverse 5'-GTC GCC GGC CAG AAT GAC CAT-3' ; For R11D, forward, 5'-GCG GAT GAT GCA ATG GCC TAT GTC CTC -3' and reverse 5'-CAA AGG CTG AAC TCT TTT TTC CGA CAT-3'; For R11K, forward, 5'-GCG AAG GAT GCA ATG GCC TAT GTC CTC G-3' and reverse 5'-CAA AGG CTG AAC TCT TTT TTC CGA CAT-3'. The mutations were verified by genetic sequencing performed by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility in Chicago, Illinois.

Expression and Purification

The wild-type and the mutants were expressed in *Escherichia coli* BL21 DE3 cells. The transformations were performed using the wild-type, and each different mutant genes. The transformation mixes were plated on selection plates treated with X-gal for blue-white colony selection. White colonies were screened for the correct ligation of the mutated gene into the pet28c vector using colony PCR. The Bl21 cells with correct plasmids were grown to an OD_{600} between 1.1-1.3 at 37ºC and cooled on ice. The culture was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 25ºC for 16 hours with shaking at 250 rpm. The cells were harvested by centrifugation and sonicated in buffer C (50 mM HEPES (pH 7.5), 10% glycerol, 200 mM NaCl, and 10 mM imidazole). Crude

extracts were loaded onto a pre-equilibrated 5 ml His-Trap FF column (Ni^{2+} Sepharose column) and eluted with a linear gradient 0-50% of buffer E (50 mM HEPES pH 7.5, 10% glycerol, 200 mM NaCl, and 750 mM imidazole). Active fractions were pooled, and concentrated after SDS-PAGE and enzyme assay, which was stored at -80 ^{o}C in buffer E in aliquots and used for kinetic characterization.

The concentrated proteins were used for further purification using equilibrated gel filtration column (Superdex^{TM}200, 10/300 GL) with buffer X (50 mM HEPES (pH7.5), 5% glycerol, and 200 mM NaCl). The list of the markers that were used is as follows; Sucrose synthase from *Nitrosomonas europaea*of 360 kDa; ADP-Glc PPase from *Agrobacterium tumefaciens* of 196 kDa; Gab-R from *Bacillus Subtilis* of 112 kDa; Pyrophosphatase from *Yeast* of 71 kDa; Bovine serum albumin from

Bovine plasma of 66 kDa; and Glycogen synthase from *Escherichia coli*of 50 kDa. The retention volume of the D141R and the WT is shown in Figure 22.

Enzyme assay

MG-AM assay: Malachite Green-Ammonium Molybdate solution is prepared by using 1.5% MG solution, and 34 mM Ammonium Molybdate. Tween – 20 is used to maintain the color. The total reaction volume was 50 µl, from which 30 µl was premix, 10 µl of enzyme, and 10 µl of the variant (which in case of the activator saturation curve were activators or in the substrate saturation curve were substrate).

Substrate saturation assay. We used the various concentrations of substrate (ATP) for this assay. The premix used for control reaction contains 50 mM HEPES (pH-7.5), 14 mM $MgCl_2$, 1.5 mM glucose 1-phosphate, 0.005 U/µL pyrophosphatase, and 0.2 mg/ml BSA. The reaction starts with 10µl of the enzyme dilution. We added the MG-AM solution, after incubation of 10 minutes to bind with PPi formed in the reaction; and we measured the data using a spectrophotometer at 595 nm absorbance.

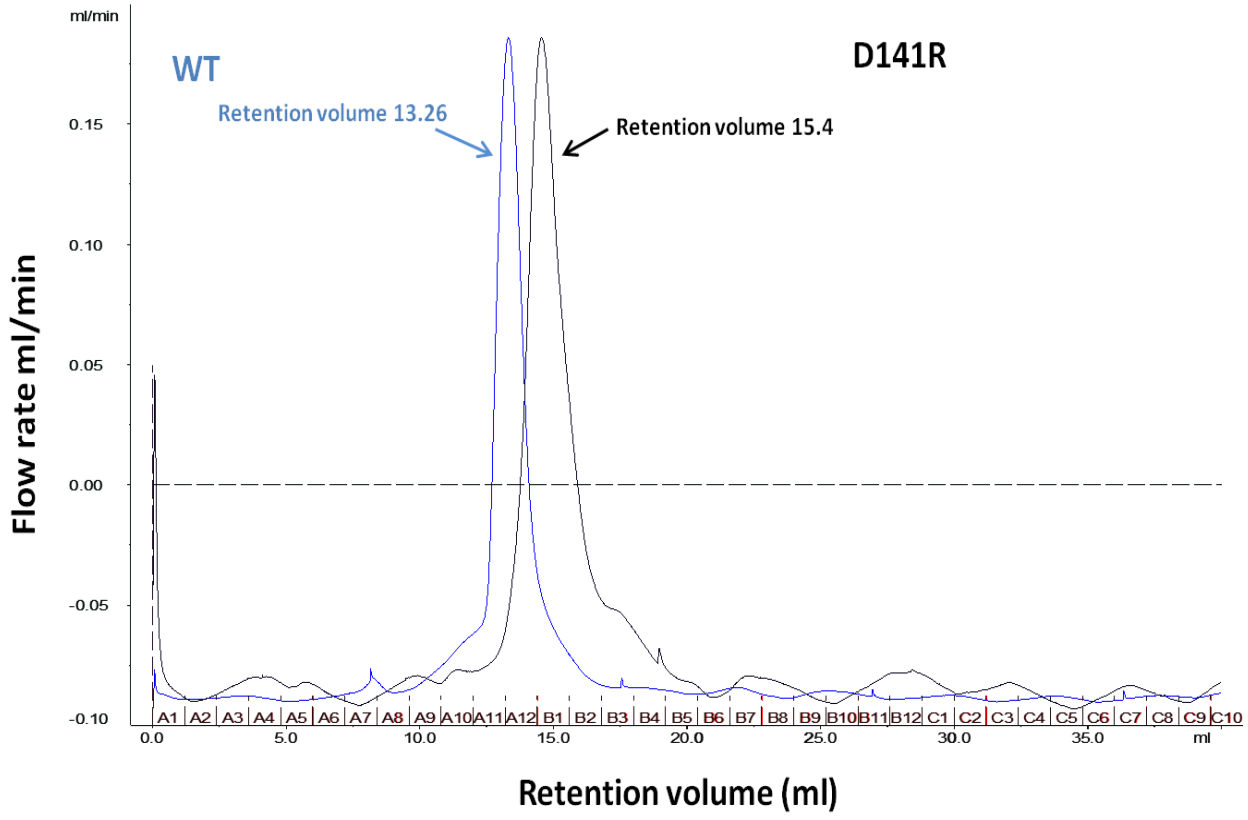


Figure 22. Gel filtration chromatography of wild-type (WT) and the D141R mutant ADP-Glc PPase. The retention volume for the wild-type enzyme based on molecular weight (198 kDa) was 13.26 ml (Blue line). While the retention volume for the D141R was 15.4 ml (Black line). The details are described in the Methods and Materials.

We did the same reaction with the presence of 1.5 mM activators (Fru6P, and Pyruvate) in the premixes.

Activator saturation assay. The premix used for this reaction contains 50 mM HEPES (pH-7.5), 7 mM $MgCl_2$, 1.5 mM glucose 1-phosphate, 1.5 mM ATP, 0.005 U/µL pyrophosphatase, and 0.2 mg/ml BSA. The reaction starts with 10µl of the enzyme dilution. After incubation of 10 minutes the MG-AM solution is added to bind with PPi formed in the reaction, the data are observed with a spectrophotometer at 595nm absorbance.

Kinetics

$S_{0.5}$and $A_{0.5}$ values indicate the concentration of substrate or activator needed to give 50% of the maximal velocity. The Hill coefficient (*nH*) was calculated from fitting data in Origin™ 6.0 to the non-linear least square formula, by using the Hill equation $V= V_0 + [(Vmax*S^n) / (K_m+ S^n)]$ for activator saturation plots, and for substrate saturation plots the Hill equation $V= (V_{max}*S^n) / (K_m+ S^n)$ was used. All values were the average of duplicate reactions with errors of 10% or more. We used these plots to calculate the substrate concentration ($S_{0.5}$), and activator concentration ($A_{0.5}$), which gives 50% of the maximum velocity (V_{max}) and the Hill number (n_H). Catalytic efficiency (k_{cat}/K_m) (*Ce*) was calculated with Origin™ 6.0, and using the Hill equation $V=Ce*V_{max}*S^n/ (K_m+ S^n)$.

Note: For all enzyme assays, standard PPi curves were performed, which are used to calculate the PPi formed in the reaction.

CHAPTER FOUR

CONCLUSIONS AND FUTURE IMPLICATIONS

Understanding the allosteric regulation of ADP-Glc PPase

In this thesis, we mainly focused on the structure-function relationship of the enzyme ADP-Glc PPase from *A. tumefaciens*. Our second chapter focuses on the allosteric regulation of this enzyme by pyruvate. According to the recently-solved crystal structure of the pyruvate-bound P96A mutant ADP-Glc PPase (PDB: 5W5R), the pyruvate binds at the C-terminal interface of two neighboring subunits [126]. Based on the structural data, the pyruvate molecule is stacked between the S328 and G329 residues from two subunits [126]. The G329 residue of *A. tumefaciens* ADP-Glc PPase is the homologue residue to the G336 of the *E-coli* ADP-Glc PPase. Interestingly, the G336D mutant of the *Escherichia coli* ADP-Glc PPase has been shown to exist in a pre-activated state [114]. Expression of this G336D mutant in potato tuber resulted in significantly increased starch production [30]. We characterized the homologous mutant G329D in *A. tumefaciens* ADP-Glc PPase, which we found pre-activated in a similar manner [126]. In addition, we found two K43 residues from two adjacent subunits that make hydrogen bonds with the oxygen atoms of the pyruvate. According to the analysis of site-directed mutants, K43 and G329 are essential in the regulation of the ADP-Glc PPase and control of the glycogen synthesis pathway.

An important conclusion is that in the G329D *A. tumefaciens* ADP-Glc PPase, the aspartate residue introduced mimics the presence of pyruvate [126]. In addition, the K43A and K43N was insensitive to pyruvate and confirmed the role of lysine in regulation. In addition, the double mutation

K43A/G329D provides further support to the importance of lysine in triggering the pyruvate activation.

Overall, K43 is an important residue, which is not only involved in binding, but also in triggering the activation signal. In our conclusion, a positive charge is important at position 43 for the binding of pyruvate in ADP-Glc PPase of *A. tumefaciens*. The structure comparison of mutant K43A and WT provides support to our conclusion that the changes in pyruvate regulation are not caused by a structural alteration (Table 2).

A future goal of this project is to accumulate more information for the specific residues involved in the allosteric regulation. This is critical to understand how the regulatory site has been evolving in different species to accommodate different metabolic needs. If we understand the architecture of the regulatory site in different species, it would be possible to manipulate them for biotechnological purposes. The ongoing research will try to replicate the pyruvate binding in the heterotetramer ADP-Glc PPase from *Ostreococcus tauri*. We did the sequence alignment of the *Ostreococcus tauri* and *A. tumefaciens*. As a result, we found N34 (in the position of the conserved K43) and D317 (in the position of the conserved G329) in the small subunit of that algal species (*O. tauri*). Similarly, in the large subunit we found A52 (replacing K43) and H341 (replacing G329). To mimic the pyruvate binding in the ADP-Glc PPase of *Ostreococcus tauri*, we will make the mutation N34K in the small subunit and A52K-H341D double mutation in the large subunit. The further kinetic characterization of the mutant enzymes might provide important information for regulation of ADP-Glc PPase. A hyperactive enzyme in unicellular algae (or any other photosynthetic eukaryote) will be of great significance for the enhanced production of starch and biofuels.

Interface inter-subunit interaction in regulation

Based on structural analysis, R11 and D141 residues are involved in the inter-subunit interaction at one of the dimer interfaces (NN-1) in the *A. tumefaciens* ADP-Glc PPase (Figure 15, Figure 21). Previous studies found that mutation of the R11 residue to alanine loses the activation by pyruvate [109]. According to our mutagenesis studies for R11 and D141, we conclude that this inter-subunit surface interaction holds the enzyme in a proper position for activation. Our study provides detailed information about the inter-subunit surface interaction related to the stability and the allosteric regulation of the ADP-Glc PPase of *A. tumefaciens*. In the mutation D141R, we can say that the two positive charges facing each other destroyed the interface interaction with R11, and formed a homodimer. The R11 and D141 interface interaction is also involved in allosteric regulation. For example, the R11A mutant enzyme lost the activation by pyruvate, whereas the D141A mutant enzyme lost the activation by Fru6P. From our results based on the conservative mutations R11K and D141E, we conclude that it is not essential to have Asp or Arg at these positions, but the charges and hydrogen bonds are important.

At the same dimer interface NN-1 of the potato tuber enzyme there is a disulfide bond present between the two Cys12 from neighboring subunits [70]. The mutation or reduction of the disulfide bond resulted in the constitutively active form of the enzyme [29]. Overall, the C12 disulfide bond keeps the two catalytic domains apart from each other in the potato tuber ADP-Glc PPase. The conservative D151 and R19 of the potato tuber small subunit are found to be in the similar arrangement as of the D141 and R11 of *A. tumefaciens* ADP-Glc PPase. We can hypothesize that after the breakage of the disulfide bond between C12 in potato tuber ADP-Glc PPase, the conserved residues R19 and D151 are coming together to form the active form of the enzyme. In a future project, we will perform mutagenesis studies for the R19 and D151 residues of the potato tuber ADP-Glc PPase to prove this

hypothesis.

The provided structure information about the ADP-Glc PPase will be important for the protein engineering in future.

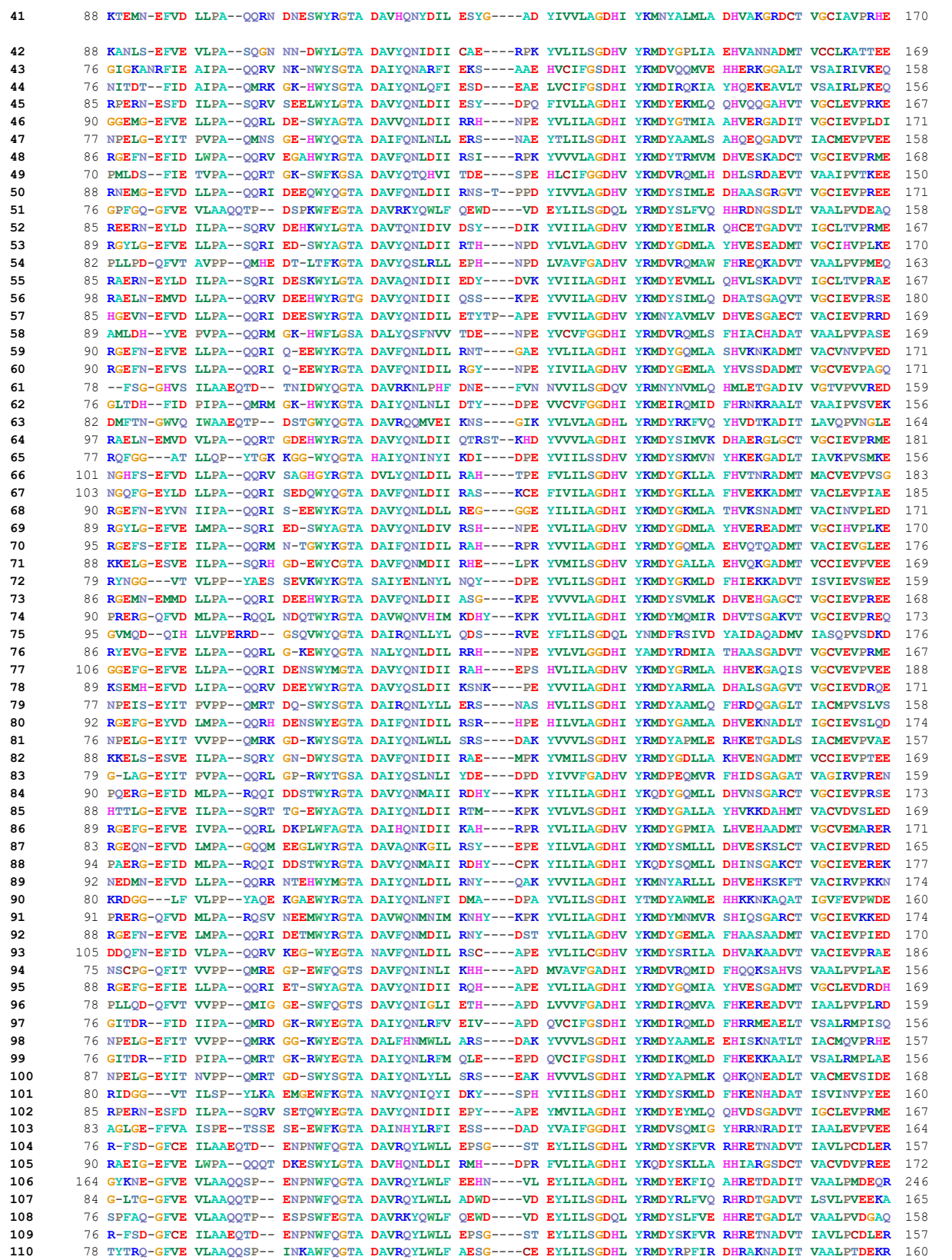
APPENDIX A

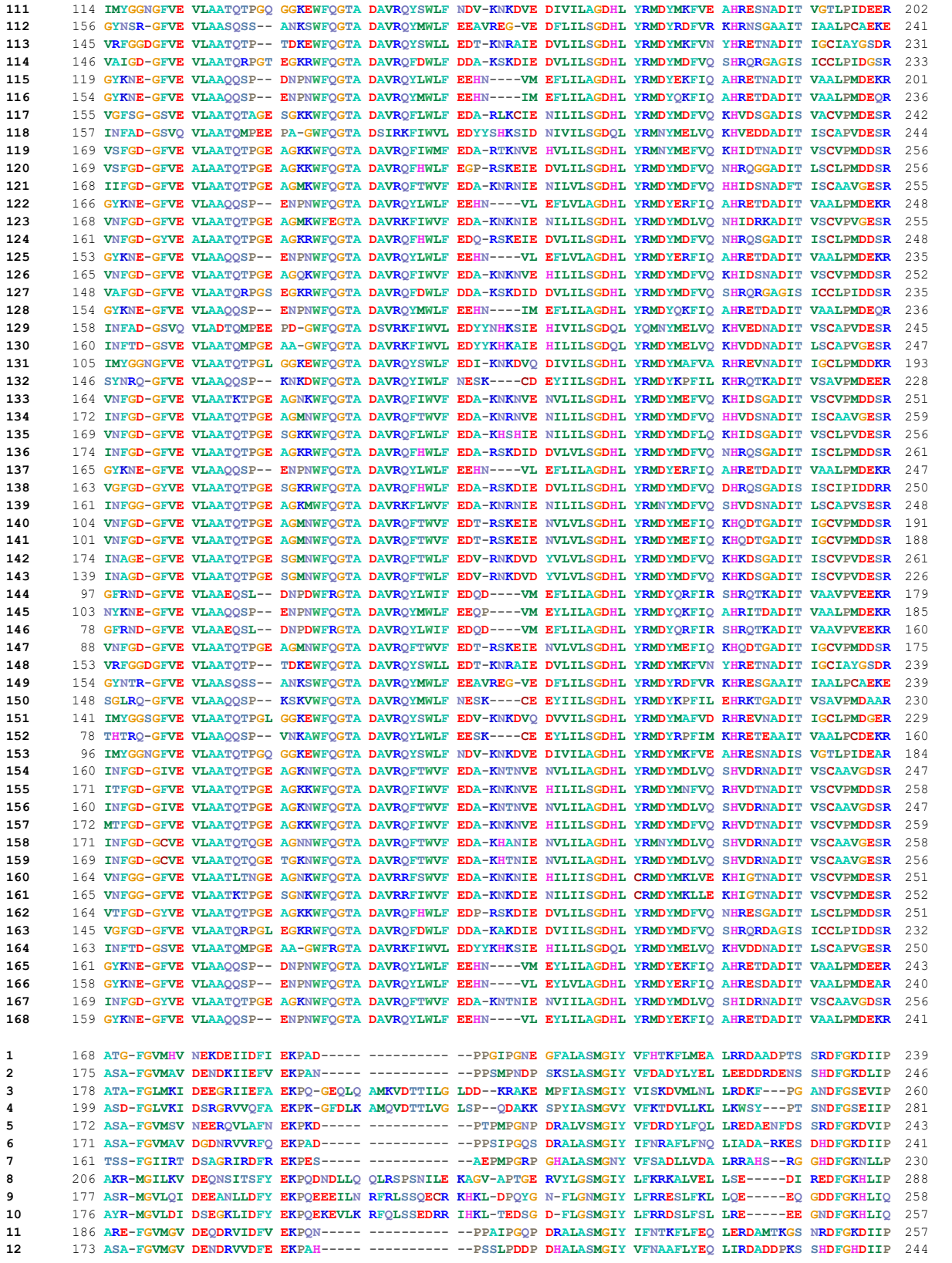
SEQUENCE ALIGNMENT OF PLANTS AND BACTERIAL ADP-GLC PPASE

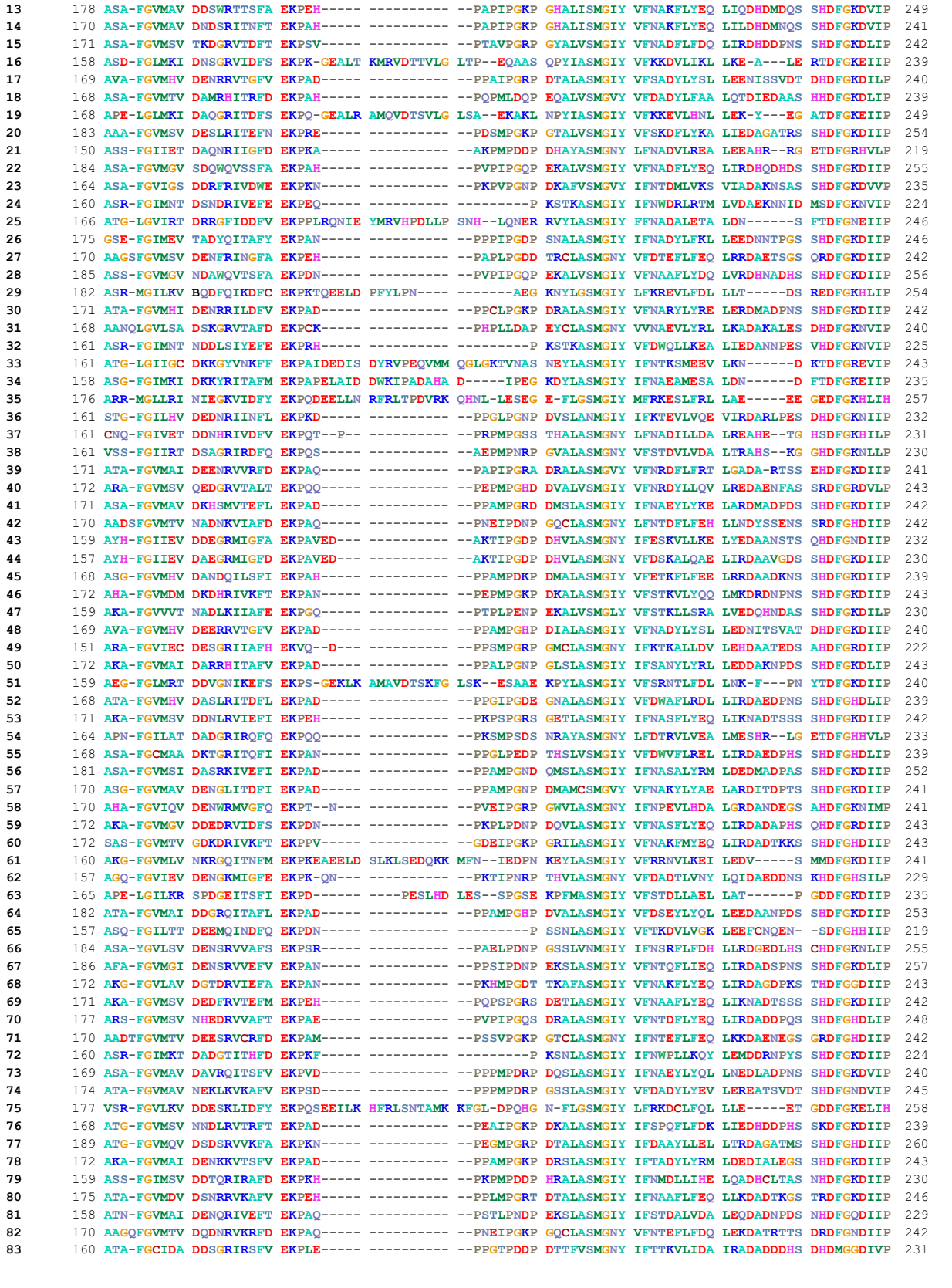

1 4 --------RV QPLARDAMAY VLAGGRGSRL KELTDRRAKP AVYFGGKARI IDFALSNALN SGIRRIGVAT QYKAHSLIRH LQRG--WDFF 84
2 11 --------LA RQLPLKSVAL ILAGGRGTRL KDLTNKRAKP AVHFGGKFRI IDFALSNCIN SGIRRMGVIT QYQSHTLVQH IQRG--WSFF 91
3 5 DSQNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG 94
4 26 P----RLERR RANPKDVAAV ILGGGEGTKL FPLTSRTATP AVPVGGCYRL IDIPMSNCIN SAINKIFVLT QYNSAPLNRH IARTY-FGNG 110
5 7 -------RFV SRLTRDTLAL IMAGGRGGRL SNLTDWRTKP AVPFGGKFRL IDFPLSNCIN SGIRRIEVLT QYKAHSLIQH IQRG--WGFL 88
6 7 -------RFV SRLTRNTLVL ILAGGRGSRL MDLTTWRAKP AVPFGGKFRI IDFTLSNCIN SGIRRIGVLT QYKAHSLIRH LRLG--WGSL 88
7 1 ---------- MIAGKSVLAF VMAGGEGSRL HPLTAERSKP SVPFNGRHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH TRRA--WVMS 78
8 37 SQVVETSDSQ SVDMRQVASL ILSGGEGTRL HPLTLARCKP AINFGGKYRL IDVPISNSLH AGCKKVFLLT QFLSSSLHQH VFQTYMQ--G 124
9 7 QGYPPNYQAS HFYRDKVGVI VLCGGEGKRL SPLTCWRCKP TVSFGGRYKL IDVPISHAIA SGFSKIFVIG QYLTYTLQQH IVKTYFY--H 94
10 7 P-EASNFESS HFYRDKVGVI ILCGGEGKRL SPLTNCRCKP TVSFGGRYKL IDIPISHAIS AGFSKIFVIG QYLTYTLQQH LFKTYFY--H 93
11 15 MREQTDMRFI SHLTRNTFAI ILAGGRGTRL KQLTDFRSKP AVPFAGKFRI LDFTLSNCVN SGIRKIGVAT QYKAHSLIRH IQRG--WSFL 102
12 5 ---LSSSRFV STLTKNTVAL ILAGGKGSRL RDLTNWTAKP AVPFGGKFRI IDFPLSNCIN SGVRRIGVVT QYKAHTLIQH IQRG--WGFL 90
13 7 VQTNDNPRFV STLTRNTLAL ILAGGRGTRL KNLTDWRAKP AVPFGGKFRI IDFTLSNCVN SGVRRIGVVT QYKAQSLIRH IQRG--WSFL 94
14 1 --MNNNPRFV STLTRNTLAL ILAGGRGTRL KNLTDWRAKP AVPFGGKFRI IDFTLSNCVN SGVRRIGVVT QYKAQSLIRH IQRG--WSFL 86
15 1 MKLPADLRSY SQITRNSIAM ILAGGRGTRL RQLTDWRAKP AVPFGGKFRI IDFPLSNCVN SGIRRIGVAT QYKAQSLISH IQQG--WGFL 88
16 1 ---------- ---MKKVLAI ILGGGAGTRL YPLTKLRAKP AVPVAGKYRL IDIPVSNCIN SEIFKIYVLT QFNSASLNRH IARTYN--FS 75
17 5 --------RL NDLQRTTLAI VLAGGRGTRL GPLTNKRVKP AVHFGGKYRI IDFALSNCLN SGIRRIAVVT QYKAHSLLRH LQRG--WSFL 85
18 4 --------AH HRPVRRTISL VLAGGRGSRL QDLTENCAKP AVHFGGKFRI IDFVLSNCVN SGLHRIGVLT QYKSHSLLRH LQHG--WSFL 84
19 1 ---MCCWQSR GLLVKRVLAI ILGGGAGTRL YPLTKLRAKP AVPLAGKYRL IDIPVSNCIN SEIVKIYVLT QFNSASLNRH ISRAYN--FS 85
20 12 YRYYTEPTLV TELTRKTLAL VLAGGEGSRL KDLTAWRAKP AVPIGGKYRI IDFPLSNCVN SGIRRIGVLT QYKSHSLIRH LQRA--WGLM 99
21 1 ---------- ---------- -MAGGEGSRL QPLTADRCKP AVPFGSRYRI VDFVLSNLTN SDIRSIYVLV QYRPQSLIEH VRKA--WTVT 67
22 13 AEYKSSSRFH SNISHETLAL ILAGGRGSRL HQLTDWRAKP AVPFGGKFRI IDFPLSNCVN SGFRRIGVVT QYKAQSLIRH IQHG--WSFL 100
23 1 -------MNL RSEMKKTLCI LMAGGKGERL YPLTKARAKP SVRFGGIYRI VDFTLSNCLN SDIRRVYVLT QYRSVSLDRH IRLG--WNIF 81
24 1 ---------- --MKNEMLAL ILAGGQGTRL GKLTQSIAKP AVQFGGRYRI IDFALSNCAN SGINNVGIIT QYQPLALNSH IGNGSSWGLD 78
25 1 -----MRRRV DPGSNDVLSI ILGGGKGSRL YPLTKDRAKP AVPFGGKYRL VDIPISNSIN SDFKKIYILT QFNSASLHLH LSSTYLFDT- 84
26 11 --------LA RQLPNKTVAL ILAGGRGSRL KDLTATRAKP AVHFGGKFRI IDFALSNCLN SGVRRIGVIT QYQSHTLVQH IQRG--WSFL 91
27 7 --------YI SNLTRDTYAL ILAGGRGSRL HELTTWRAKP ALYFGGKFRI IDFPLSNCIN SGIRRVGVVT QYKSHSLIRH LVRG--WGHF 87
28 14 AECTTSTRFH NSVTYNTLAL ILAGGRGSRL HQLTDWRAKP AVPFGGKFRI IDFPLSNCVN SGVRRIGVVT QYKSQSLIRH IQHG--WSFL 101
29 12 TPLTQTINLH THRTDRVASI ILGGGEGVRL FPLTLSRCKP AIPVGGRYRL IDFSISNSLN SGYQKIFILT QFLSSSLHQH IFRTYQF--D 99
30 7 --------SS FDLPRRSIAL VLAGGRGTRL KNLTDNRAKP AVYFGGKFRI VDFALSNCLN SGIRRIGVIT QYKSHSLLRH LQRG--WAFL 87
31 6 ---------L QQIIENTMVL VLAGGQGSRL KALTETRAKP VLEFGSHCRI IDFPLSNCVN SGLNRIAVLT QYKSQCLIRH LMQH--WGTL 85
32 1 ---------- -MRKREVVAM ILAGGQGSRL GVLTQNLAKP AVPFGGRYRI IDFTLSNCSN SGIYTVGVLT QYMPMELHTH IGVGSPWDLD 78
33 1 -----MKQ-- ---KNRAIAI VLGGGKGTRL YPLTMDRSKP AVPFAGKYRL VDIPISNCIN SGIRQIYILT QFNSASLHNH IANTYVFDN- 79
34 1 ---------- ---MAKVLSI ILGGGKGTRL YPLTKERSKP AVPFGGNHRI VDIPISNCIN SGFRQIYLLT QFNSASLHMH ISNAYNFDR- 76
35 6 QNHLSGGKTS YRYRDRVGVI VLCGGEGKRL SPLTDSRCKP TVSFGGRYKL IDVPISHAIS SGFSKIFVIG QYLTYTLQQH LFKTYFY--H 93
36 1 ---------- MDRLRKLTTL IMAGGRGERL FPLTREKAKP AVTFGGIYKI IDFTLSNCIN SGIRQIYVLT QYGSFSLDHH LRMA--WEVV 78
37 1 ---------- MSSTKNVLAI VMAGGEGSRL HPLTAERCKP AVPFNGKHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH IRQS--WTMT 78
38 1 ---------- MIAGKSVLAF VMAGGEGSRL HPLTAERSKP SVPFNGRHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH IRRS--WVLT 78
39 7 -------RFV SRLTRDTLAL ILAGGRGTRL HELTQWRAKP AVPFGGKFRI IDFPLSNCIN SGVRRIGVLT QYKAHSLIRH IRQG--WSSL 88
40 7 -------RFV SRLTRETLAL IMAGGRGGRL SSLTDWRTKP AIPFGGKFRL IDFPLSNCIN SGIRRVGILT QYKAHSLIQH VQRG--WGFL 88
41 7 --------QG LDLPKRAIAL VLAGGRGSRL MNLTDSRAKP AVYFGGKFRI IDFALSNCLN SGIRRIGVIT QYKSHSLLRH LQRG--WAFL 87
42 7 --------YI SNLTKDTYAL VLAGGRGSRL FELTDSRAKP AVYFGGKFRI IDFSLSNCIN SGINRIGVAT QYKSHSLIRH INRG--WGNF 87
43 1 ---------- ---MAGILAM ILAGGEGSRL FPLTQTRTKP AVPFGGNYRL VDFALNNFVN ADLMKIYVLT QFKSQSLNIH LRKA--WRLS 75
44 1 ---------- ---MAKVLSM ILAGGEGSRL YPLTQSRTKP SVPFGGSYRL VDFALNNFVN SDLLRIYVLT QFKSQSLNQH LRKA--WELN 75
45 4 --------IE PPFARHAMAY VLAGGRGSRL MELTDRRAKP AVYFGGKSRI IDFALSNALN SGIRRIAVAT QYKAHSLIRH LQHG--WNFF 84
46 8 -------RFV SRLTRETLAV ILAGGRGSRL HQLTDWRAKP AVHFGGKFRI IDFPLSNCVN SGIRRISVLT QYKSHSLDRH IQRG--WGFL 89
47 1 ---------- --MQNQTLTF LLAGGVGSRL HPLTSSRSKP SVPFGGKYRI IDFTLANCLH SGLRRILVLT QYKSHSLNKH LRDG--WSIF 76
48 5 --------SL NDLQHTTLAI VLAGGRGTRL GPLTNKRVKP AVHFGGKYRI IDFALSNCLN SGIRRIAVVT QYKAHSLLRH VQRG--WGFL 85
49 1 ---------- ---------M ILAGGEGRRL GPLTHDRAKP AVPFGGRYRI IDIVLSNFVN SGLHRIKILT QYKSASLDEH IARA--WRLS 69
50 7 --------RR RILTRRTLAL VLAGGRGSRL RDLTNVRAKP AVHFGGKFRI IDFALSNCMN SGLRRIGVIT QYKSHSLLRH LQRG--WSFL 87
51 1 ---------- ---MKRVLAI ILGGGKGSRL YPLTKMRAKP AVPLAGKYRL IDIPISNCIN SGIEKMYVLT QFNSASLNRH IGRTYN--LN 75
52 4 --------PP LRLTSQAMAF VLAGGRGSRL KELTDRRAKP AVYFGGKARI IDFALSNAMN SGIRKMAIAT QYKAHSLIRH IQRG--WNFF 84
53 8 --------FV SRLTRDTLAL ILAGGRGSRL KHLTAWRAKP AVPIGGKFRI IDFPLSNCVN SGIRRIGVLT QYKAHSLVRH IQQG--WGFM 88
54 1 -------MRK KRYGGRIIAF VMAGGEGKRL HPLTAERCKP AVPFGARYRL VDFVLSNLIN SEIRSIYLLV QYKPQALIEH IRRA--WVIS 81
55 4 --------PN QRLSSQAMAF VLAGGRGSRL KELTDRRAKP AVYFGGKTRI IDFALSNALN SGIRKMAIAT QYKAHSLIRH MQRG--WNFF 84
56 17 --------QP HMLVRRSIAL VLAGGRGSRL KQLTDKRAKP AVYFGGKFRI IDFALSNCVN SGIRRIGVIT QYKSHSLLRH LQRG--WSFL 97
57 4 --------KT FQLPRKAIAL VLAGGRGSRL HELTDRRAKP AVHFGGKFRI IDFALSNCVN SQIRRVGVVT QYKSHSLLRH LQRG--WNFL 84
58 1 MEGDYAQHIH VRPQPRVLAV VLAGGEGTRL YPLSAHRAKP AVPFGGSYRI IDFVLSNFVN SGFHRIKVLT QYKSDSLVNH ISRG--WRLS 88
59 4 ---LTSTRFI SALTKNTVAL ILAGGKGSRL KDLTNWRAKP AVPFGGKFRI IDFPLSNCMN SGVRRIGVVT QYKSHSLMQH IQRG--WGFL 89
60 4 ---MRSPRFI SALTKNTVAL VLAGGRGSRL HNLTDWNAKP AVQFGGKFRI IDFPLSNCMN SGIRRIGVVT QYKAHTLIEH IQKG--WGFL 89
61 1 ---------- -MRAFNTVAL ILGGGRGTRL YPLVKARSKP AVSLGGQYRM IDIPVSNCIN SGFRNIYVIT QFNSASLNNH IYNAYRFDN- 78
62 1 ---------- ---MGGILSM ILAGGEGTRL FPFTSLRAKP AVPFGGNYRI IDFVLNNFIN SDLLQIFVLT QFKSHSLMKH LSQA--WRIS 75
63 1 ------MEIT AMIMKDVLGL IMGGGRGTRL YPLTKKRSKP AVPLAGKYRL IDVPISNCLH SGIDKISILT QFNSVSLHRH IFQTYR---R 81
64 16 --------QA HQLVRRTIAL VLAGGRGSRL KQLTDRRAKP AVYFGGKFRI IDFALSNCLN SGIRRMAVVT QYKSHSLMRH LQRG--WSFL 96
65 1 ---------- ----METLAM VLAGGKGSRL DILSADRAKP SVPFAGKFRI IDFALSNCVN SGIYDVGILT QYLPRSLNRH IGIGKPWDLD 76
66 15 ASSARPS--V DQLTKNTYAM VLAGGQDCRL QQLTECCAKA AVPFGGKFRI IDFVLSNCMN SGIRHIGVAT QYRSQNLIQH IKRG--WSFL 100
67 15 ASGDPEPRFV SRLTKNTYAM VLAGGRGSRL HELTNWRAKP AVPFGGKFRI IDFVLSNCVN SGIRRIGVAT QYKSHSLIQH IQRG--WSFL 102

```
68    4   ---QHSDRFI STLTKNTAAI ILAGGRGSRL KNLTDWRAKP AVQFGGKFRI IDFPLSNCIN SGIRRINVAT QYKAQSLIQH LQRG--WGFL  89
69    8   --------FV SRLTRDTLAL ILAGGRGSRL KHLTAWRAKP AVPIGGKFRI IDFPLSNCVN SGVRRIGVLT QYKAHSLVRH IQQG--WGFM  88
70    7   TGVANSPRFV SNLTKNSLAL VLAGGRGSRL GPLTDWRAKP AVPFGGKFRI IDFCLSNCIN SGIRRVGVLT QYKAHSLIRH LQLG--WGFL  94
71    7   --------YI SSLTRETYAL ILAGGRGSRL HELTDWRAKP AVYFGGKHRI IDFPLSNCIN SGVRRVGIAT QYKSHSLIRH VNRA--WGHF  87
72    1   ---------- -MKK-QCVAM LLAGGKGSRL SELTKNMAKP AVSFGGKYRI IDFTLSNCSN SGIDTVGILT QYQPLELNSY IGIGSAWDLD  78
73    5   --------QS TQLVRRTIAL VLAGGRGSRL KALTDRRAKP AVYFGGKFRI VDFAMSNCVN SGIRRIGVIT QYKSHSLIRH LQRG--WSFL  85
74    7   ------LPST EEMANETLVL ILAGGRGSRL YELTDQRAKP AVYFGGGHRI IDFALSNCIN SGLLKIGVIT QYEAHSLLRH LQHG--WSFL  89
75    7   KEEQINRKRS HFYRDNVGVI VLCGGEGKRL SPLTCWRCKP TVSFGGRYKL IDVPISHAFA SEFSKIFVIG QYLTYTLQQH LFKTYFY--H  94
76    6   ---------V SELTRNTLAL VLAGGEGSRL KELTQWRAKP AVPFGGKYRI IDFVLSNCVN SDIRRIGVLT QYKSHSLIRH IQRA--WSFM  85
77   24  -------RFV SRLTKSTLAL VMAGGRGSRL GPMTQWRAKP AVPIAGKFRI IDFSLSNCIN SGIRRIGVLT QYKSHSLIQH VQKA--WNFL 105
78    8   --------LE RRLPKRAMAL ILAGGRGSRL KQLTDTRCKP AVYFGGHFRI IDFVLSNCMN SGLRRIGVLT QYKSHSLLRH LQRG--WNFL  88
79    1   ---------- --MLDKTLTI ILAGGVGSRL HPLTADRAKP AVPFGGNYRI IDFTLSNCLH SGLRRMLVLT QYKSHSLQKH LRDG--WSIF  76
80    9   ------DHNI SHLTRNTIAL ILAGGRGSRL MNMTDWRAKP AVPFGGKFRI IDFPLSNCMN SGIRKIGVLT QYKSDSLIRH IQQG--WGFL  91
81    1   ---------- ---MQDTLTV ILAGGMGARL APLTDNRAKP AVPFGGKYRI IDFTLTNCLH SGLRRILVLT QYKSHSLQKH LRDG--WSIF  75
82    7   --------YI SNLTKDTYAL ILAGGRGSRL HELTDWRAKP AVFFGGKFRI IDFPLSNCIN SGIRRVGIAT QYKSHSLIRH VNRG--WGHF  87
83    1   ---------- MREVPHVLGI VLAGGEGKRL YPLTADRAKP AVPFGGAYRL IDFVLSNLVN ARYLRICVLT QYKSHSLDRH ISQN--WRLS  78
84    7   ------VSSK YTLAKDTLVL ILAGGRGSRL HELTDKRAKP ALYFGGNRRI IDFALSNCIN SGLNHIGVIT QYAAHSLLRH LQKG--WSFL  89
85    7   --------FV SRLTRDTLAL VLAGGRGSRL YELTDSRAKP AVYFGGKFRI IDFPLSNCVN SGIRRIGVLT QYKAHSLIRH LVNG--WGSF  87
86    7   -------RFV SRLTRDTLAI ILAGGRGGRL ANLTDWRTKP ALPFGGKFRL IDFPLSNCIN SGVRRIQIVT QYKAHSLIQH VQRG--WGFL  88
87    2   --------VA SQLPKRTVAL VLAGGRGTRL HQLTDNRAKP AVYFGGKFRI IDFALSNCIN SGIRRVGVIT QYCPHSLIRH LQRG--WSFL  82
88   11  ------LPNK YELVKDTLVL ILAGGRGSRL YELTDKRAKP ALYFGGNRRI IDFALSNCLN SGLNRIGVVT QYAAHSLLRH LQNG--WSFL  93
89   11  --------LA QQLPKEAIAL VLAGGRGTRL KALTSKRAKP AVFFGGKFRI IDFTLSNCLN SGIRRIGVIT QYQSHSLVQH IQRG--WSFF  91
90    1   ---------- -MKKTECLAM ILAGGQGSRL GALTKRVAKP AVPFGGKYRI IDFPLSNCVN SGIEKVGVLT QYRPLELNQY LGSGSAWDLD  79
91    8   ------FLHN TDIAKDTLVL ILAGGRGSRL HEMTDERAKP AVYFGGNWRI IDFTLSNCLN SNLLRIGVIT QYEAHSLLRH LQHA--WSFL  90
92    3   ----EYPRFV SLLTKNTVAL ILAGGRGSRL KNLTEWRAKP AVPFAGKFRI IDFPLSNCVN SGIRRIGVIT QYKAHSLLQH LHRG--WSFL  87
93   17 YRAELAPRFI GALTKKTYAM VLAGGRGSRL QQMTDWRAKP AVPFGGKLRI IDFPLSNCIN SGIRRIGVAT QYMAHSLIHH IQRG--WSFL 104
94    1   ---------- ----MKVLAM VLAGGQGSRL HPLTAERSKP AVPFGSRYRI ADFVLSNMMN SNIRGVYLLV QYKSQSLIEH VNKA--WGYA  74
95    6   -------RFV SRLTRETLAL ILAGGRGSRL KHLTLWRAKP AVPFGGKFRI IDFPLSNCIN SGIRRVGVLT QYKAHSLIQH VQRG--WSFL  87
96    1   -------MQT ----DRIIAF VMAGGQGSRL QPLTTARSKP SVPFGSRYRI VDFVLSNLVN SQIRTIYLLV QYKSQSLIEH VRKS--WTIS  77
97    1   ---------- ---MAGVLGM ILAGGEGSRL KPLTETRTKP AVPFGGSYRL IDFALNNFVN ADLMRIYVLT QFKSQSLYIH MKKG--WNLS  75
98    1   ---------- ---MQDTLAV ILAGGMGSRL SPLTDDRAKP AVPFGGKYRI IDFTLTNCLH SGLRRILVLT QYKSHSLHKH LRNG--WSIF  75
99    1   ---------- ---MAGVLGM ILAGGEGSRL RPLTESRSKP SVPFGGSYRL IDFALNNFVN ADLMKIYVLT QFKSQSLFHH MKKG--WNIS  75
100   1   --MHNHKYIE VLTMQDTLTI VLAGGVGSRL SPLTDNRAKP AVPFGGKYRI IDFTLANCLH SGLRQILVLT QYKSHSLQKH LRDG--WSVL  86
101   1   ---------- -MIRKEMIAM LLAGGQGSRL GVLTKNVAKP AVLYGGKYRI IDFSLSNCTN SGIDTVGVLT QYQPLKLNAH IGIGKPWDMD  79
102   4   --------RV QPLARDAMAY VLAGGRGSRL KELTDRRAKP AVYFGGKARI IDFALSNALN SGIRRIGVAT QYKAHSLIRH LQRG--WDFF  84
103   1   ------MYVT GNLAGNTIAM VLAGGKGERL APLTLRRPKP GVAFGGKYKI IDFVLSNMFN SGIKKVYILT QYRAYSLMKH IRES--WGKW  82
104   1   ---------- ---MRQVTAI ILGGGRGTRL YPLTKRRAKP AVPIGGKYRL IDIPVSNCIN SGIQHIYILT QFNSASLNRH VSQTYQ--FS  75
105   8   -------LDI NRALKETLAL VLAGGRGSRL RDLTNRESKP AVPFGGKYRI IDFPLSNCMN SGIRRMCVIT QYRAHTLIHH IQRG--WGFL  89
106  74 DSQNSQTCLD PDASSSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG 163
107   1   -----MAVQM EFSMKRVLAI ILGGGAGTRL YPLTKRRAKP AVPLAGKYRL IDIPVSNCIN SEIHNIYVLT QFNSASLNRH IARTYT--FP  83
108   1   ---------- ---MKRVLAI ILGGGKGSRL YPLTKMRAKP AVPLAGKYRL IDIPISNCIN SNITKMYVLT QFNSASLNRH LAQTYN--LS  75
109   1   ---------- ---MRQVTAI ILGGGRGTRL YPLTKRRAKP AVPIGGKYRL IDIPVSNCIN SGIQHIYILT QFNSASLNRH VSQTYQ--FS  75
110   1   ---------- ---MDNVLSI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCIN SDINKVYCLT QFNSASLNRH LSQAYNTNIG  77
111  37 S---------- ----KTVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKVYILT QFNSASLNRH LARTYNFGNG 113
112  66 AAGTGQNDPA GDISKTVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNVTKIYCLT QFNSASLNRH LSQAYNSSVG 155
113  60 R-----EQDR QPRGEVCSSI ILGGGAGTRL FPLTKSRAKP AVPIGGAYRL IDVPMSNCIN SGISKIYILT QFNSTSLNRH LGRAYNMGSG 144
114  58 ASE--DQALE ARNSKTVVAV ILGGGAGTRL FPLTRRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYILT QFNSQSLNRH LSRAYDFSNG 145
115  33 DS----TYLN PQAHDSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYGSNIG 118
116  64 DSRSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNVSKIYVLT QFNSASLNRH LSRAYGNNIA 153
117  66 AAAA-AAARR DVSPDTVASI ILGGGAGTRL FPLTRTRAKP AVPVGGCYRL IDIPMSNCIN SKINKIYVLT QFNSQSLNRH IARTYNFGEG 154
118  72 Q----SSRKN YADANRVSAI ILGGGTGSQL FPLTSTRATP AVPVGGCYRL IDIPMSNCFN SGINKIFVMS QFNSTSLNRH IHRTY-LEGG 156
119  82 AP---VFETP QADPSNVASI ILGGGAGTRL FPLTSRRAKP AVPIGGCYRL IDIPMSNCIN SGIKKIFILT QFNSFSLNRH LARTYNFGNG 168
120  81 RLR--DLEME KRDPRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGSYRL IDVPMSNCIN SGINKVYILT QYNSASLNRH LARAYNLGNG 168
121  83 P----RFERR KADPKNVASI ILGGGAGTQL FPLTRRAATP AVPLGGCYRL IDIPMSNCIN SGINKIFVLT QFNSTSLNRH LARTY-FGNG 167
122  76 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG 165
123  83 P----IFERR RADPKNVASI ILGGGAGTQL FPLTIRQATP AVPVGGCYRL IDIPMSNCIN SNINKIFILT QFNSASLNRH IARTY-FGNG 167
124  73 KLR--DLEME KRDPRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGSYRL IDVPMSNCIN SGINKVYILT QFNSASLNRH LARAYNFGHG 160
125  63 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG 152
126  78 GP---VFEKQ HADPSSVAAI ILGGGAGTRL FPLTSRRAKP AVPIGGCYRL IDVPMSNCIN SGIRKIFILT QFNSASLNRH IARIYNFGNG 164
127  60 ASE--DQALE ARNSRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYILT QFNSQSLNRH LSRAYDCTNG 147
128  64 DSRSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKKRAKP AVPLRANYRL IDIPVSNCLN SNVSKIYVLT QFNSASLNRH LSRAYGNNIA 153
129  73 Q----SSRKS YADANHVSAI ILGGGTGSQL FPLTSTRATP AVPVGGCYRL IDIPMSNCFN SGINKIFVMT QFNSTSLNRH IHRTY-LGGE 157
130  75 -----SFRRN YADPNEVAAV ILGGGTGTQL FPLTSTRATP AVPIGGCYRL IDIPMSNCFN SGINKIFVMT QFNSASLNRH IHRTY-LGGG 159
131  28 S---------- ----KSVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKMYILT QFNSVSLNRH LARTYNFGNG 104
132  57 IATETATEVN DN-TDNVLGI ILGGGAGTRL YPLTKTRAKP AVPLGANYRL IDLPVSNCIN SDINKMYCLT QFNSASLNRH LSQAYNNNVG 145
133  77 TP---VFETP RADPKKVASI ILGGGAGTRL FPLTSKRAKP AVPIGGCYRL IDIPMSNCIN SGIRKIFIMT QFNSFSLNRH LARTYNFGNG 163
134  83 PPPPPRFERR KVDPKNVASI ILGGGAGTQL FPLTRRAATP AVPVGGCYKL IDIPMSNCIN SGINKIFVLT QFNSASLNRH LARTY-FGNG 171
135  80 FQAT-VLRKQ EADPKTVASI ILGGGAGTRL FPLTRTRAKP AVPIGGCYRL IDVPMSNCIN SGINKIYILT QFNSQSLNRH IARTYNSGNG 168
136  86 KLR--DLEME KRDPRTVVAI ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYILT QFNSASLNRH LARAYNFGNG 173
137  75 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG 164
138  76 KVH--ELETE KRDSRTVASI ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYILT QYNSASLNRH LARAYN-SNG 162
```

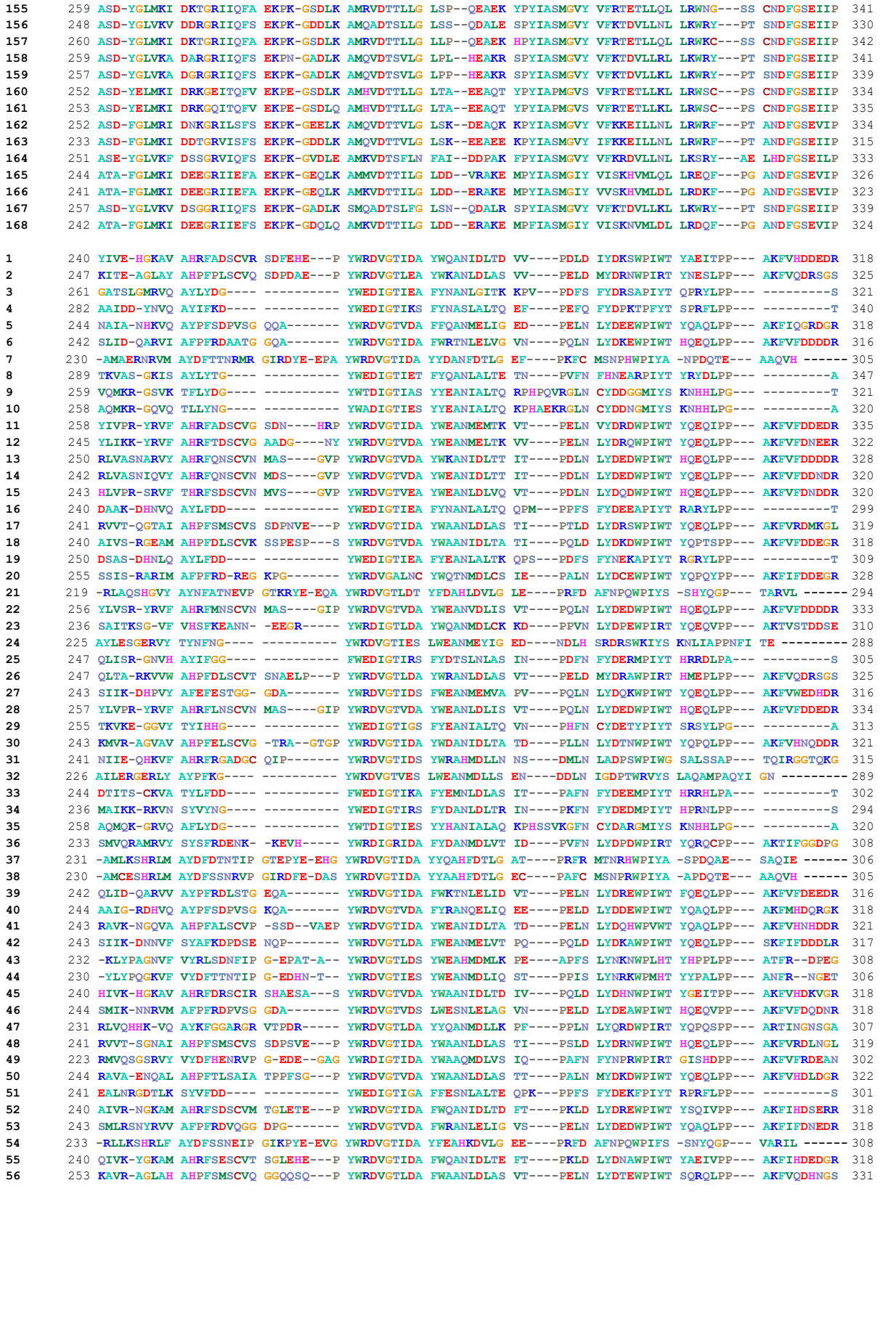
```
139     76 S----MFERR KADPQNVAAI ILGGGNGAKL FPLTMRAATP AVPVGGCYRL IDIPMSNCIN SCINKIFVLT QFNSASLNRH LARTY-FGNG  160
140     19 ------IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNSASLNRH LARTYNFGNG  103
141     16 ------IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNSASLNRH LARTYNFGNG  100
142     88 LPFS-VFETP RVDPKSVVSI ILGGGVGTRL FPLTKQRAKP AVPIGGGYRL IDVPMSNCIN SGINRVFVLT QFNSASLNRH LARTYNF---  173
143     53 LPFS-VFETP RVDPKSVVSI ILGGGVGTRL FPLTKQRAKP AVPIGGGYRL IDVPMSNCIN SGINRVFVLT QFNSASLNRH LARTYNF---  138
144      7  ARDESSPYLE PDARSSVLAV ILGGGAGTRL HPLTRERAKP AVPLGANYRL IDIPVSNCIN SNIPRIYVLT QYNSTSLNSH LYRAYAGNMG  96
145     13 DSMEEEICLK PDAGVSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCIN SNIRKIYVLT QFNSASLNRH LSRAYSSNMG  102
146      1  ---------- ---MQSVLAV ILGGGAGTRL HPLTKERAKP AVPLGANYRL IDIPVSNCIN SNIPRIYVLT QYNSTSLNSH LYRAYAGNMG  77
147      3  ------IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNSASLNRH LARTYNFGNG  87
148     63 APQLRYEPAT KARTNTVLSI ILGGGAGTRL FPLTKQRAKP AVPIGGAYRL IDVPMSNCIN SGISKIYILT QFNSTSLNRH LARAYNMGSG  152
149     64 TAAPASYDYA GDISKTVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNVTKIYCLT QFNSASLNRH LSQAYNSSVG  153
150     58 TTTTTAGKVV DDSTDNVLAI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCIN SDINKMYCLT QFNSASLNRH LSQAYNSNVG  147
151     64 S---------- ----KSVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKMYILT QFNSVSLNRH LARTYNFGNG  140
152      1  ---------- ---MDNVLSI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCIN SDINKVYCLT QFNSASLNRH LAQAYNTNIG  77
153     19 T--------- ----KTVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKVYILT QFNSASLNRH LARTYNFGNG  95
154     75 P----SFLRR RADPKNVISI ILGGGPGTQL FPLTKRAATP AVPVGGCYRL IDIPMSNCLN SGINKIFVLT QFNSASLNRH IARTY-FGNG  159
155     84 VP---TFEKP EVDPKSVASI ILGGGAGTRL FPLTGRRAKP AVPIGGCYRL IDIPMSNCIN SGIRKIFILT QFNSFSLNRH LSRAYSFGNG  170
156     75 P----SFLRR RADPKNVVSI ILGGGPGTQL FPLTKRAATP AVPVGGCYRL IDIPMSNCLN SGINKIFVLT QFNSASLNRH IARTY-FGNG  159
157     85 VP---TFEKP EVDPKSVASI ILGGGAGTRL FPLTGRRAKP AVPIGGCYRL IDIPMSNCIN SGIRKIFILT QFNSFSLNRH LSRAYSFGNG  171
158     86 P----SFLRR KADPKNVVSV ILGGGPGIQL FPLTKRAATP AVPVGGCYRL IDIPMSNCIN SGLNKIFVLT QFNSASLNRH ISRTY-FGNG  170
159     84 P----SFLRR KADPKNVVSI ILGGGPGIQL FPLTKRAATP AVPVGGCYRL IDIPMSNCIN SGLNKIFVLT QFNSASLNRH ISRTY-FGNG  168
160     77 GP---IFQSP KANPENVVAI ILGGGAGTRL FPLTSTRAKQ AVPIAGCYRL IDIPMSNCIN SGIRKVYVLT QFNSFSLNGH LSRTYNFGNG  163
161     78 GP---IFQNP KANPENVAAI ILGGGAGTRL FPLTSTRAKQ AVPIAGCYRL IDIPMSNCIN SGIRKVYVLT QFNSFSLNGH LSRTYNFGNG  164
162     76 KLR--DLDME RRNPRTVLAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYILT QFNSASLNRH IARAYNSGNG  163
163     57 ASE--DADTE TRNARTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKVYVLT QFNSASLNRH LSRAYNFSNG  144
164     78 -----SFRRN YADPNEVAAV ILGGGTGTQL FPLTSTRATP AVPIGGCYRL IDIPMSNCFN SGINKIFVMT QFNSASLNRH IHRTY-LGGG  162
165     71 DSKSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYGSNIG  160
166     68 DSKNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNMG  157
167     84 P----SFIRR KADPKNVASI VLGGGPGVQL FPLTKRAATP AVPVGGCYRL IDIPMSNCIN SGINKIFVLT QFNSASLNRH IARTY-FGNG  168
168     69 DSQNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANYRL IDIPVSNCLN SNISKIYVLT QFNSASLNRH LSRAYASNLG  158

1       85 RPERN-ESFD ILPA--SQRV SETQWYEGTA DAVYQNIDII EPY----APE YMVILAGDHI YKMDYEYMLQ QHVDSGADVT IGCLEVPRME  167
2       92 NEEMN-EFVD LLPA--QQRM KGENWYRGTA DAVTQNLDII RRY----KAE YVVILAGDHI YKQDYSRMLI DHVEKGARCT VACMPVPIEE  174
3       95 GYKNE-GFVE VLAAQQSP-- ENPDWFQGTA DAVRQYLWLF EEHT----VL EYLILAGDHL YRMDYEKFIQ AHRETDADIT VAALPMDEKR  177
4      111 VSFGD-GFVE VLAATQTPGE AGKKWFQGTA DAVRKFIWVF EDA-KNKNIE NIVVLSGDHL YRMDYMELVQ NHIDRNADIT LSCAPAEDSR  198
5       89 RGEFG-EFVE LVPA--QQRM DKPLWYAGTA DAVYQNIDII KAH----NPS YVLVLAGDHV YKMDYGGMIA RHAESGAAMT VGCVEVPRKR  171
6       89 RGDFG-EFVE ILPA--QQRT EG-SWYRGTA DAVYQSLDIV RMH----DPD YVLILAGDHV YKMDYGPMLA RHVETGADVT VGCLEVPVEE  170
7       79 PLLPH-HFVT VVPP--QMQH GP-EWFQGTA DSVYQNLHLV ELV----KPD LVVVFGADHV YRMDLRQMIE FHIATQADAT VAALPVPLEH  160
8      125 PGAG---SIE ILTAEQKP-- SKKNWFQGTA DAVRQNIDYL LES----PFE YFLILSGDQL YNIDFQEMVH FAKKNDSDVV VATIPVNTQD  205
9       95 GVLQD--QIH LLAPEGRD-- GSQVWYKGTA DAIRQNLLYL EDT----GIE YFLVLSGDQL YNMDFRKIVD YALSMQSDMV IVAQPIQEKD  176
10      94 GVLQD--QIH LLAPEARQ-- GDQIWYQGTA DAIRKNLLYF EDT----EIE YFLILSGDQL YNMDFRSIVD TAIRTHVDMV LVAQPIPEKD  175
11     103 DGRFD-EFIQ LLPA--QQQI DETQWYQGTA DAVYQNLHFL RRY----QPD HILVVAGDHI YKMDYGRMLA HHVKHHADMT VACIDVPLDE  185
12      91 RGEFN-EFVE LLPA--QQRI Q-EEWYKGTA DAVFQNLDIL RQT----NIE FVLILAGDHV YKMDYGQMLA AHVRNKADMT VACINVPLKE  172
13      95 DGRFQ-EFIE LLPA--QQRT EEGTWYQGTA DAVFQNLDIL RTH----NPG YVLILGGDHI YKMDYGRILA EHVERQADLT IACLEVPVED  177
14      87 DGRFH-EFIE LLPA--QQRT EEGTWYQGTA DAVFQNLDII RTH----NPS YVLVLGGDHI YKMDYGQILA EHVEKQADLT IACLEVPIED  169
15      89 DGRFQ-EFIE LLPA--QQRT EES-WYQGTA DAVYQNIDIL RSH----NPD YVLILGGDHV YKMDYAKLLA DHIAKSAEMT IACIDLPLEE  170
16      76 G-FSE-GFVE VLAAQQTP-- ENPNWFQGTA DAVRQYLWML QEWD----VD EFLILSGDHL YRMDYRLFIQ RHRETNADIT LSVIPIDDRR  157
17      86 RGEFG-EFID LWPA--QQRV EGAHWYRGTA DAVFQNLDII RSI----RPK YVVVLAGDHI YKMDYTRMIA DHAESGADCT VGCIEVPRMD  168
18      85 RNEVN-EFID LLPA--QQRV DEASWYRGTA DAVYQNIDIL REH----DPK YILVLAGDHV YKMNYASLIE DHVALGAPCT VACIEVPLAE  167
19      86 G-FQE-GFVE VLAAQQTK-- DNPDWFQGTA DAVRQYLWLF REWD----VD EYLILSGDHL YRMDYAQFVK RHRETNADIT LSVVPVDDRK  167
20     100 RTEVG-EFVE ILPA--QQRT HKKEWYQGTA DALFQNLDIM QRH----HPE YVLVLGGDHV YTMDYTQMLL YHVQTGADVT VGSVEVPVAE  182
21      68 SLFAD-QFLT VVPP--QMTK TS-TVFGGTA DAVYQSLDLM NMH----RPD LVAIFGADHI YRMDVRQMVR FHCEHDAEAT VAALPVSLNQ  149
22     101 DGRFK-EFVE LLPA--QQRT VEETWYQGTA DAIFQNIDIL LRH----DAK YVLILGGDHI YKMDYSKLLA EHIEKSADMT VACLEIPLQE  183
23      82 NHELG-EFIE CIPP--QQRN VD-RWYRGTA DSIYQNIHIL QRE----RPE RVLILSGDHV YKMNYNDMLA FHIEKNAQLT VAGVEVDRSE  163
24      79 GINSG---AT ILQP--YSAT EGNRWFQGTS HAIYQNIDYI DSI----NPE YVLILSGDHI YKMDYDDMLQ THKDNMASLT VAVIDVPLKE  159
25      84 --FSR-GFVE ILAAEQTF-- DHSGWYEGTA DAVRKNFQHF RTQ----NPS HYLILSGDQL YRMDLAEMYR RHLESGAQVT IAGTLVTREQ  165
26      92 NEEMN-EFVD LLPA--QQRL STEQWYKGTA DAVCQNLDII RRY----DAE YIVILAGDHI YKMDYSRMLL DHVEKGAECT VACIPVPISE  174
27      88 KKELG-ESVE ILPA--SQRF SD-SWYEGTA DAVFQNIDII RDE----LPK YVMILSGDHI YRMDYGTMLA RHVESGAKMT VSCMSVPIEE  169
28     102 DGRFK-EFVE LLPA--QQRT AEETWYQGTA DAVFQNVDIL QRH----DAK YVLILGGDHI YKMDYSKLLD EHIEKAADMT VACLEVPVEE  184
29     100 PFSGG--FIE LLPAEQKP-- HKKTWYQGTA DAVRQSLECF IET----PVD YFLILSGDQL YNMDFRPMLQ FAHENDADLV VASHPVNAKD  181
30      88 KSEMN-EFVD LLPA--QQRV DEESWYRGTA DAVYQNQDIL AAYK----AD YVVVLAGDHV YKQNYALMLA DHVAQGRECT VGCIEVPRQD  170
31      86 NNSFG-GRLD ILPA--SQQQ SE-SWYQGTA DACFQNIDYI KSR----APK YVMILSGDHV YQMDYRKLLA EHVKNGAEMT VSCIEVPTKN  167
32      80 RLNGG---VF ILPP--HQKA SGANWYQGTA DAVYQNIPFI DAY----EPN LVLVLSGDHI YKMNYNKMIA FHKDKGARCT IAVTDVPLSE  160
33      79 --FSN-GFVE ILAAEQTY-- HSETWYQGTA DAVRKNLKHF RDQ----AAD YYIILSGDQL YRMDFQLMLK KHIESGAELT IAAKPISREQ  160
34      76 --FSH-GFVE ILAAEQTL-- EHSGWYEGTA DAVRKNFIHF KTQ----NPT HYIILSGDQL YRMDLKKFLD KHIESGADIT IATTSVTRED  157
35      94 GVLQD--HIH LLVPEGRQ-- GNQIWYRGTA DAIRQNLLYL KDL----DLD YFLILSGDQL YNMDFHVIVE SMISSQADMI LVAQPVSEKD  175
36      79 NPEMG-EYIY SIPP--QQVT VN-RWYRGTA DSIYQNISIL QSE----RPD YVLILSGDHV YKMNYMEMLN YHIDKRADMT AASVEFPRLE  160
37      79 RFIPQ-HFVT VVPP--QMRN GP-EWFQGTA DSVYQNIHLI ESF----KPD IVAVFGADHI YRMDVRQMID FHVKNDAHVS VATLPVKLAD  160
38      79 PLIPH-HFIT VVPP--QMQC GP-EWFQGTA DSVYQNLHLI DLI----KPD LVVVFGADHV YRMDIRQMVQ YHVDTHADAT VAAIPIPLEQ  160
39      89 SSDFG-EFIE LLPA--QQRI AD-SWYLGTA DAVYQSLDIV RLH----DPD YVLILAGDHI YKMDYGPLLA YHVERGADVT VSCLEVAIEE  170
40      89 RGEFG-EFVE LIPA--QQRM DKPLWYSGTA DSVYQNIDII QAH----DPS YVLILAGDHV YKMDYGAMIA RHVESGADVT VGCVQVTLEQ  171
```

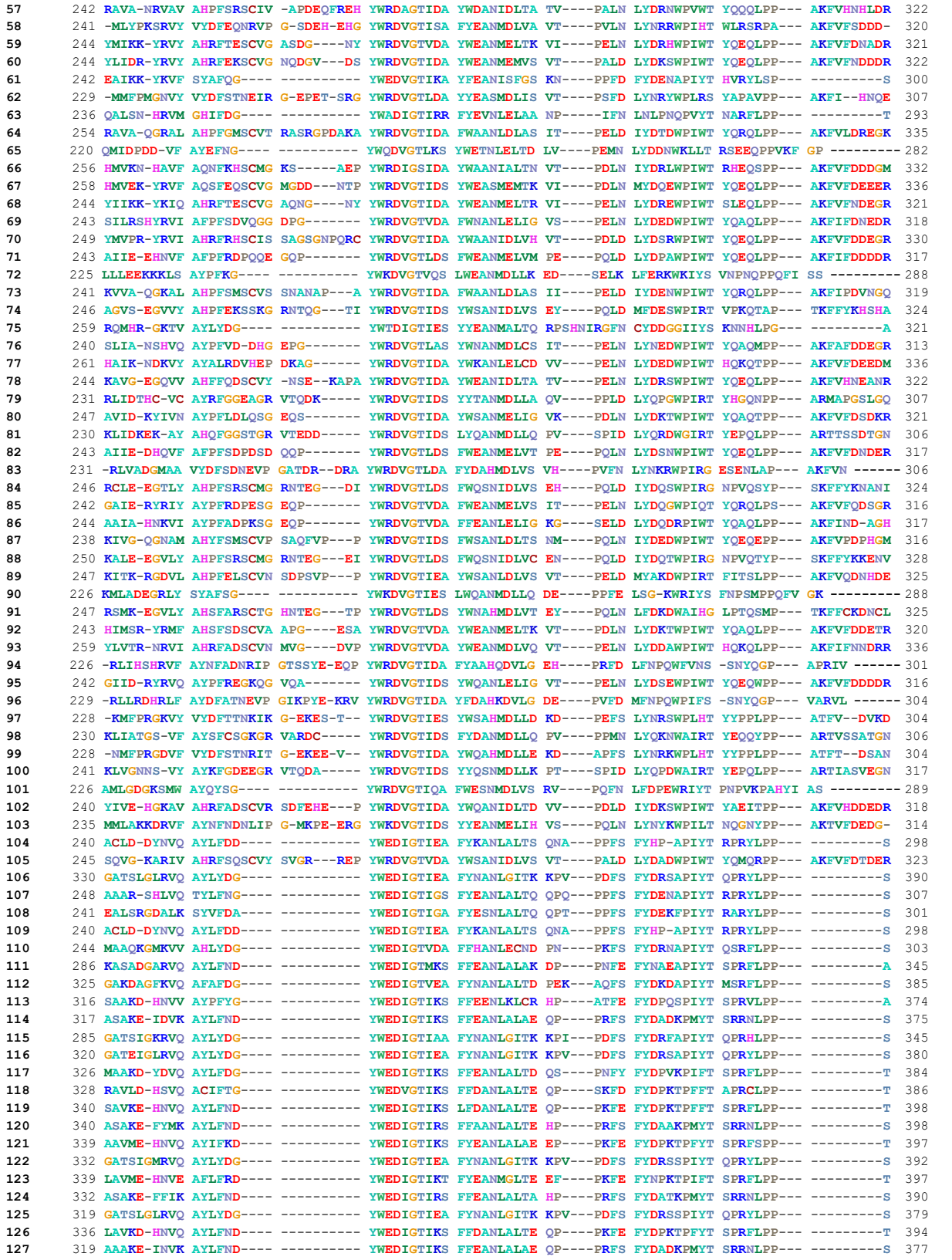


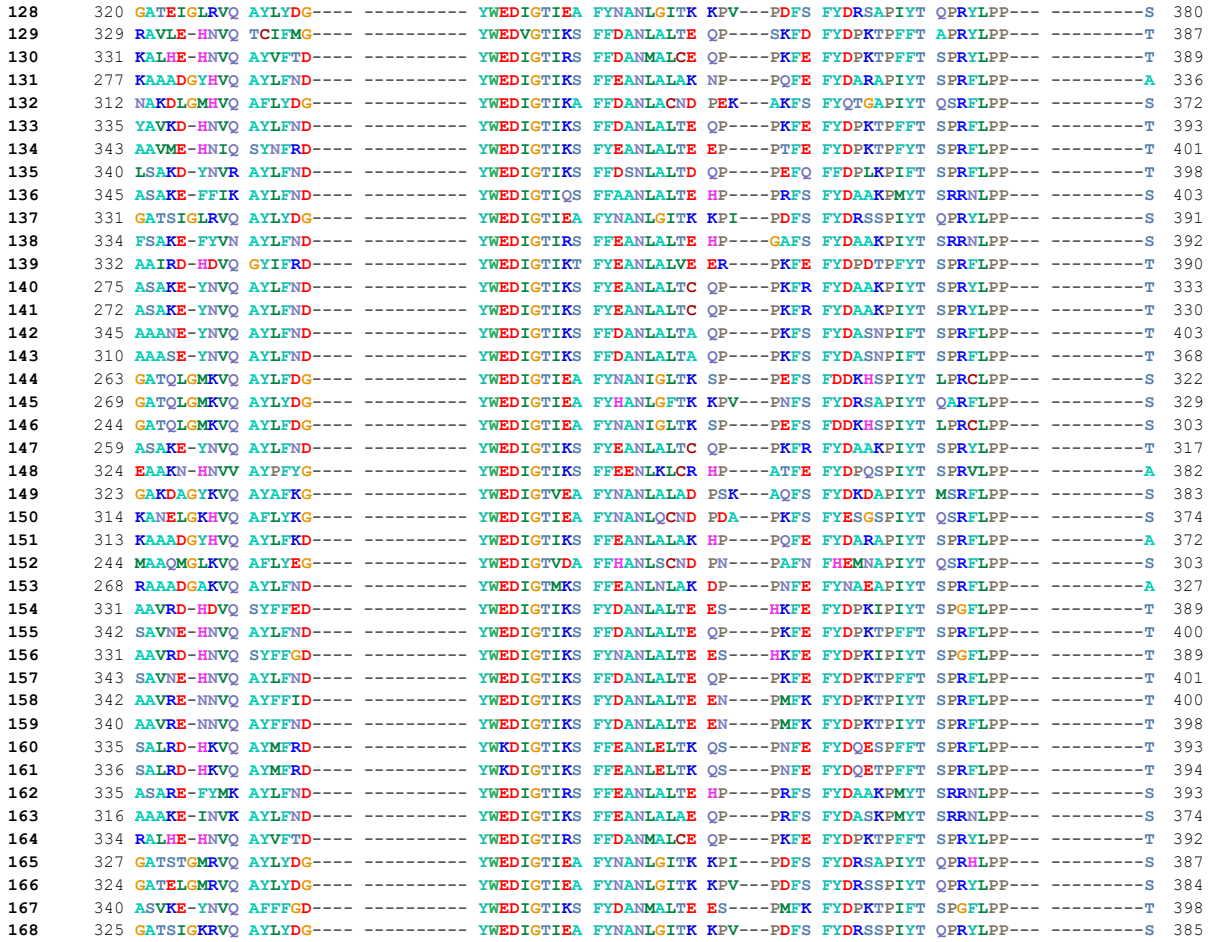

111 114 IMYGGNGFVE VLAATQTPGQ GGKEWFQGTA DAVRQYSWLF NDV-KNKDVE DIVILAGDHL YRMDYMKFVE AHRESNADIT VGTLPIDEER 202
112 156 GYNSR-GFVE VLAASQSS-- ANKSWFQGTA DAVRQYMWLF EEAVREG-VE DFLILSGDHL YRMDYRDFVR KHRNSGAAIT IAALPCAEKE 241
113 145 VRFGGDGFVE VLAATQTP-- TDKEWFQGTA DAVRQYSWLL EDT-KNRAIE DVLILSGDHL YRMDYMKFVN YHRETNADIT IGCIAYGSDR 231
114 146 VAIGD-GFVE VLAATQRPGT EGKRWFQGTA DAVRQFDWLF DDA-KSKDIE DVLILSGDHL YRMDYMDFVQ SHRQRGAGIS ICCLPIDGSR 233
115 119 GYKNE-GFVE VLAAQQSP-- DNPNWFQGTA DAVRQYLWLF EEHN----VM EFLILAGDHL YRMDYEKFIQ AHRETNADIT VAALPMDEKR 201
116 154 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYMWLF EEHN----IM EFLILAGDHL YRMDYQKFIQ AHRETDADIT VAALPMDEQR 236
117 155 VGFSG-GSVE VLAATQTAGE SGKKWFQGTA DAVRQFLWLF EDA-RLKCIE NILILSGDHL YRMDYMDFVQ KHVDSGADIS VACVPMDESR 242
118 157 INFAD-GSVQ VLAATQMPEE PA-GWFQGTA DSIRKFIWVL EDYYSHKSID NIVILSGDQL YRMNYMELVQ KHVEDDADIT ISCAPVDESR 244
119 169 VSFGD-GFVE VLAATQTPGE AGKKWFQGTA DAVRQFIWMF EDA-RTKNVE HVLILSGDHL YRMNYMEFVQ KHIDTNADIT VSCVPMDDSR 256
120 169 VSFGD-GFVE ALAATQTPGE AGKKWFQGTA DAVRQFHWLF EGP-RSKEIE DVLILSGDHL YRMDYMDFVQ NHRQGGADIT LSCLPMDDSR 256
121 168 IIFGD-GFVE VLAATQTPGE AGMKWFQGTA DAVRQFTWVF EDA-KNRNIE NILVLSGDHL YRMDYMDFVQ HHIDSNADFT ISCAAVGESR 255
122 166 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYLWLF EEHN----VL EFLVLAGDHL YRMDYERFIQ AHRETDADIT VAALPMDEKR 248
123 168 VNFGD-GFVE VLAATQTPGE AGMKWFEGTA DAVRKFIWVF EDA-KNKNIE NILILSGDHL YRMDYMDLVQ NHIDRKADIT VSCVPVGESR 255
124 161 VNFGD-GYVE ALAATQTPGE AGKRWFQGTA DAVRQFHWLF EDQ-RSKEIE DVLILSGDHL YRMDYMDFVQ NHRQSGADIT ISCLPMDDSR 248
125 153 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYLWLF EEHN----VL EFLVLAGDHL YRMDYERFIQ AHRETDADIT VAALPMDEKR 235
126 165 VNFGD-GFVE VLAATQTPGE AGQKWFQGTA DAVRQFIWVF EDA-KNKNVE HILILSGDHL YRMDYMDFVQ KHIDSNADIT VSCVPMDDSR 252
127 148 VAFGD-GFVE VLAATQRPGS EGKRWFQGTA DAVRQFDWLF DDA-KSKDID DVLILSGDHL YRMDYMDFVQ SHRQRGAGIS ICCLPIDDSR 235
128 154 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYMWLF EEHN----IM EFLILAGDHL YRMDYQKFIQ AHRETDADIT VAALPMDEQR 236
129 158 INFAD-GSVQ VLADTQMPEE PD-GWFQGTA DSVRKFIWVL EDYYNHKSIE HIVILSGDQL YQMNYMELVQ KHVEDNADIT VSCAPVDESR 245
130 160 INFTD-GSVE VLAATQMPGE AA-GWFQGTA DAVRKFIWVL EDYYKHKAIE HILILSGDQL YRMDYMELVQ KHVDDNADIT LSCAPVGESR 247
131 105 IMYGGNGFVE VLAATQTPGL GGKEWFQGTA DAVRQYSWLF EDI-KNKDVQ DIVILSGDHL YRMDYMAFVA RHREVNADIT IGCLPMDDKR 193
132 146 SYNRQ-GFVE VLAAQQSP-- KNKDWFQGTA DAVRQYIWLF NESK----CD EYIILSGDHL YRMDYKPFIL KHRQTKADIT VSAVPMDEER 228
133 164 VNFGD-GFVE VLAATKTPGE AGNKWFQGTA DAVRQFIWVF EDA-KNKNVE NVLILSGDHL YRMDYMEFVQ KHIDSGADIT VSCVPMDDSR 251
134 172 INFGD-GFVE VLAATQTPGE AGMNWFQGTA DAVRQFTWVF EDA-KNRNVE NILILSGDHL YRMDYMDFVQ HHVDSNADIT ISCAAVGESR 259
135 169 VNFGD-GFVE VLAATQTPGE SGKKWFQGTA DAVRQFLWLF EDA-KHSHIE NILILSGDHL YRMDYMDFLQ KHIDSGADIT VSCLPVDESR 256
136 174 INFGD-GFVE VLAATQTPGE AGKRWFQGTA DAVRQFHWLF EDA-RSKDID DVLVLSGDHL YRMDYMDFVQ NHRQSGADIT ISCLPMDDSR 261
137 165 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYLWLF EEHN----VL EFLILAGDHL YRMDYERFIQ AHRETDADIT VAALPMDEKR 247
138 163 VGFGD-GYVE VLAATQTPGE SGKRWFQGTA DAVRQFHWLF EDA-RSKDIE DVLILSGDHL YRMDYMDFVQ DHRQSGADIS ISCIPIDDRR 250
139 161 INFGG-GFVE VLAATQTPGE AGKMWFQGTA DAVRKFLWVF EDA-KNRNIE NILILSGDHL YRMNYMDFVQ SHVDSNADIT LSCAPVSESR 248
140 104 VNFGD-GFVE VLAATQTPGE AGMNWFQGTA DAVRQFTWVF EDT-RSKEIE NVLVLSGDHL YRMDYMEFIQ KHQDTGADIT IGCVPMDDSR 191
141 101 VNFGD-GFVE VLAATQTPGE AGMNWFQGTA DAVRQFTWVF EDT-RSKEIE NVLVLSGDHL YRMDYMEFIQ KHQDTGADIT IGCVPMDDSR 188
142 174 INAGE-GFVE VLAATQTPGE SGMNWFQGTA DAVRQFTWLF EDV-RNKDVD YVLVLSGDHL YRMDYMDFVQ KHKDSGADIT ISCVPVDESR 261
143 139 INAGD-GFVE VLAATQTPGE SGMNWFQGTA DAVRQFTWLF EDV-RNKDVD YVLVLSGDHL YRMDYMDFVQ KHKDSGADIT ISCVPVDESR 226
144 97 GFRND-GFVE VLAAEQSL-- DNPDWFRGTA DAVRQYLWIF EDQD----VM EFLILAGDHL YRMDYQRFIR SHRQTKADIT VAAVPVEEKR 179
145 103 NYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYMWLF EEQP----VM EYLILAGDHL YRMDYQKFIQ AHRITDADIT VAALPMDEKR 185
146 78 GFRND-GFVE VLAAEQSL-- DNPDWFRGTA DAVRQYLWIF EDQD----VM EFLILAGDHL YRMDYQRFIR SHRQTKADIT VAAVPVEEKR 160
147 88 VNFGD-GFVE VLAATQTPGE AGMNWFQGTA DAVRQFTWVF EDT-RSKEIE NVLVLSGDHL YRMDYMEFIQ KHQDTGADIT IGCVPMDDSR 175
148 153 VRFGGDGFVE VLAATQTP-- TDKEWFQGTA DAVRQYSWLL EDT-KNRAIE DVLILSGDHL YRMDYMKFVN YHRETNADIT IGCIAYGSDR 239
149 154 GYNTR-GFVE VLAASQSS-- ANKSWFQGTA DAVRQYMWLF EEAVREG-VE DFLILSGDHL YRMDYRDFVR KHRESGAAIT IAALPCAEKE 239
150 148 SGLRQ-GFVE VLAAQQSP-- KSKVWFQGTA DAVRQYMWLF NESK----CE EYIILSGDHL YRMDYKPFIL EHRKTGADIT VSAVPMDAAR 230
151 141 IMYGGSGFVE VLAATQTPGL GGKEWFQGTA DAVRQYSWLF EDV-KNKDVQ DVVILSGDHL YRMDYMAFVD RHREVNADIT IGCLPMDGER 229
152 78 THTRQ-GFVE VLAAQQSP-- VNKAWFQGTA DAVRQYLWLF EESK----CE EYLILSGDHL YRMDYRPFIM KHRETEAAIT VAALPCDEKR 160
153 96 IMYGGNGFVE VLAATQTPGQ GGKEWFQGTA DAVRQYSWLF NDV-KNKDVE DIVILAGDHL YRMDYMKFVE AHRESNADIS VGTLPIDEAR 184
154 160 INFGD-GIVE VLAATQTPGE AGKNWFQGTA DAVRQFTWVF EDA-KNTNVE NVLILAGDHL YRMDYMDLVQ SHVDRNADIT VSCAAVGDSR 247
155 171 ITFGD-GFVE VLAATQTPGE AGKKWFQGTA DAVRQFIWVF EDA-KNKNVE HILILSGDHL YRMDYMNFVQ RHVDTNADIT VSCVPMDDSR 258
156 160 INFGD-GIVE VLAATQTPGE AGKNWFQGTA DAVRQFTWVF EDA-KNTNVE NVLILAGDHL YRMDYMDLVQ SHVDRNADIT VSCAAVGDSR 247
157 172 MTFGD-GFVE VLAATQTPGE AGKKWFQGTA DAVRQFIWVF EDA-KNKNVE HILILSGDHL YRMDYMDFVQ RHVDTNADIT VSCVPMDDSR 259
158 171 INFGD-GCVE VLAATQTQGE AGNNWFQGTA DAVRQFTWVF EDA-KHANIE NVLILAGDHL YRMNYMDLVQ SHVDRNADIT VSCAAVGESR 258
159 169 INFGD-GCVE VLAATQTQGE TGKNWFQGTA DAVRQFTWVF EDA-KHTNIE NVLILAGDHL YRMDYMDLVQ SHVDRNADIT VSCAAVGESR 256
160 164 VNFGG-GFVE VLAATLTNGE AGNKWFQGTA DAVRRFSWVF EDA-KNKNIE HILIISGDHL CRMDYMKLVE KHIGTNADIT VSCVPMDESR 251
161 165 VNFGG-GFVE VLAATKTPGE SGNKWFQGTA DAVRRFIWVF EDA-KNKDIE NILIISGDHL CRMDYMKLLE KHIGTNADIT VSCVPMDESR 252
162 164 VTFGD-GYVE VLAATQTPGE AGKKWFQGTA DAVRQFHWLF EDP-RSKDIE DVLILSGDHL YRMDYMDFVQ NHRESGADIT LSCLPMDDSR 251
163 145 VGFGD-GFVE VLAATQRPGL EGKRWFQGTA DAVRQFDWLF DDA-KAKDIE DVIILSGDHL YRMDYMDFVQ SHRQRDAGIS ICCLPIDDSR 232
164 163 INFTD-GSVE VLAATQMPGE AA-GWFRGTA DAVRKFIWVL EDYYKHKSIE HILILSGDQL YRMDYMELVQ KHVDDNADIT LSCAPVGESR 250
165 161 GYKNE-GFVE VLAAQQSP-- DNPNWFQGTA DAVRQYLWLF EEHN----VM EYLILAGDHL YRMDYEKFIQ AHRETDADIT VAALPMDEER 243
166 158 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYLWLF EEHN----VL EYLVLAGDHL YRMDYERFIQ AHRESDADIT VAALPMDEAR 240
167 169 INFGD-GYVE VLAATQTPGE AGKNWFQGTA DAVRQFTWVF EDA-KNTNIE NVIILAGDHL YRMDYMDLVQ SHIDRNADIT VSCAAVGDSR 256
168 159 GYKNE-GFVE VLAAQQSP-- ENPNWFQGTA DAVRQYLWLF EEHN----VL EYLILAGDHL YRMDYEKFIQ AHRETDADIT VAALPMDEKR 241
1 168 ATG-FGVMHV NEKDEIIDFI EKPAD----- ---------- --PPGIPGNE GFALASMGIY VFHTKFLMEA LRRDAADPTS SRDFGKDIIP 239
2 175 ASA-FGVMAV DENDKIIEFV EKPAN----- ---------- --PPSMPNDP SKSLASMGIY VFDADYLYEL LEEDDRDENS SHDFGKDLIP 246
3 178 ATA-FGLMKI DEEGRIIEFA EKPQ-GEQLQ AMKVDTTILG LDD--KRAKE MPFIASMGIY VISKDVMLNL LRDKF---PG ANDFGSEVIP 260
4 199 ASD-FGLVKI DSRGRVVQFA EKPK-GFDLK AMQVDTTLVG LSP--QDAKK SPYIASMGVY VFKTDVLLKL LKWSY---PT SNDFGSEIIP 281
5 172 ASA-FGVMSV NEERQVLAFN EKPKD----- ---------- --PTPMPGNP DRALVSMGIY VFDRDYLFQL LREDAENFDS SRDFGKDVIP 243
6 171 ASA-FGVMAV DGDNRVVRFQ EKPAD----- ---------- --PPSIPGQS DRALASMGIY IFNRAFLFNQ LIADA-RKES DHDFGKDIIP 241
7 161 TSS-FGIIRT DSAGRIRDFR EKPES----- ---------- --AEPMPGRP GHALASMGNY VFSADLLVDA LRRAHS--RG GHDFGKNLLP 230
8 206 AKR-MGILKV DEQNSITSFY EKPQDNDLLQ QLRSPSNILE KAGV-APTGE RVYLGSMGIY LFKRKALVEL LSE-----DI REDFGKHLIP 288
9 177 ASR-MGVLQI DEEANLLDFY EKPQEEEILN RFRLSSQECR KHKL-DPQYG N-FLGNMGIY LFRRESLFKL LQE-----EQ GDDFGKHLIQ 258
10 176 AYR-MGVLDI DSEGKLIDFY EKPQEKEVLK RFQLSSEDRR IHKL-TEDSG D-FLGSMGIY LFRRDSLFSL LRE-----EE GNDFGKHLIQ 257
11 186 ARE-FGVMGV DEQDRVIDFV EKPQN----- ---------- --PPAIPGQP DRALASMGIY IFNTKFLFEQ LERDAMTKGS NRDFGKDIIP 257
12 173 ASA-FGVMGV DENDRVVDFE EKPAH----- ---------- --PSSLPDDP DHALASMGIY VFNAAFLYEQ LIRDADDPKS SHDFGHDIIP 244

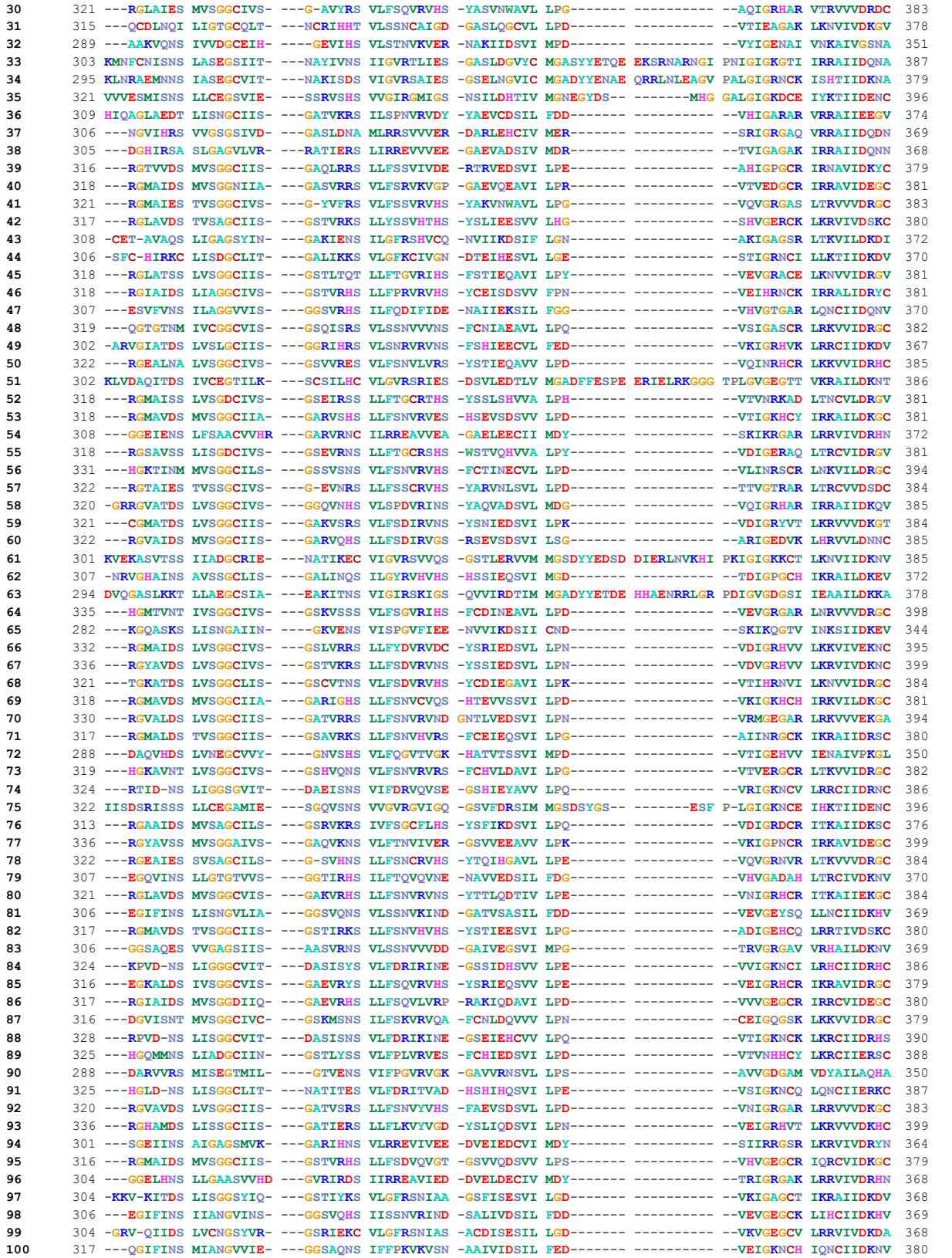


```
84    174 ASE-FGVMAV NENLKVKDFV EKPKD----- ---------- --PPPMVGKP DVSLASMGIY VFDADYLYEM LNKEVNTPYT SHDFGKDVLP  245
85    170 ARG-FGVMSV DKDQRVIGFD EKPAN----- ---------- --PSPQPGIP DKALASMGNY VFNTEFLYEQ LEKDAGESSS AHDFGHNIIP  241
86    172 ARA-FGVMTV DENGRVLRFT EKPQE----- ---------- --PNPVPGKP DTALVSMGIY VFEREYLFEQ LRADAENIDS SRDFGRDVIP  243
87    166 ASG-FGVMDV DENRKITRFL EKPAD----- ---------- --PPGMPDNP DKSLASMGIY VFNADYLYRL LDEDCGDNTS SHDFGKDIIP  237
88    178 ASE-FGVMAV NENLKVKSFV EKPKD----- ---------- --PPAMVGKP NTSLASMGIY VFDADYLYDV LEREVTSPYT SHDFGKDVLP  249
89    175 AFQ-FGIMDI DGNRRVLNFL EKPSN----- ---------- --PPCIPNDP EHSLASMGIY VVDRDYLFDL LEEDSRDPNS HHDFGQDIIP  246
90    161 APR-FGIMNT DESGRIVEFE EKPAK----- ---------- ---------P KSNLASMGIY IFNRDYLAEY LTADARSETS SHDFGKDIIP  225
91    175 AKE-FGIMSV NDKWQVQAFV EKPQD----- ---------- --PPTMREKP DTSMASMGIY VFDSNYLYDV LEDEIQKQTK NLDFGKHIMP  246
92    171 ARE-FGVMSV DEGHRVVAFN EKPEH----- ---------- --PQSTPGNP DMALASMGIY VFNAEFLYEQ LARDADDPNS SHDFAKDIIP  242
93    187 GRE-FGIIGV DEEGRVTYFH EKPEH----- ---------- --PAAMPGRP DSALASMGIY VFGARFLYEQ LIRDHDDPES GHDFGKDLIP  258
94    157 ASA-FGIIDA DAQGQIKGFL EKPKN----- ---------- --PPPIPSDP SRAYGSMGNY LFNTDVLLKA LYEAKE--RG EHDFGGNILP  226
95    170 ARA-FGVMAV DGDGRVTDFL EKPDD----- ---------- --PPEMPGKP GTSLASMGIY VFNTAFLFER LIRDADDSRS SHDFGKDIIP  241
96    160 ARG-FGIISA APDGRVQAFQ EKPAN----- ---------- --PTPIPGDP ERAFASMGNY VFRADVLMAA LEEAHR--NG ETDFGGHILP  229
97    157 ASQ-FGVIEV DENGKMVGFE EKPS--N--- ---------- --PKSIPGEP EWALVSMGNY IFEAETLSKE LREDAENNQS SHDFGKDIIP  228
98    158 ASA-FGVMAI DDDSRITCFV EKPAD----- ---------- --PPCIPNRP DHSLASMGIY IFNMDVLKKA LTEDAEIEQS SHDFGKDVIP  229
99    157 ASE-FGVIEV DAEGRMIGFE EKPA--N--- ---------- --PKPIPGDP DSALVSMGNY VFEANELFAE LVEDADREGS SHDFGKDIIP  228
100   169 AKE-FGVMEI DESLQINNFT EKPRY----- ---------- --PACVPGRP TRSMASMGIY IFDKEVLTQA LLADAEDPDS SHDFGKDIIP  240
101   161 ASR-YGIMNC HENGKIYEFE EKPKN----- ---------- ---------P KSTLASMGVY IFTWSTLREY LIKDNECSDS VNDFGKNIIP  225
102   168 ATG-FGVMHV NEKDEIIDFI EKPAD----- ---------- --PPGIPGNE GFALASMGIY VFHTKFLMEA LRRDAADPTS SRDFGKDIIP  239
103   165 ARR-FGVFCV DDDNRVTAFE EKPAN----- ---------- --PVTIPGR- ETCFASMGNY IFSTRRLIEV LQEGKK-LHA DLDFGKHVIP  234
104   158 ASD-FGLIKT DADGRVVQFT EKPK-GAELE RMRVDTTTLG LTL--EEAER RPFVASMGIY VFRHDVMLKL LRD-D---PS RTDFGKEILP  239
105   173 ATG-YGCVEV DNDDNIVHFL EKPAN----- ---------- --PPGIPGRP DRAFASMGIY IFNADFLYEI LESDALNEAS QHDFGRDIIP  244
106   247 ATA-FGLMKI DEEGRIIEFA EKPK-GEHLK AMKVDTTILG LDD--QRAKE MPFIASMGIY VVSRDVMLDL LRNQF---PG ANDFGSEVIP  329
107   166 ASG-FGLLKV DGTGRVTDFR EKPT-GDALR DMRVDTTRYG LTI--EEAHR KPYIASMGIY VFKRQVLIDL LQQ-M---AD ATDFGKEIIP  247
108   159 AEG-FGLMRT DNDGNIREFK EKPS-GEALK AMAVDTSRFG LSP--DSAKE RPYLASMGIY VFSRSTLFDL LNK-Y---PS YKDFGKEVIP  240
109   158 ASD-FGLIKT DADGRVVQFT EKPK-GAELE RMRVDTTTLG LTL--EEAER RPFVASMGIY VFRHDVMLKL LRD-D---PS RTDFGKEILP  239
110   161 ASS-FGLMKI NEHATIIEFS EKPK-GDALK AMQCDTTILG LDA--ERAKE MPYIASMGIY VFNAKAMEQV LQDDF---PE ANDFGGEIIP  243
111   203 ASD-FGLMKI DSSGRIVEFT EKPK-GDALQ AMKVDTTILG LTA--AEAEA KPFIASMGIY VFKKSMLVKF LDDDY---PE DNDFGGEIIP  285
112   242 ASA-FGLMKI DEEGRVIEFA EKPK-GEALT KMRVDTGILG VDP--ATAAA KPYIASMGIY VMSAKALREL LLNRM---PG ANDFGNEVIP  324
113   232 AKE-FGLMKI DEKRRVTSFA EKPKTQEALD AMKVDTTVLG LTP--EEAAE KPYIASMGIY VFKKSVLLQL LNDSY---AK ANDFGGEIIP  315
114   234 ASD-FGLMKI DDTGRVISFS EKPK-GDELK AMQVDTTVLG LSK--EEAEN KPYIASMGIY IFKKDILLNL LRWRF---PT ANDFGSEIIP  316
115   202 ATA-FGLMKI DEEGRIIEFA EKPK-GEQLK AMMVDTTILG LDD--VRAKE MPYIASMGIY VFSKDVMLQL LREQF---PE ANDFGSEVIP  284
116   237 ATA-FGLMKI DDEGRIVEFA EKPK-GEKLR SMMVDTTILG LDP--ERAKE LPYIASMGIY VFSKDVMLRL LRENF---PA ANDFGSEVIP  319
117   243 ASD-FGLMKA DRNGHITDFL EKPK-GADLE SMQVDMGLFG LSP--EFAST YKYMASMGIY VFKADVLRKL LRGHY---PT ANDFGLEVIP  325
118   245 ASK-NGLVKI DHTGRVLQFF EKPK-GADLN SMRVETNFLS YAI--DDAQK YPYLASMGIY VFKKDALLDL LKSKY---TQ LHDFGSEILP  327
119   257 ASD-YGLMKI DSTGRIIQFA EKPK-GTDLK AMQVDTTLLG LSK--QEAMQ FPYIASMGVY VFRTDVLLKL LRCSY---PS CNDFGSEIIP  339
120   257 ASD-FGLMKI DNKGRVLSFS EKPK-GVDLK AMEVDTTVLG LSK--EEALK KPYIASMGVY VFKKEILLNL LRWRF---PT ANDFGSEIIP  339
121   256 ASD-YGLVKI DGRGQVFQFA EKPK-GSELR EMRVDTTRLG LSP--QDAMK SPYIASMGVY VFKTDILLKL LRWRY---PT ANDFGSEIIP  338
122   249 ATA-FGLMKI DEEGRIIEFA EKPK-GEQLK AMKVDTTILG LDD--ERAKE MPYIASMGIY VVSKNVMLDL LREKF---PG ANDFGSEVIP  331
123   256 ASD-YGLLKM DNRGRIIQFA EKPK-GADLK AMKVDTTRLG LSP--QEAMK SPYIASMGVY VFKTDILLNL LRWRY---PT SNDFGSEIIP  338
124   249 ASD-FGLMKI DNKGRVLFFS EKPK-GEDLK AMEVDTKVLG LSR--EEAEK KPYIASMGVY VFKKEILLNL LRWRF---PT ANDFGSEIIP  331
125   236 ATA-FGLMKI DEEGRIIEFA EKPK-GEQLK AMKVDTTILG LDD--ERAKE MPYIASMGIY VVSKDVMLDL LRDQF---PG ANDFGSEVIP  318
126   253 ASD-YGLMKI DNTGRIIQFS EKPK-GPNLK AMKVNTTLLG LSE--KEAEK CPYIASMGVY VFRTDVLLKL LTRKY---LS CNDFGSEIIP  335
127   236 ASD-FGLMKI DDTARVISFS EKPK-GDELK AMQVDTTVLG LSK--EEAEK KPYIASMGVY IFKKDILLNL LRWRF---PT ANDFGSEIIP  318
128   237 ATA-FGLMKI DDEGRIVEFA EKPK-GEKLR SMMVDTTILG LDP--ERAKE LPYIASMGIY VFSKDVMLRL LRENF---PA ANDFGSEVIP  319
129   246 ASN-NGLVKC DHTGRVLQFF EKPK-GADLN SMRVDTNFLS YAI--GDAQK YQYIASMGIY VFKKDALLDL LKSKY---TQ LHDFGSEILP  328
130   248 ASD-YGLVKF DSSGRVIQFS EKPK-GAALE EMKVDTSFLN FAI--DDPTK YPYIASMGVY VFKRDVLLDL LKSRY---AE LHDFGSEILP  330
131   194 ASD-FGLMKI DDTGRITEFA EKPN-GDALK AMEVDTTILG LTA--EEATS SPYIASMGIY VFKKSALLNF LNAEY---PK DNDFGGEIIP  276
132   229 AAA-FGLMKI DDTGKIIDFA EKPT-GDALK AMMVDTTILG LDA--ERAKE MPYIASMGIY VFNARAMEKL LMEDF---PT CHDFGGEIIP  311
133   252 ASD-YGLMKI DNTGRIIQFA EKPK-GLDLK AMQIDTKLLG LSK--QDALQ YPYIASMGVY VFRTEVLCKL LRWSY---PS CIDFGSEVIP  334
134   260 ASD-YGLVKI DSRGRIVHFA EKPG-GAELK SLKADTTQLG LSP--QDALK SPYIASMGVY VFRTEILLKL LRWRF---PT SNDFGSEIIP  342
135   257 ASD-FGLIKI DETGQIRQFL EKPK-GESLK SMRVDTSTLG LSI--SDARK LPYIASMGIY MFKTDVLLKL LRWHY---PT ANDFGSEIIP  339
136   262 ASD-FGLMNI DNKGRVLSFS EKPK-GADLK AMAVDTTVLG LSK--EEAEK KPYIASMGVY VFKKEILLNL LRWRF---PT ANDFGSEIIP  344
137   248 ATA-FGLMKI DEEGRIIEFA EKPK-GEQLK AMKVDTTILG LDD--ERAKE MPYIASMGIY VVSKNVMLDL LRDKF---PG ANDFGSEVIP  330
138   251 ASD-FGLMKI DDKGRVISFS EKPR-GDELK AMAVDTTILG LSK--EEAEK KPYIASMGVY VFKKEILLNL LRWRF---PT ANDFGSEIIP  333
139   249 ASN-FGLVKI DRGGRVIHFS EKPT-GVDLK SMQTDTTMLG LSH--QEATD SPYIASMGVY CFKTEALLNL LTRQY---PS SNDFGSEVIP  331
140   192 ASD-FGLMKI DANGQILYFS EKPK-GADLK AMQVDTTVLG LTP--EEAIE KPYIASMGIY VFKKDILLKL LRWRY---PT ANDFGSEILP  274
141   189 ASD-FGLMKI DANGQILYFS EKPK-GADLK AMQVDTTVLG LTP--EEAIE KPYIASMGIY VFKKDILLKL LRWRY---PT ANDFGSEILP  271
142   262 ASD-FGLVKT DARGRIISFS EKPK-GMDLK AMQVDTTALG LSR--EEAKK MPYIASMGIY VFRKDVLLKL LRWRY---PT SNDFGSEIIP  344
143   227 ASD-FGLVKT DARGRIISFS EKPK-GMDLK AMQVDTTALG LSR--EEAKK MPYIASMGIY VFRKDVLLKL LRWRY---PT SNDFGSEIIP  309
144   180 ATN-FGLMKI DSEGKITEFA EKPK-GGILQ GMKVDTTILG LDP--KRAEA LPYIASMGIY VISKEAMYKL LHEKF---PN ANDFGSEIIP  262
145   186 ATA-FGLMKI DDEGRITEFS EKPK-GSALK AMEVDTTILG LDP--ERAKE MPYIASMGIY VVSKDVMSRL LRDEF---PN CNDFGSEVIP  268
146   161 ATN-FGLMKI DSEGKITEFA EKPK-GGILQ AMKVDTTILG LDP--KRAEA LPYIASMGIY VISKEAMYKL LHEKF---PN ANDFGSEIIP  243
147   176 ASD-FGLMKI DANGQILYFS EKPK-GADLK AMQVDTTVLG LTP--EEAIE KPYIASMGIY VFKKDILLKL LRWRY---PT ANDFGSEILP  258
148   240 AKE-FGLMKI DDKRRVLSFA EKPKTQEALD AMKVDTTVLG LTP--DEAAD KPYIASMGIY VFKKSVLCKL LNETY---AK ANDFGGEIIP  323
149   240 ASA-FGLMKI DDAGRVVEFA EKPK-GEALQ RMKVDTSILG VDP--ATAQS KPFIASMGIY VMSAKALREL LLNRM---PG ANDFGNEVIP  322
150   231 AEA-FGLMKI DDSGRIIDFA EKPK-GKELE AMAVDTTILG LDK--KLAKE MPYIASMGIY VFKASAMDEL LTEKF---PD CHDFGGEIIP  313
151   230 ASD-FGLMKI DKTGRITEFA EKPE-GNDLL AMQVDTTVLG LSP--EESQA SPYIASMGIY VFKKSALISF LNSEY---PK DNDFGGEIIP  312
152   161 ASS-FGLMKI DNTGRVIEFA EKPK-GAELQ AMKVDTTVLG LDA--DKAQE MPFIASMGIY VFDAKKMREC LLENF---KE ADDFGGEIIP  243
153   185 ASD-FGLMKI DSTGRIVEFT EKPK-GDALQ AMKVDTTVLG LTA--DEAKE KPFIASMGIY VFKKSALVKF LEKDY---PE DNDFGGEIIP  267
154   248 ASD-YGLVKV DDRGRIIQFS EKPN-GDDLK AMQADTSLLG LSP--QDALK SPYIASMGVY VFKTDVLLNL LKWRY---PT SNDFGSEIIP  330
```









```
3     407 RIGDNVKIIN KDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP 446
4     426 KIGKNVSIIN KDGVQEADRP EE------GF YIRSG----- ---------- ---------- ---------- ---------I IIILEKATIR 465
5     382 RIPAGTVIGE DPAED----- ----RRR-FF VTPKD----- ---------- ---------- ---------- ---------V VLVTAEMLGQ 417
6     380 RLAAGTVIGE DPEED----- ----ARR-FH LSPGG----- ---------- ---------- ---------- ---------V VLVTPDMLGQ 415
7     369 FVPPGMKIGY DPEAD----- ----RQR-FH ISESG----- ---------- ---------- ---------- ---------V VVVGKGQLKP 404
8     430 RIGKGVQLIN KQQLTRYES- -E------LV FIRDG----- ---------- ---------- ---------- ---------I IVVPRGSVLP 467
9     398 RIGNGVKLTN IQGYKDYDSP DG------KL VVRDG----- ---------- ---------- ---------- ---------I IIIPRGTKIP 437
10    397 CIGNGVKLQN LKGYIKYDSP DK------KL FVRDN----- ---------- ---------- ---------- ---------I IIVPQGTHIP 436
11    399 RIPEGMVIGV DPEED----- ----RKR-FH VSPKG----- ---------- ---------- ---------- ---------I TLVTAEMLGQ 434
12    386 RIPDGMEIGV NLELD----- ----RKR-FH ITEQG----- ---------- ---------- ---------- ---------V VLVTPDMLGQ 421
13    392 QIPEGLKVGF NPDED----- ----RKH-FY VTDDG----- ---------- ---------- ---------- ---------I TLITPEMLGQ 427
14    384 QIPEGLEIGF DPVAD----- ----KKH-FY VTDNG----- ---------- ---------- ---------- ---------I TLITPEMLRQ 419
15    384 IIPPGLEVGF DPVED----- ----RKH-FY VTETG----- ---------- ---------- ---------- ---------V TLVTPEMLGQ 419
16    385 RIGHDVKIIN KDNVQEADRE SQ------GF YIRSG----- ---------- ---------- ---------- ---------I VVVLKNAVIT 424
17    383 AIPDGTVIGE DPVSD----- ----AER-FY RTDDG----- ---------- ---------- ---------- ---------V VLVTPEALRQ 418
18    382 RIPAGMTIGF DAEDD----- ----ARR-FH VSADG----- ---------- ---------- ---------- ---------V VLVTVAMLEA 417
19    395 RIGKNVMIVN KENVQEANRE EL------GF YIRNG----- ---------- ---------- ---------- ---------I VVVIKNVTIA 434
20    392 VIPDGTVIGE DPVED----- ----AKR-FH VTPEG----- ---------- ---------- ---------- ---------I VLVTPRMLGQ 427
21    358 VVPPGARIGY DHDAD----- ----RAAGYH VTEGG----- ---------- ---------- ---------- ---------V VVAPLGDVRF 394
22    397 VIPEGLVIGF DPEQD----- ----RKR-FY VTEKG----- ---------- ---------- ---------- ---------I TLITPEMLGQ 432
23    377 HVPDDFVIGH DPEAD----- ----KAR-FT ISNAG----- ---------- ---------- ---------- ---------V VIVPRGMSL- 411
24    351 KIGEDVEIDG T--------- ------EEVQ VIGYN----- ---------- ---------- ---------- ---------E VVGVPNED-- 379
25    391 RIGENCSIGY EREGYEDGDY G-------YY HVKDG----- ---------- ---------- ---------- ---------I IVIAKNTVLP 429
26    389 HIPEGMVIGE NADED----- ----SAR-FY RSEGGGGVSD SGYAGKVRGK IEPLGFLFVR LDLLIRLSLL IRLNLFIRMN LLIILTLFFK 468
27    380 VIPEGTVIGY NHDDD----- ----RARGFR VSEKG----- ---------- ---------- ---------- ---------V VLVTREMLGQ 416
28    398 VIPEGLVAGY DVKAD----- ----RKR-FY VTEKG----- ---------- ---------- ---------- ---------I TLITPEMLGQ 433
29    395 NIGDGVQLIN KDRLTTYDG- -E------HV FIRDG----- ---------- ---------- ---------- ---------V IIVPRGADLP 432
30    384 VIPDGMVIGE DPAAD----- ----AAR-FY RTESG----- ---------- ---------- ---------- ---------I VLVTREMLRA 419
31    379 TVPANLAIDD INIAK----- ----HFG-FS VSEKG----- ---------- ---------- ---------- ---------V VLITQQVIDS 414
32    352 IIRKNSVVGD G--------- ------RNVT AIGLY----- ---------- ---------- ---------- ---------E EIKTGMVAN- 381
33    388 RIGNGCRIGI DNIPRAEGDY P-------MY SIHDG----- ---------- ---------- ---------- ---------I IVINKNAVIA 426
34    380 RIGDNCQIGV SGKTYEDGEH GPH---G-EF YSSAG----- ---------- ---------- ---------- ---------I IVIRKNAIIP 421
35    397 SIGNGVKLSN LKGYSHYDSP DG------KL FVRDG----- ---------- ---------- ---------- ---------I TIIPRGTKLP 436
36    375 TVPPGFSIGY DREAD----- ----VRR-FP VSESG----- ---------- ---------- ---------- ---------V VIVPNNIDLR 410
37    370 DIPAHERIGF DLEAD----- ----RKR-FH VTASG----- ---------- ---------- ---------- ---------I VVVPRQFFKP 405
38    369 FIPPGMRIGY DREAD----- ----RAR-FH VTDSG----- ---------- ---------- ---------- ---------I VVVAKGQLKP 404
39    380 HIEAGTVIGE DPEAD----- ----AQR-FT VSPNG----- ---------- ---------- ---------- ---------V VLVTPEMLGQ 415
40    382 RIPPGMVIGE DLETD----- ----RER-FH VTPGG----- ---------- ---------- ---------- ---------V VLVTAEMLGQ 417
41    384 VIPDGMVIGE DAELD----- ----SQR-FH RSPNG----- ---------- ---------- ---------- ---------I TLVTKTMLAR 419
42    381 HIPAGLTIGY DREQD----- ----IENGFR VTEKG----- ---------- ---------- ---------- ---------I TLVTSQMLKA 417
43    373 EIAPNTIIGE NLEED----- ----RKN-FT VSDEG----- ---------- ---------- ---------- ---------V IAIAKGSRIG 408
44    371 HIADNVQIGV NLEED----- ----KRR-FT VSEEG----- ---------- ---------- ---------- ---------I VVIPKGARIG 406
45    382 RIPDGLVVGE DPELD----- ----EQR-FR RTENG----- ---------- ---------- ---------- ---------I CLITQPMIDR 417
46    382 KIPEGTVIGY NLEED----- ----KKR-FH VSPKG----- ---------- ---------- ---------- ---------V VLVTPDMLGQ 417
47    371 SIPPGEQIGY DLVKD----- ----RNR-FT VSEKG----- ---------- ---------- ---------- ---------I VVVPKDFVFV 406
48    383 HIPDGTVIGE DPVRD----- ----GER-FY RTDSG----- ---------- ---------- ---------- ---------V VLVTAEALKR 418
49    368 EIPAGAEIGF NLEED----- ----RKK-WF VSEGG----- ---------- ---------- ---------- ---------I VVIPKRAKID 403
50    386 VIPERTVIGE DAEAD----- ----ARR-FH RTEGG----- ---------- ---------- ---------- ---------V VLVTREMLDR 421
51    387 RIGDNVVIIN KDRVEEADKP EL------GF YIRNG----- ---------- ---------- ---------- ---------I VVVVKNATIA 426
52    382 VIPEGLVIGQ EPEED----- ----ARW-FR RSEGG----- ---------- ---------- ---------- ---------I VLVTQDMLDA 417
53    382 KVPDGMVIGE DLEED----- ----KKR-FY VTEEE----- ---------- ---------- ---------- ---------V VLVTPEMLGQ 417
54    373 IIAPGTQIGY DYQED----- ----TRR-YH VSPSG----- ---------- ---------- ---------- ---------I VVVPRGKLSF 408
55    382 KVPPGLVVGV DPEED----- ----AKW-FR RTEGG----- ---------- ---------- ---------- ---------I VLITQSMLDA 417
56    395 VLPEGMVIGE DPELD----- ----AKR-FE RTSNG----- ---------- ---------- ---------- ---------V VLVTKKMLRA 430
57    385 HIPPGMVVGE DPDDD----- ----ARR-FR RTESG----- ---------- ---------- ---------- ---------V TLITRKMLDQ 420
58    386 QVPPKVEIGY DHEQD----- ----RARGFT VTEEG----- ---------- ---------- ---------- ---------L VVVPKSYIFK 422
59    385 RIPDGMEIGV NPEQD----- ----RKR-FY VSEKG----- ---------- ---------- ---------- ---------I TLVTPDMLGQ 420
60    386 VIPDGMKIGL NPEED----- ----AQR-FH VSPGG----- ---------- ---------- ---------- ---------V TLVTPDMLGQ 421
61    386 RIGNDVVITN KKK-IQHQDS E-------FY CIRDG----- ---------- ---------- ---------- ---------I VIIPKNTIVK 423
62    373 VVAPGTIIGE DPELD----- ----RQR-FQ VSEGG----- ---------- ---------- ---------- ---------I VVIPKGARVG 408
63    379 RIGRNVHIRF LP--DRPDSE TD------QW AIRDG----- ---------- ---------- ---------- ---------L VVVPKSAIIP 416
64    399 VIPDEMVIGE DADAD----- ----AAR-FE RTDNG----- ---------- ---------- ---------- ---------V VLVTREMLKR 434
65    345 IIGSECQIGC SGEKKANFEQ PE-ILHSGLN VIGKG----- ---------- ---------- ---------- ---------A EVPAETCIEK 389
66    396 CIPDGFSAGV DPEED----- ----RRY-FY VSPGG----- ---------- ---------- ---------- ---------V TLVTAAALEK 431
67    400 KIPDGMEVGV NIEED----- ----RKR-FH VSENG----- ---------- ---------- ---------- ---------I TLITPEMLGQ 435
68    385 SIPEGMQIGV DLALD----- ----AKR-FY VSEKG----- ---------- ---------- ---------- ---------I TLVTPDMLGQ 420
69    382 NVPDGTVIGE DLEED----- ----KRR-FY VTEEG----- ---------- ---------- ---------- ---------V VLVTPEMLGQ 417
70    395 IIPPGLVVGE DPVED----- ----ARR-FH RTPGG----- ---------- ---------- ---------- ---------V TLITPESLGQ 430
71    381 EIPAGLEIGF DRKTD----- ----EENGFR VSKKG----- ---------- ---------- ---------- ---------I VLVTRDMLSE 417
72    351 VLPDGAVIRS E--------- -----KDIQE VLLVS----- ---------- ---------- ---------- ---------E EFVEKELI-- 380
73    383 TLPEGMVIGE DPVAD----- ----AQR-FD RSEGG----- ---------- ---------- ---------- ---------V VLVTREMLAN 418
```

```
74    387 TIPDGMQIGV DAELD----- ----KQR-FR VSQGG----- ---------- ---------- ---------- ---------V VLVTKTMLKA  422
75    397 CIGNGVRLQN LQGHKDYDSP DG------KL VVRDG----- ---------- ---------- ---------- ---------I IIVPRGTQIP  436
76    377 VIAPGTIIGE DRAED----- ----EKR-FY VDENG----- ---------- ---------- ---------- ---------I VLVTPDMLGQ  412
77    400 VIPEGTVIGY DAEAD----- ----RKA-YT MSAGG----- ---------- ---------- ---------- ---------V VLVTPEMLGQ  435
78    385 RIPDGLVVGE DPDDD----- ----ARR-FY RSEGG----- ---------- ---------- ---------- ---------V TLITPRMLEK  420
79    371 HIPPGERIGF NHAAD----- ----AAR-FV ISESG----- ---------- ---------- ---------- ---------I TVVPKNYHFL  406
80    385 EIPEGTVIGE NRAED----- ----EKR-FH VSPGG----- ---------- ---------- ---------- ---------V VLVTPDMLGQ  420
81    370 KIPPRTKIGV NRAED----- ----AAR-FT ISDRG----- ---------- ---------- ---------- ---------I VVVPESYKFE  405
82    381 VIPAGLIVGH DKAQD----- ----LANGFR VSPKG----- ---------- ---------- ---------- ---------I TLVTSDMLKR  417
83    370 VVGPGEMVGV DLEKD----- ----RER-FA ISAGG----- ---------- ---------- ---------- ---------V VAVGKGVWI-  404
84    387 VIPDGMRIGV DKESD----- ----RKR-FR VSSSGK---- ---------- ---------- ---------- ---------V VLVTPAMLKR  423
85    380 RLPEGTVLGE DHEAD----- ----AKR-FR VTDNG----- ---------- ---------- ---------- ---------I VLVTPKMLGQ  415
86    381 RIPRETVIGE DDVAD----- ----RER-FF VSPKG----- ---------- ---------- ---------- ---------V VLVTAEMLGQ  416
87    380 RIPEGTIIGE DPELD----- ----AQR-FY RSPNG----- ---------- ---------- ---------- ---------V VLVTQEMFAK  415
88    391 VIPDGMEIGV NLELD----- ----RQR-FR VSSGG----- ---------- ---------- ---------- ---------V VLVTPSMLKK  426
89    389 TIPEGTVIGM NAEED----- ----SAR-FH RTEEG----- ---------- ---------- ---------- ---------I VLVTREMLEQ  424
90    351 VMEEGARVIG E---------- -----ETQIT VIPEG----- ---------- ---------- ---------- ---------E TVSAASTAQR  382
91    388 VIPDNFVVGV DKEHD----- ----KARGFR LSSSGK---- ---------- ---------- ---------- ---------V TLVTPTMLTK  425
92    384 KIPDGLVVGE NPEED----- ----AKR-FH VTKNG----- ---------- ---------- ---------- ---------I TLITPEMLGQ  419
93    400 KIPDGFSAGI VPERD----- ----KQC-FH VTERG----- ---------- ---------- ---------- ---------I TLITPEMLGQ  435
94    365 DIAPNSRIGF DAAAD----- ----GAR-YT VSEGG----- ---------- ---------- ---------- ---------V VVVPKG---L  397
95    380 RIPDHTEIGV SDEED----- ----ARR-FY ISPGG----- ---------- ---------- ---------- ---------V RVVTPEMLGQ  415
96    369 HIEPGERIGF DPDVD----- ----RQR-FQ VSESG----- ---------- ---------- ---------- ---------I TVVPRGRASY  404
97    369 EIAAGTIIGE DLELD----- ----RKR-FH VSDEG----- ---------- ---------- ---------- ---------I VVIAKGSKVG  404
98    370 KIPPYTEIGL NPIED----- ----RKR-FH ISERG----- ---------- ---------- ---------- ---------V VVVPESYQFS  405
99    369 DIAPGTQIGV NLKED----- ----KKH-YH VSDDG----- ---------- ---------- ---------- ---------V VVIPKGARVG  404
100   381 KVPDGTQIGL DSLAD----- ----AKR-FH ISKQG----- ---------- ---------- ---------- ---------V IVVPSSYQFE  416
101   352 IIGEKARLGE GPDVPN-EYK PG-IYDSGIT VVGEK----- ---------- ---------- ---------- ---------S SIPADAVIGK  395
102   382 VIPEGLIVGE DPELD----- ----AKR-FR RTESG----- ---------- ---------- ---------- ---------I CLITQSMIDK  417
103   379 TIPDGTTIGY DHGED----- ----RRRGYT VTESG----- ---------- ---------- ---------- ---------I VVVSPAE---  412
104   384 RIGNDVRILN KERPDSADHP ER------GF YIRHG----- ---------- ---------- ---------- ---------I VIVPKDTVIP  423
105   387 RIPEGLVIGE DPILD----- ----AKR-FH VTEQG----- ---------- ---------- ---------- ---------I TLVTPDRLAL  422
106   476 RIGDNVKIIN SDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  515
107   393 CIGRDVKIIN KDNVEESNRE DQ------GF YIRSG----- ---------- ---------- ---------- ---------V VVIIKNAVIP  432
108   387 RIGENVTIIN KDRIEEADRA DQ------GF YIRNG----- ---------- ---------- ---------- ---------I VVVVKNASIL  426
109   384 RIGNDVRILN KERPDSADHP ER------GF YIRHG----- ---------- ---------- ---------- ---------I VIVPKDTVIP  423
110   389 RIGENCQILN EAGVMDKDCE SE------GY IIRDG----- ---------- ---------- ---------- ---------I IVVIKDAVIK  428
111   431 RVGKNCVITN KDNIEDLADE ER------GV FIRNG----- ---------- ---------- ---------- ---------I VTILRNCTIP  470
112   470 RIGPKCQIIN KDGVKEANRE DQ------GF VIKDG----- ---------- ---------- ---------- ---------I VVVIKDSHIP  509
113   460 RVGKNVKIVN KEGVTEGTRE AE------GI YIRSG----- ---------- ---------- ---------- ---------I VVIDKGALVP  499
114   461 RIGKKVVISN SEGVDEADRT SE------GF YIRSG----- ---------- ---------- ---------- ---------I TVVLKNAIIA  500
115   431 RIGDNVKILN ADNVQEAAME TD------GY FIKGG----- ---------- ---------- ---------- ---------I VTVIKDALLP  470
116   466 RIGENVKIIN FDNVQEAVRE TE------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  505
117   470 RIGKNVVIMI SENVQEADRP AE------GY YIRSG----- ---------- ---------- ---------- ---------I TVVLKNAVIL  509
118   472 RIGKNVVITN SKGIQEADHP EE------GY YIRSG----- ---------- ---------- ---------- ---------I VVILKNATIN  511
119   484 KIGKDVIITN ADGVQEADRP SE------GF YIRSG----- ---------- ---------- ---------- ---------I TAVLKNATIK  523
120   484 RIGKNVVIAN SEGIQEADRS ME------GF YIRSG----- ---------- ---------- ---------- ---------V TVILKNSVIQ  523
121   483 KIGKDVIIMN KDGVQEADRE EE------GF YIRSG----- ---------- ---------- ---------- ---------I TIISEKATIE  522
122   478 RIGDNVKIIN GDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  517
123   483 KIGKDAVIVN KDGVQEADRP DD------GF YIRSG----- ---------- ---------- ---------- ---------I TIILEKATIK  522
124   476 RIGKNVVISN SEGIQEADRS LE------GF YIRSG----- ---------- ---------- ---------- ---------I TIILKNFTIK  515
125   465 RIGDNVKIIN SDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALLP  504
126   480 KIGRDVVIAN ADGVQEADRP SE------GF YIRSG----- ---------- ---------- ---------- ---------I TVILKNATIN  519
127   463 RIGKNVIISN SEGVVEADRT SE------GF YIRTG----- ---------- ---------- ---------- ---------V TVVLKNSIIA  502
128   466 RIGENVKIIN FDNVQEAVRE TE------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  505
129   473 RIGKNVVITN SKGIQEADHP EE------GY YIKSG----- ---------- ---------- ---------- ---------I VVILKNATIK  512
130   475 RVGRNVSITN TEGVQEADRP EL------GY YIRSG----- ---------- ---------- ---------- ---------I VVILKNATIK  514
131   422 RIGKNCIITN AAGVEDLEDE EN------GI YIRSG----- ---------- ---------- ---------- ---------I VTILRNATIP  461
132   456 RIGMDCQIIN KDNVQEANHE DK------GY IIKDG----- ---------- ---------- ---------- ---------I VVICKDAIIP  495
133   479 KIGRGVVITN ADGVQEAERP EE------GF YIRSG----- ---------- ---------- ---------- ---------I TVIMENATIN  518
134   487 KIGKDVVIVN KDGVQEADRP EE------GF YIRSG----- ---------- ---------- ---------- ---------I TIIMEKATIE  526
135   484 RIGKNVVIAN KDHVEEADRP SE------GF YIRSG----- ---------- ---------- ---------- ---------I TVVLKNSEIK  523
136   489 RIGKNVVIAN SEGIQEADRS SE------GF YIRSG----- ---------- ---------- ---------- ---------V TIILKNSVIQ  528
137   477 RIGDNVKIIN SDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  516
138   478 RVGKNVIIAN SEGIQEADRS SD------GF YIRSG----- ---------- ---------- ---------- ---------I TVILKNSIIK  517
139   476 KIGKNVIIMN KGDVQEADRP EE------GF YIRLG----- ---------- ---------- ---------- ---------I TVIVEKATIQ  515
140   419 RIGKNVVIAN KDNVQEAEKP SE------GY YIRSG----- ---------- ---------- ---------- ---------I TVILKNATIA  458
141   416 RIGKNVVIAN KDNVQEAEKP SE------GY YIRSG----- ---------- ---------- ---------- ---------I TVILKNATIA  455
142   489 RIGSHVIITN TDGVQEAERP SE------GI YIRSG----- ---------- ---------- ---------- ---------I TVVVKNSILK  528
143   454 RIGSHVIITN TDGVQEAERP SE------GI YIRSG----- ---------- ---------- ---------- ---------I TVVVKNSIVK  493
144   405 RIGRNVRIIN KDNVLEAARE TE------GY FIKNG----- ---------- ---------- ---------- ---------I VTIIKDAVIP  444
```

```
145     415 RIGQN--IIN KDGVQEAARE TD------GF FINCG----- ---------- ---------- ---------- ---------I VTVIKDAVIP  452
146     393 RIGRNVRIIN KDNVLEAARE TE------GY FIKNG----- ---------- ---------- ---------- ---------I VTIIKDAVIP  432
147     403 RIGKNVVIAN KDNVQEAEKP SE------GY YIRSG----- ---------- ---------- ---------- ---------I TVILKNATIA  442
148     468 RVGKNVRIVN KDGVSEGTRE SE------GI YIRSG----- ---------- ---------- ---------- ---------I VVIDKGAKVP  507
149     468 RIGPKCQIIN KDGVKEANRE EQ------GF VIKDG----- ---------- ---------- ---------- ---------I VVVIKDSCIP  507
150     458 RIGMDCQLIN KDNVQEANEE EK------GY IIKDG----- ---------- ---------- ---------- ---------I IVIVKDSYIP  497
151     458 RIGKNCIITN ASGIDDLEDE EN------GV YIRSG----- ---------- ---------- ---------- ---------I VTILRNATIP  497
152     389 RIGENCQILN EAGVMDKDCE NE------GY IIRDG----- ---------- ---------- ---------- ---------I IVVIKDAVIK  428
153     413 RVGKNCVITN AAGVEDLADE ER------GV FIRNG----- ---------- ---------- ---------- ---------I ITILRNCTIP  452
154     475 KIGKDVIIAN KDGVQEADRP ED------GF YIRSG----- ---------- ---------- ---------- ---------I TIIMEKATIE  514
155     486 KIGRNVVIEN IDGVQEADRA KE------GF YIRSG----- ---------- ---------- ---------- ---------I TITLKNATIK  525
156     475 KIGKDVIIAN KDGVQEADRP ED------GF YIRSG----- ---------- ---------- ---------- ---------I TIIMEKATIE  514
157     487 KIGRNVIIAN TDGVQEADRA KE------GF YIRSG----- ---------- ---------- ---------- ---------I TVTLKNATIK  526
158     486 KIGKDVIIMN KDGVQEADRP ED------GF YIRSG----- ---------- ---------- ---------- ---------I TVILEKATIE  525
159     484 KIGKDVIIMN KDGVQEADRP ED------GF YIRSG----- ---------- ---------- ---------- ---------I TVILEKATIE  523
160     479 RIGRNVIIAN TDGVQEADRP ME------GF YIRSG----- ---------- ---------- ---------- ---------I VVVANNATIE  518
161     480 RIGRNVIIAN TDGVQEADRP AE------GF YIRSG----- ---------- ---------- ---------- ---------I VVVVKNATIE  519
162     479 RIGKNVVIAN SEGIQEADRS SE------GF YIRSG----- ---------- ---------- ---------- ---------V TIVLKNSVIE  518
163     460 RIGKNVTIAN SEGVQEADRT SE------GF HIRSG----- ---------- ---------- ---------- ---------I TVVLKNSVIA  499
164     478 RIGRDVIIAN SEGVEEADRA EE------GY YIRSG----- ---------- ---------- ---------- ---------I VVILKNATIK  517
165     473 RIGDNVKIIN VDNVQEAARE TD------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALLP  512
166     470 RIGDDVKIIN SDNVQEAARE TE------GY FIKSG----- ---------- ---------- ---------- ---------I VTVIKDALIP  509
167     484 RIGKDVIIMN KDGVQEADRP ED------GF YIRSG----- ---------- ---------- ---------- ---------I TIVMEKATIE  523
168     471 RIGENVKIIN SDNVQEAARE TE------GY FIKSG----- ---------- ---------- ---------- ---------I VTIIKDALIP  510

1       418 LDLSPX------------------------------------------ 423
2       425 LGHKQER---------------------------------------- 431
3       447 SGIII----------------------------------------- 451
4       466 DGTVI----------------------------------------- 470
5       418 EVAHVR---------------------------------------- 423
6       416 EIHNVYT--------------------------------------- 422
7       404 ---------- ---------- ---------- ---------- ---- 404
8       468 DGFIL----------------------------------------- 472
9       438 DNYVF----------------------------------------- 442
10      437 DNYIF----------------------------------------- 441
11      435 GAAEGR---------------------------------------- 440
12      422 NLHHIR---------------------------------------- 427
13      428 GIHYIR---------------------------------------- 433
14      420 QIHYTR---------------------------------------- 425
15      420 KIHYVR---------------------------------------- 425
16      425 DGTII----------------------------------------- 429
17      419 MPK------------------------------------------- 421
18      418 LRVSQA---------------------------------------- 423
19      435 DGTVI----------------------------------------- 439
20      428 NIYARWEDEY DG--------------------------------- 439
21      395 YAREASR--- HSGGYSE ---------------------------- 408
22      433 EIHKIC---------------------------------------- 438
23      411 ---------- ---------- ---------- ---------- ---- 411
24      379 ---------- ---------- ---------- ---------- ---- 379
25      430 PESRI----------------------------------------- 434
26      469 LASIQASH-------------------------------------- 476
27      417 PVGG------ -----LSPN ------------------------- 424
28      434 QIHKVR---------------------------------------- 439
29      433 DGFII----------------------------------------- 437
30      420 LQ-------------------------------------------- 421
31      415 FADVEIYFDV PQKSTSVSPI RPMHSELSTR NIVRTHVPSA LNER 458
32      381 ---------- ---------- ---------- ---------- ---- 381
33      427 DNTVM----------------------------------------- 431
34      422 PGTVI----------------------------------------- 426
35      437 DNYVF----------------------------------------- 441
36      411 VE-------------------------------------------- 412
37      406 QSLQSTTPSI GRARPSVASA SEGSGLKVAA ------------- 435
38      404 ---------- ---------- ---------- ---------- ---- 404
39      416 EIHVVY---------------------------------------- 421
40      418 EVAHVR---------------------------------------- 423
41      420 IAA------------------------------------------- 422
42      418 LAKK------ ----QMIID ------------------------- 426
43      409 F--------------------------------------------- 409
44      407 F--------------------------------------------- 407
45      418 LDL------------------------------------------- 420
46      418 DDIYG----------------------------------------- 422
```

```
47      407 S---------------------------------------------- 407
48      419 QTAR------------------------------------------- 422
49      404 SP--------------------------------------------- 405
50      422 L---------------------------------------------- 422
51      427 NGTVI------------------------------------------ 431
52      418 RARALG----------------------------------------- 423
53      418 RYRYIR----------------------------------------- 423
54      409 YARNSRG--- KGLGYAE ---------------------------- 422
55      418 RAAALG----------------------------------------- 423
56      431 LGETELNEED DD ---------------------------------- 442
57      421 IA--------------------------------------------- 422
58      423 D---------------------------------------------- 423
59      421 SIHQAR----------------------------------------- 426
60      422 SIHHIR----------------------------------------- 427
61      424 SGTII------------------------------------------ 428
62      409 FV--------------------------------------------- 410
63      417 DGTVI------------------------------------------ 421
64      435 LLAA------------------------------------------- 438
65      390 NCRIFPWIEE DDFRKKAIQS GSTVRPQT ----------------- 417
66      432 MAEWRA----------------------------------------- 437
67      436 IVHSIR----------------------------------------- 441
68      421 DLYRVI----------------------------------------- 426
69      418 KYRYIR----------------------------------------- 423
70      431 QLHFVR----------------------------------------- 436
71      418 LAKK-----L ERQTQENKKL A ----------------------- 433
72      380 ---------- ---------- ---------- ---------- ---- 380
73      419 LAA-------------------------------------------- 421
74      423 LSDKREKAAE D ----------------------------------- 433
75      437 DNYVF------------------------------------------ 441
76      413 HLHMVR----------------------------------------- 418
77      436 PHMVR------------------------------------------ 440
78      421 LQG-------------------------------------------- 423
79      407 TNAPVQR---------------------------------------- 413
80      421 KRHYVR----------------------------------------- 426
81      405 ---------- ---------- ---------- ---------- ---- 405
82      418 MAEKDAKAAL EKAAKQTAEL V ------------------------ 438
83      404 ---------- ---------- ---------- ---------- ---- 404
84      424 LEGGNVPEEG HLD --------------------------------- 436
85      416 KADA------ ------TA ---------------------------- 421
86      417 EVAHVR----------------------------------------- 422
87      416 LYNFEHIPKV E ----------------------------------- 426
88      427 LNGEKIVSEA HLD --------------------------------- 439
89      425 LARKEKRKSN RKKCRTKNFK MRRHFRECPS LLF ------------ 457
90      383 VG--------------------------------------------- 384
91      426 LAKQNEIGKP SI ----------------------------------- 437
92      420 KLHHLR----------------------------------------- 425
93      436 TVHRVR----------------------------------------- 441
94      398 DIPDVTR--- ----YQI ---------------------------- 407
95      416 VSRYVR----------------------------------------- 421
96      405 FARGHTGGGT SGGGYAE ---------------------------- 421
97      405 F---------------------------------------------- 405
98      406 TE--------------------------------------------- 407
99      405 Y---------------------------------------------- 405
100     417 E---------------------------------------------- 417
101     396 NVMIDIGASA VDFTSLNVQS GKSVFKGGVA E ------------- 426
102     418 LDL-------------------------------------------- 420
103     412 ---------- ---------- ---------- ---------- ---- 412
104     424 DGTVI------------------------------------------ 428
105     423 L---------------------------------------------- 423
106     516 TGTVI------------------------------------------ 520
107     433 DGTII------------------------------------------ 437
108     427 DGTII------------------------------------------ 431
109     424 DGTVI------------------------------------------ 428
110     429 AGTVI------------------------------------------ 433
111     471 DGTVI------------------------------------------ 475
112     510 AGTII------------------------------------------ 514
113     500 DNTTI------------------------------------------ 504
114     501 DGLVI------------------------------------------ 505
115     471 SGTVI------------------------------------------ 475
116     506 SGTII------------------------------------------ 510
117     510 NGTKI------------------------------------------ 514
```

```
118     512 DGSVI---------------------------------------- 516
119     524 DGTII---------------------------------------- 528
120     524 DGTVI---------------------------------------- 528
121     523 DGTVI---------------------------------------- 527
122     518 SGTVI---------------------------------------- 522
123     523 DGTVI---------------------------------------- 527
124     516 DGFVI---------------------------------------- 520
125     505 SGTII---------------------------------------- 509
126     520 DGTII---------------------------------------- 524
127     503 DGLVI---------------------------------------- 507
128     506 SGTII---------------------------------------- 510
129     513 DGSVI---------------------------------------- 517
130     515 DGTVI---------------------------------------- 519
131     462 DGTVI---------------------------------------- 466
132     496 NGTVI---------------------------------------- 500
133     519 DGTII---------------------------------------- 523
134     527 DGTVI---------------------------------------- 531
135     524 DGTII---------------------------------------- 528
136     529 DGFVI---------------------------------------- 533
137     517 SGTVI---------------------------------------- 521
138     518 DGVVI---------------------------------------- 522
139     516 DGTVI---------------------------------------- 520
140     459 DGTVI---------------------------------------- 463
141     456 DGTVI---------------------------------------- 460
142     529 DGTVI---------------------------------------- 533
143     494 DGTVI---------------------------------------- 498
144     445 NGTTI---------------------------------------- 449
145     453 NSVVI---------------------------------------- 457
146     433 NGTTI---------------------------------------- 437
147     443 DGTVI---------------------------------------- 447
148     508 DNATI---------------------------------------- 512
149     508 AGTII---------------------------------------- 512
150     498 NGTII---------------------------------------- 502
151     498 DGTVI---------------------------------------- 502
152     429 PGTVI---------------------------------------- 433
153     453 DGTII---------------------------------------- 457
154     515 DGTIV---------------------------------------- 519
155     526 DGTVI---------------------------------------- 530
156     515 DGTVI---------------------------------------- 519
157     527 DGTVI---------------------------------------- 531
158     526 DGTVI---------------------------------------- 530
159     524 DGTVI---------------------------------------- 528
160     519 DGTVI---------------------------------------- 523
161     520 DGTVI---------------------------------------- 524
162     519 DGFII---------------------------------------- 523
163     500 DGLVI---------------------------------------- 504
164     518 DGTVV---------------------------------------- 522
165     513 SGTVI---------------------------------------- 517
166     510 SGTVI---------------------------------------- 514
167     524 DGTVI---------------------------------------- 528
168     511 SGTVI---------------------------------------- 515
```

APPENDIX B

THE LIST OF SEQUENCES USED FOR ALIGNMENT

Sequence #	GI #	Accession #	Organism
1	756169480	WP_042617607	*Agrobacterium tumefaciens*
2	62288123	P0A6V1.2	*Escherichia coli*
3	232164	P23509.2	*Solanum tuberosum (potato)small subunit*
4	232166	Q00081.1	*Solanum tuberosum (potato)large subunit*
5	499947973	WP_011628707	*Alkalilimnicola ehrlichii*
6	499558932	WP_011239715	*Aromatoleum aromaticum*
7	46399384	CAF22833	*Candidatus Protochlamydia amoebophila UWE25*
8	499777686	WP_011458420	*Chlamydia felis*
9	499185705	WP_010883245	*Chlamydia pneumoniae*
10	71845846	AAZ45342	*Dechloromonas aromatica RCB*
11	499798856	WP_011479590	*Methylobacillus flagellatus*
12	499425082	WP_011112546	*Nitrosomonas europaea*
13	499953628	WP_011634362	*Nitrosomonas eutropha*
14	499699341	WP_011380075	*Nitrosospira multiformis*
15	499308001	WP_010998776	*Nostocaceae*
16	499807020	WP_011487754	*Paraburkholderia xenovorans*
17	499782104	WP_011462838	*Rhodoferax ferrireducens*
18	451782402	AGF53371	*Synechocystis sp. PCC 6803*
19	499631839	WP_011312573	*Thiobacillus denitrificans*
20	74056688	AAZ97128	*Thiobacillus denitrificans ATCC 25259*
21	503413850	WP_013648511.1	*Nitrosomonas sp. AL212*
22	504622405	WP_014809507	*Desulfomonile tiedjei*
23	488191742	WP_002262950	*Streptococcus mutans*
24	504267776	WP_014454878	*Spirochaeta africana*
25	488143381	WP_002214589	*Yersinia pestis*
26	503550450	WP_013784526	*Alteromonas naphthalenivorans*

27	503732917	WP_013966993	*Nitrosomonas sp. Is79A3*
28	503690534	WP_013924610	*Parachlamydia acanthamoebae*
29	504242005	WP_014429107	*Rubrivivax gelatinosus*
30	503549068	WP_013783144	*Alteromonas naphthalenivorans*
31	503547512	WP_013781588	*Mahella australiensis*
32	503504895	WP_013739556	*Sphaerochaeta coccoides*
33	503524667	WP_013758887	*Treponema brennaborense*
34	503477853	WP_013712514	*Chlamydia pecorum*
35	503471890	WP_013706551	*Desulfobacca acetoxidans*
36	500120054	WP_011796059	*Acidovorax citrulli*
37	500089517	WP_011765530	*Azoarcus*
38	500137900	WP_011813905	*Halorhodospira halophila*
39	500138647	WP_011814650	*Halorhodospira halophila*
40	500124512	WP_011800517	*Polaromonas naphthalenivorans*
41	500093673	WP_011769686	*Psychromonas ingrahamii*
42	500095282	WP_011771289	*Psychromonas ingrahamii*
43	500095283	WP_011771290	*Psychromonas ingrahamii*
44	499981531	WP_011662249	*Rhodopseudomonas palustris*
45	501464056	WP_012487501	*Cellvibrio japonicus*
46	501878000	WP_012662521	*Desulfobacterium autotrophicum*
47	501369623	WP_012401189	*Paraburkholderia phymatum*
48	161165958	CAN97263	*Sorangium cellulosum So ce56*
49	490447333	WP_004318234	*Thauera*
50	500142271	WP_011818274.1	*Prochlorococcus marinus*
51	500247830	WP_011908172.1	*Rhodobacter sphaeroides*
52	502797186	WP_013032162	*Nitrosococcus halophilus*
53	502797555	WP_013032531	*Nitrosococcus halophilus*
54	502832588	WP_013067564	*Rhodobacter capsulatus*

55	502887802	WP_013122778	*Thiomonas intermedia*
56	506246837	WP_015766612	*Candidatus Accumulibacter phosphatis*
57	502593296	WP_012830868	*Haliangium ochraceum*
58	503207469	WP_013442130	*Methylovorus*
59	502794959	WP_013029935	*Sideroxydans lithotrophicus*
60	501919867	WP_012669639	*Brachyspira hyodysenteriae*
61	506385032	WP_015904751	*Desulfobacterium autotrophicum*
62	223690692	ACN13975	*Desulfobacterium autotrophicum HRM2*
63	502262346	WP_012745576	*Variovorax paradoxus*
64	503043080	WP_013278056	*Acetohalobium arabaticum*
65	503057632	WP_013292608	*Gallionella capsiferriformans*
66	503058355	WP_013293331	*Gallionella capsiferriformans*
67	502913215	WP_013148191	*Methylotenera versatilis*
68	502984855	WP_013219831	*Nitrosococcus watsonii*
69	340555420	AEK57174	*Acidithiobacillus caldus SM-1*
70	119447833	EAW29099	*Alteromonadales bacterium TW-7*
71	296152771	EFG93637	*Bacillus subtilis subsp. spizizenii ATCC 6633*
72	374666046	EHR70831	*Burkholderiales bacterium JOSHI_001*
73	258520883	EEV89742	*Cardiobacterium hominis ATCC 15826*
74	339461259	AEJ77762	*Chlamydia trachomatis L2c*
75	95133625	EAT15287	*Desulfuromonas acetoxidans DSM 684*
76	391858795	EIT69324	*Hydrocarboniphaga effusa AP103*
77	319741407	EFV93832	*Lautropia mirabilis ATCC 51599*
78	344259477	EGW19750	*Methylobacter tundripaludum SV96*
79	344261375	EGW21646	*Methylobacter tundripaludum SV96*
80	149805755	EDM65752	*Moritella sp. PE36*
81	149806046	EDM66029	*Moritella sp. PE36*

82	289712874	EFD76886	*Mycobacterium tuberculosis T85*
83	298282049	EFI23537	*Neisseria sp. oral taxon 014 str. F0314*
84	88789203	EAR20337	*Nitrococcus mobilis Nb-231*
85	88791003	EAR22116	*Nitrococcus mobilis Nb-231*
86	229380302	EEO30393	*Oxalobacter formigenes OXCC13*
87	260631727	EEX49905	*Pasteurella dagmatis ATCC 43325*
88	225200874	EEG83228	*Proteus penneri ATCC 35198*
89	292648071	EFF66043	*Selenomonas noxia ATCC 43541*
90	294483192	EFG30878	*Simonsiella muelleri ATCC 29453*
91	540594576	BAN35736	*Sulfuricella denitrificans skB26*
92	540594643	BAN35803	*Sulfuricella denitrificans skB26*
93	540595079	BAN36239	*Sulfuricella denitrificans skB26*
94	570727232	AHE98540	*Thioalkalivibrio paradoxus ARh 1*
95	570727359	AHE98667	*Thioalkalivibrio paradoxus ARh 1*
96	229333019	EEN98505	*Vibrio cholerae 12129(1)*
97	229334837	EEO00323	*Vibrio cholerae 12129(1)*
98	145963199	EDK28466	*Vibrionales bacterium SWAT-3*
99	145965071	EDK30321	*Vibrionales bacterium SWAT-3*
100	238054288	Q9L385	*[Clostridium] cellulolyticum H10*
101	15890896	NP_356568	*Agrobacterium fabrum str. C58*
102	27381569	NP_773098	*Bradyrhizobium diazoefficiens USDA 110*
103	78195219	ABB32986	*Geobacter metallireducens GS-15*
104	37523829	NP_927206	*Gloeobacter violaceus PCC 7421*
105	83593581	YP_427333	*Rhodospirillum rubrum ATCC 11170*
106	15238933	NP_199641	*Arabidopsis thaliana*
107	22298830	NP_682077	*Thermosynechococcus elongatus BP-1*
108	33240292	NP_875234	*Prochlorococcus marinus subsp. marinus str. CCMP1375*

109	37523829	NP_927206	*Gloeobacter violaceus PCC 7421*
110	145349062	XP_001418959	*Ostreococcus lucimarinus CCE9901*
111	145356323	XP_001422382	*Ostreococcus lucimarinus CCE9901*
112	159467349	XP_001691854	*Chlamydomonas reinhardtii*
113	159470605	XP_001693447	*Chlamydomonas reinhardtii*
114	162460455	NP_001106017	*Zea mays*
115	162461970	NP_001105038	*Zea mays*
116	162462257	NP_001105178	*Zea mays*
117	162463875	NP_001106058	*Zea mays*
118	189027076	NP_001121104	*Zea mays*
119	224062107	XP_002300758	*Populus trichocarpa*
120	224100249	XP_002311802	*Populus trichocarpa*
121	224103389	XP_002313036	*Populus trichocarpa*
122	224131934	XP_002321214	*Populus trichocarpa*
123	225428422	XP_002283855	*Vitis vinifera*
124	225432564	XP_002281069	*Vitis vinifera*
125	225447450	XP_002263255	*Vitis vinifera*
126	225458219	XP_002281223	*Vitis vinifera*
127	242033053	XP_002463921	*Sorghum bicolor*
128	242048788	XP_002462140	*Sorghum bicolor*
129	242053733	XP_002456012	*Sorghum bicolor*
130	242088961	XP_002440313	*Sorghum bicolor*
131	255070935	XP_002507549	*Micromonas commoda*
132	255080070	XP_002503615	*Micromonas commoda*
133	255538708	XP_002510419	*Ricinus communis*
134	255543725	XP_002512925	*Ricinus communis*
135	255548169	XP_002515141	*Ricinus communis*
136	255552303	XP_002517196	*Ricinus communis*

137	255567204	XP_002524583	*Ricinus communis*
138	297812109	XP_002873938	*Arabidopsis lyrata subsp. lyrata*
139	297821353	XP_002878559	*Arabidopsis lyrata subsp. lyrata*
140	302769466	XP_002968152	*Selaginella moellendorffii*
141	302773934	XP_002970384	*Selaginella moellendorffii*
142	302783933	XP_002973739	*Selaginella moellendorffii*
143	302788037	XP_002975788	*Selaginella moellendorffii*
144	302798196	XP_002980858	*Selaginella moellendorffii*
145	302802313	XP_002982912	*Selaginella moellendorffii*
146	302815217	XP_002989290	*Selaginella moellendorffii*
147	302825850	XP_002994500	*Selaginella moellendorffii*
148	302840808	XP_002951950	*Volvox carteri f. nagariensis*
149	302849075	XP_002956068	*Volvox carteri f. nagariensis*
150	303271247	XP_003054985	*Micromonas pusilla CCMP1545*
151	303273364	XP_003056043	*Micromonas pusilla CCMP1545*
152	308806175	XP_003080399	*Ostreococcus tauri*
153	308814250	XP_003084430	*Ostreococcus tauri*
154	356508352	XP_003522921	*Glycine max*
155	356509672	XP_003523570	*Glycine max*
156	356517038	XP_003527197	*Glycine max*
157	356518710	XP_003528021	*Glycine max*
158	356538761	XP_003537869	*Glycine max*
159	356545193	XP_003541029	*Glycine max*
160	356553863	XP_003545270	*Glycine max*
161	356562361	XP_003549440	*Glycine max*
162	356563435	XP_003549968	*Glycine max*
163	357116651	XP_003560093	*Brachypodium distachyon*
164	357132398	XP_003567817	*Brachypodium distachyon*

165	357145851	XP_003573789	*Brachypodium distachyon*
166	357462397	XP_003601480	*Medicago truncatula*
167	357473317	XP_003606943	*Medicago truncatula*
168	357495273	XP_003617925	*Medicago truncatula*

RFERENCE LIST

1. Taylor, S., M.N. Sivak, and J. Preiss, *Starch: Basic Science to Biotechnology*. 1998: Elsevier Science.

2. Preiss, J. and C. Levi, *Starch Biosynthesis and Degradation*. Vol. 3. 1980. 371-423.

3. Whistler, R.L. and E.F. Paschall, *Starch: Chemistry and technology: Fundamental aspects*. 1965: Academic Press.

4. Leloir, L.F. and C.E. Cardini, *BIOSYNTHESIS OF GLYCOGEN FROM URIDINE DIPHOSPHATE GLUCOSE1.* Journal of the American Chemical Society, 1957. **79**(23): p. 6340-6341.

5. Recondo, E. and L.F. Leloir, *Adenosine diphosphate glucose and starch synthesis.* Biochem Biophys Res Commun, 1961. **6**: p. 85-8.

6. Recondo, E., M. Dankert, and L.F. Leloir, *Isolation of adenosine diphosphate D-glucose from corn grains.* Biochem Biophys Res Commun, 1963. **12**: p. 204-7.

7. Ghosh, H.P. and J. Preiss, *Biosynthesis of starch in spinach chloroplasts.* Biochemistry, 1965. **4**(7): p. 1354-61.

8. Preiss, J., E. Greenberg, and A. Sabraw, *Biosynthesis of bacterial glycogen. Kinetic studies of a glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) from a glycogen-deficient mutant of Escherichia coli B.* Journal of Biological Chemistry, 1975. **250**(19): p. 7631-7638.

9. Sivak, M.N., and Preiss, J. , *Starch: basic science to biotechnology, .* in *Advances In Food and Nutrition Reasearch*, 1998((Taylor, S.L., Ed.)): p. 1-199.

10. Espada, J., *Enzymic Synthesis of Adenosine Diphosphate Glucose from Glucose 1-Phosphate and Adenosine Triphosphate.* Journal of Biological Chemistry, 1962. **237**: p. 3577-3581.

11. Preiss, J., *Bacterial glycogen synthesis and its regulation.* Annu Rev Microbiol, 1984. **38**: p. 419-58.

12. Manners, D.J., *Recent developments in our understanding of glycogen structure.* 1991. ***Carbohydrate Polymers 16***: p. 37-82.

13. Strange, R.E., *Bacterial "glycogen" and survival.* Nature, 1968. **220**(5167): p. 606-7.

14. Buleon, A., Colonna, P., Planchot, V., and Ball, S., *Starch granules: structure and biosynthesis.* International Journal of Biological Macromolecules, 1998. **23**: p. 85-112.

15. J., P., *Biology and molecular biology of starch synthesis and its regulation*. Plant Mol. Vol. 7, Oxford University Press. 1991. 59-114.

16. Greenberg, E. and J. Preiss, *The Occurrence of Adenosine Diphosphate Glucose: Glycogen Transglucosylase in Bacteria.* J Biol Chem, 1964. **239**: p. 4314-5.

17. Dickinson, D.B. and J. Preiss, *ADP glucose pyrophosphorylase from maize endosperm.* Arch Biochem Biophys, 1969. **130**(1): p. 119-28.

18. Müller-Röber, B., U. Sonnewald, and L. Willmitzer, *Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes.* Embo j, 1992. **11**(4): p. 1229-38.

19. Lin, T.P., et al., *A Starch Deficient Mutant of Arabidopsis thaliana with Low ADPglucose Pyrophosphorylase Activity Lacks One of the Two Subunits of the Enzyme.* Plant Physiol, 1988. **88**(4): p. 1175-81.

20. Lin, T.P., et al., *Isolation and Characterization of a Starchless Mutant of Arabidopsis thaliana (L.) Heynh Lacking ADPglucose Pyrophosphorylase Activity.* Plant Physiol, 1988. **86**(4): p. 1131-5.

21. Preiss, J., et al., *Biosynthesis of bacterial glycogen. IV. Activation and inhibition of the adenosine diphosphate glucose pyrophosphorylase of Escherichia coli B.* Biochemistry, 1966. **5**(6): p. 1833-45.

22. Preiss, J., *Regulation of the biosynthesis and degradation of starch. Annual review of Plant Physiology*, 1982. **54**: p. 431-454.

23. Preiss, J., *Biochemistry of the starch and its regulation. The Biochemistry of Plants*, 1988. **(Preiss, J., Ed.)**(Academic Press, New York): p. 181-254.

24. preiss, j., and Romeo, T. (1994), *Molecular biology and regulatory aspects of glycogen biosynthesis in bacteria. Progress in nucleic acid researchand molecular biology*, 1994. **(Cohn, W.E., and Moldave, K., Eds.)**(Academic Press, New York): p. 299-329.

25. Figueroa, C.M., et al., *Understanding the allosteric trigger for the fructose-1,6-bisphosphate regulation of the ADP-glucose pyrophosphorylase from Escherichia coli.* Biochimie, 2011. **93**(10): p. 1816-23.

26. Meyer, C.R., Ghosh, P., Remy, E., and Preiss, J., *Clonning, expression, and nucleotide sequence of a mutant glgC gene from Escherichia coli B. Journal of Bacteriology*, 1992. **174**: p. 4509-4512.

27. Meyer, C.R., Ghosh, P., Nadler, S., and Preiss, J., *Cloning, expression, and sequence of an allosteric mutant ADP glucose pyrophosphorylase from Escherichia coli B. Arch. Biochem. Biophys.*, 1993. **302**: p. 64-71.

28. Giroux, M.J., et al., *A single mutation that increases maize seed weight.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 5824-9.

29. Ballicora, M.A., et al., *Activation of the potato tuber ADP-glucose pyrophosphorylase by thioredoxin.* J Biol Chem, 2000. **275**(2): p. 1315-20.

30. Stark, D.M., et al., *Regulation of the Amount of Starch in Plant Tissues by ADP Glucose Pyrophosphorylase.* Science, 1992. **258**(5080): p. 287-92.

31. Li, L. and J. Preiss, *Characterization of ADPglucose pyrophosphorylase from a starch-deficient mutant of Arabidopsis thaliana (L.).* Carbohydrate Research, 1992. **227**: p. 227-239.

32. Neuhaus, H.E., et al., *Control of photosynthate partitioning in spinach leaves.* Planta, 1990. **181**(4): p. 583-592.

33. Smidansky, E.D., et al., *Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield.* Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1724-9.

34. Ball, S., et al., *A Chlamydomonas reinhardtii low-starch mutant is defective for 3-phosphoglycerate activation and orthophosphate inhibition of ADP-glucose pyrophosphorylase.* Planta, 1991. **185**(1): p. 17-26.

35. Geigenberger, P., M. Stitt, and A.R. Fernie, *Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers.* Plant, Cell & Environment, 2004. **27**(6): p. 655-673.

36. Figueroa, C.M., et al., *Unraveling the activation mechanism of the potato tuber ADP-glucose pyrophosphorylase.* PLoS One, 2013. **8**(6): p. e66824.

37. Kumar, A., C.E. Larsen, and J. Preiss, *Biosynthesis of bacterial glycogen. Primary structure of Escherichia coli ADP-glucose:alpha-1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the glgA gene.* J Biol Chem, 1986. **261**(34): p. 16256-9.

38. Furukawa, K., et al., *Identification of lysine 15 at the active site in Escherichia coli glycogen synthase. Conservation of Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes.* J Biol Chem, 1990. **265**(4): p. 2086-90.

39. Holmes, E., C. Boyer, and J. Preiss, *Immunological characterization of Escherichia coli B glycogen synthase and branching enzyme and comparison with enzymes from other bacteria.* J Bacteriol, 1982. **151**(3): p. 1444-53.

40. Holmes, E. and J. Preiss, *Characterization of Escherichia coli B glycogen synthase enzymatic reactions and products.* Archives of Biochemistry and Biophysics, 1979. **196**(2): p. 436-448.

41. Buschiazzo, A., et al., *Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation.* Embo j, 2004. **23**(16): p. 3196-205.

42. Coutinho, P.M., et al., *An Evolving Hierarchical Family Classification for Glycosyltransferases.* Journal of Molecular Biology, 2003. **328**(2): p. 307-317.

43. Roach, P.J., *Glycogen and its metabolism.* Curr Mol Med, 2002. **2**(2): p. 101-20.

44. Ball, S.G. and M.K. Morell, *From Bacterial Glycogen to Starch: Understanding the Biogenesis of the Plant Starch Granule.* Annual Review of Plant Biology, 2003. **54**(1): p. 207-233.

45. Baba, T., et al., *Identification, cDNA cloning, and gene expression of soluble starch synthase in rice (Oryza sativa L.) immature seeds.* Plant Physiol, 1993. **103**(2): p. 565-73.

46. John, W., S. Downton, and J.S. Hawker, *Enzymes of starch and sucrose metabolism in Zea mays leaves.* Phytochemistry, 1973. **12**(7): p. 1551-1556.

47. Macdonald, F.D. and J. Preiss, *Partial purification and characterization of granule-bound starch synthases from normal and waxy maize.* Plant Physiol, 1985. **78**(4): p. 849-52.

48. Flipse, E., et al., *Expression of a wild-type GBSS gene introduced into an amylose-free potato mutant by Agrobacterium tumefaciens and the inheritance of the inserts at the microsporic level.* Theoretical and Applied Genetics, 1994. **88**(3): p. 369-375.

49. Nakamura, Y., et al., *Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties.* Plant Mol Biol, 2005. **58**(2): p. 213-27.

50. Yep, A., et al., *Identification and characterization of a critical region in the glycogen synthase from Escherichia coli.* J Biol Chem, 2004. **279**(9): p. 8359-67.

51. Borovsky, D., E.E. Smith, and W.J. Whelan, *On The Mechanism of Amylose Branching by Potato Q-Enzyme.* European Journal of Biochemistry, 1976. **62**(2): p. 307-312.

52. Baecker, P.A., E. Greenberg, and J. Preiss, *Biosynthesis of bacterial glycogen. Primary structure of Escherichia coli 1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-alpha-D-(1, 4-*

alpha-D-glucano)-transferase as deduced from the nucleotide sequence of the glg B gene. J Biol Chem, 1986. **261**(19): p. 8738-43.

53. Kiel, J.A.K.W., et al., *Nucleotide sequence of the Synechococcus sp. PCC7942 branching enzyme gene (glgB): expression in Bacillus subtilis.* Gene, 1990. **89**(1): p. 77-84.

54. Kiel, J.A.K.W., et al., *Molecular cloning and nucleotide sequence of the glycogen branching enzyme gene (glgB) from Bacillus stearothermophilus and expression in Escherichia coli and Bacillus subtilis.* Molecular and General Genetics MGG, 1991. **230**(1): p. 136-144.

55. Rumbak, E., et al., *Characterization of the Butyrivibrio fibrisolvens glgB gene, which encodes a glycogen-branching enzyme with starch-clearing activity.* J Bacteriol, 1991. **173**(21): p. 6732-41.

56. Takata, H., et al., *Properties and active center of the thermostable branching enzyme from Bacillus stearothermophilus.* Appl Environ Microbiol, 1994. **60**(9): p. 3096-104.

57. Schwall, G.P., et al., *Production of very-high-amylose potato starch by inhibition of SBE A and B.* Nature Biotechnology, 2000. **18**: p. 551.

58. Kawabata, Y., et al., *Preparation of Highly Branched Starch by Glycogen Branching Enzyme from Neurospora crassa N2-44 and Its Characterization*. Vol. 49. 2002. 273-279.

59. Takata, H., et al., *Properties of Branching Enzyme from Hyperthermophilic Bacterium, Aquifex aeolicus, and Its Potential for Production of Highly-branched Cyclic Dextrin*. Vol. 50. 2003. 15-20.

60. Takata, H., et al., *Production and some properties of a dextrin with a narrow size distribution by the cyclization reaction of branching enzyme.* Journal of Fermentation and Bioengineering, 1997. **84**(2): p. 119-123.

61. Ballicora, M.A., A.A. Iglesias, and J. Preiss, *ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis.* Microbiol Mol Biol Rev, 2003. **67**(2): p. 213-25, table of contents.

62. Ballicora, M.A., A.A. Iglesias, and J. Preiss, *ADP-Glucose Pyrophosphorylase: A Regulatory Enzyme for Plant Starch Synthesis.* Photosynth Res, 2004. **79**(1): p. 1-24.

63. Igarashi, R.Y. and C.R. Meyer, *Cloning and sequencing of glycogen metabolism genes from Rhodobacter sphaeroides 2.4.1. Expression and characterization of recombinant ADP-glucose pyrophosphorylase.* Arch Biochem Biophys, 2000. **376**(1): p. 47-58.

64. Iglesias, A.A., et al., *Characterization of the kinetic, regulatory, and structural properties of ADP-glucose pyrophosphorylase from Chlamydomonas reinhardtii.* Plant Physiol, 1994. **104**(4): p. 1287-94.

65. Iglesias, A.A. and J. Preiss, *Bacterial glycogen and plant starch biosynthesis.* Biochemical Education, 1992. **20**(4): p. 196-203.

66. Takata, H., et al., *Characterization of a gene cluster for glycogen biosynthesis and a heterotetrameric ADP-glucose pyrophosphorylase from Bacillus stearothermophilus.* J Bacteriol, 1997. **179**(15): p. 4689-98.

67. Uttaro, A.D., et al., *Cloning and expression of the glgC gene from Agrobacterium tumefaciens: purification and characterization of the ADPglucose synthetase.* Arch Biochem Biophys, 1998. **357**(1): p. 13-21.

68. Asención Diez, M.D., et al., *Characterization of Recombinant UDP- and ADP-Glucose Pyrophosphorylases and Glycogen Synthase To Elucidate Glucose-1-Phosphate Partitioning into Oligo- and Polysaccharides in Streptomyces coelicolor.* J Bacteriol, 2012. **194**(6): p. 1485-93.

69. Okita, T.W., et al., *The Subunit Structure of Potato Tuber ADPglucose Pyrophosphorylase.* Plant Physiol, 1990. **93**(2): p. 785-90.

70. Jin, X., et al., *Crystal structure of potato tuber ADP-glucose pyrophosphorylase.* Embo j, 2005. **24**(4): p. 694-704.

71. Hannah, L.C. and O.E. Nelson, Jr., *Characterization of ADP-glucose pyrophosphorylase from shrunken-2 and brittle-2 mutants of maize.* Biochem Genet, 1976. **14**(7-8): p. 547-60.

72. Burton, R.A., et al., *Characterization of the genes encoding the cytosolic and plastidial forms of ADP-glucose pyrophosphorylase in wheat endosperm.* Plant Physiol, 2002. **130**(3): p. 1464-75.

73. Chen, B.Y. and H.W. Janes, *Multiple forms of ADP-glucose pyrophosphorylase from tomato fruit.* Plant Physiol, 1997. **113**(1): p. 235-41.

74. Chen, B.Y., H.W. Janes, and T. Gianfagna, *PCR cloning and characterization of multiple ADP-glucose pyrophosphorylase cDNAs from tomato.* Plant Sci, 1998. **136**(1): p. 59-67.

75. Crevillen, P., et al., *Differential pattern of expression and sugar regulation of Arabidopsis thaliana ADP-glucose pyrophosphorylase-encoding genes.* J Biol Chem, 2005. **280**(9): p. 8143-9.

76. Kim, I.J., et al., *Molecular characterization of cDNA clones for ADP-glucose pyrophosphorylase from Citrus.* Biochim Biophys Acta, 2001. **1518**(3): p. 324-8.

77. La Cognata, U., L. Willmitzer, and B. Muller-Rober, *Molecular cloning and characterization of novel isoforms of potato ADP-glucose pyrophosphorylase.* Mol Gen Genet, 1995. **246**(5): p. 538-48.

78. Muller-Rober, B., G. Nast, and L. Willmitzer, *Isolation and expression analysis of cDNA clones encoding a small and a large subunit of ADP-glucose pyrophosphorylase from sugar beet.* Plant Mol Biol, 1995. **27**(1): p. 191-7.

79. Smith-White, B.J. and J. Preiss, *Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources.* J Mol Evol, 1992. **34**(5): p. 449-64.

80. Weber, H., et al., *Cell-type specific, coordinate expression of two ADP-glucose pyrophosphorylase genes in relation to starch biosynthesis during seed development of Vicia faba L.* Planta, 1995. **195**(3): p. 352-61.

81. Crevillen, P., et al., *The different large subunit isoforms of Arabidopsis thaliana ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme.* J Biol Chem, 2003. **278**(31): p. 28508-15.

82. Kuhn, M.L., C.A. Falaschetti, and M.A. Ballicora, *Ostreococcus tauri ADP-glucose Pyrophosphorylase Reveals Alternative Paths for the Evolution of Subunit Roles.* Journal of Biological Chemistry, 2009. **284**(49): p. 34092-34102.

83. Cupp-Vickery, J.R., R.Y. Igarashi, and C.R. Meyer, *Preliminary crystallographic analysis of ADP-glucose pyrophosphorylase from Agrobacterium tumefaciens.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2005. **61**(Pt 3): p. 266-8.

84. Kumar, A., et al., *Biosynthesis of bacterial glycogen. Use of site-directed mutagenesis to probe the role of tyrosine 114 in the catalytic mechanism of ADP-glucose synthetase from Escherichia coli.* J Biol Chem, 1988. **263**(29): p. 14634-9.

85. Cupp-Vickery, J.R., et al., *Structural analysis of ADP-glucose pyrophosphorylase from the bacterium Agrobacterium tumefaciens.* Biochemistry, 2008. **47**(15): p. 4439-51.

86. Preiss, J., L. Shen, and M. Partridge, *The Activation of Escherichia Coli Adp-Glucose Pyrophosphorylase.* Biochem Biophys Res Commun, 1965. **18**: p. 180-5.

87. Murata, R.M., et al., *Inhibition of Streptococcus mutans biofilm accumulation and development of dental caries in vivo by 7-epiclusianone and fluoride.* Biofouling, 2010. **26**(7): p. 865-72.

88. Wilson, W.A., et al., *Regulation of glycogen metabolism in yeast and bacteria.* FEMS Microbiol Rev, 2010. **34**(6): p. 952-85.

89. Ghosh, H.P. and J. Preiss, *Adenosine diphosphate glucose pyrophosphorylase. A regulatory enzyme in the biosynthesis of starch in spinach leaf chloroplasts.* J Biol Chem, 1966. **241**(19): p. 4491-504.

90. Asención Diez, M.D., et al., *A Novel Dual Allosteric Activation Mechanism of Escherichia coli ADP-Glucose Pyrophosphorylase: The Role of Pyruvate.* PLoS One, 2014. **9**(8).

91. Preiss, J., *Bacterial Glycogen Inclusions: Enzymology and Regulation of Synthesis*, in *Inclusions in Prokaryotes*, J.M. Shively, Editor. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 71-108.

92. Yung, S.G. and J. Preiss, *Biosynthesis of bacterial glycogen: purification and structural and immunological properties of Rhodopseudomonas sphaeroides ADPglucose synthetase.* J Bacteriol, 1982. **151**(2): p. 742-9.

93. Preiss, J. and E. Greenberg, *Biosynthesis of starch in Chlorella pyrenoidosa. I. Purification and properties of the adenosine diphosphoglucose: alpha-1, 4-glucan, alpha-4-glucosyl transferase from Chlorella.* Arch Biochem Biophys, 1967. **118**(3): p. 702-8.

94. Sanwal, G.G. and J. Preiss, *Biosynthesis of starch in Chlorella pyrenoidosa. II. Regulation of ATP: alpha-D-glucose 1-phosphate adenyl transferase (ADP-glucose pyrophosphorylase) by inorganic phosphate and 3-phosphoglycerate.* Arch Biochem Biophys, 1967. **119**(1): p. 454-69.

95. Asencion Diez, M.D., et al., *The ADP-glucose pyrophosphorylase from Streptococcus mutans provides evidence for the regulation of polysaccharide biosynthesis in Firmicutes.* Mol Microbiol, 2013. **90**(5): p. 1011-27.

96. Cereijo, A.E., et al., *On the Kinetic and Allosteric Regulatory Properties of the ADP-Glucose Pyrophosphorylase from Rhodococcus jostii: An Approach to Evaluate Glycogen Metabolism in Oleaginous Bacteria.* Front Microbiol, 2016. **7**.

97. Iglesias, A.A., G. Kakefuda, and J. Preiss, *Regulatory and Structural Properties of the Cyanobacterial ADPglucose Pyrophosphorylases.* Plant Physiol, 1991. **97**(3): p. 1187-95.

98. Ballicora, M.A., et al., *Adenosine 5'-diphosphate-glucose pyrophosphorylase from potato tuber. Significance of the N terminus of the small subunit for catalytic properties and heat stability.* Plant Physiol, 1995. **109**(1): p. 245-51.

99. Iglesias, A.A., et al., *Expression of the potato tuber ADP-glucose pyrophosphorylase in Escherichia coli.* J Biol Chem, 1993. **268**(2): p. 1081-6.

100. Gomez-Casati, D.F. and A.A. Iglesias, *ADP-glucose pyrophosphorylase from wheat endosperm. Purification and characterization of an enzyme with novel regulatory properties.* Planta, 2002. **214**(3): p. 428-34.

101. Kleczkowski, L.A., et al., *Insensitivity of barley endosperm ADP-glucose pyrophosphorylase to 3-phosphoglycerate and orthophosphate regulation.* Plant Physiol, 1993. **101**(1): p. 179-86.

102. Fu, Y., et al., *Mechanism of reductive activation of potato tuber ADP-glucose pyrophosphorylase.* J Biol Chem, 1998. **273**(39): p. 25045-52.

103. E W Nester, et al., *Crown Gall: A Molecular and Physiological Analysis.* Annual Review of Plant Physiology, 1984. **35**(1): p. 387-413.

104. de la Riva, G.A., J. Gonzalez-Cabrera, R. Vazquez-Padron, and C. AyraPardo, *Agrobacterium tumefaciens: a natural tool for plant transformation.* Electron. J. Biotechnol. , 1998. **1**: p. 1-16.

105. Xu, X., *Starch Meets Biotechnology - In Planta Modification of Starch Composition and. Functionalities.* 2016.

106. Ballicora, M.A., et al., *Characterization of chimeric ADPglucose pyrophosphorylases of Escherichia coli and Agrobacterium tumefaciens. Importance of the C-terminus on the selectivity for allosteric regulators.* Biochemistry, 2002. **41**(30): p. 9431-7.

107. Cifuente, Javier O., et al., *Structural Basis of Glycogen Biosynthesis Regulation in Bacteria.* Structure. **24**(9): p. 1613-1622.

108. Benjamin L. Hill, R.M., Hiral Patel, Matías D. Asencion Diez, Rui Wu, and D.L. Alberto A. Iglesias, and Miguel A. Ballicora, *The Pyruvate Regulatory Site of the Agrobacterium tumefaciens ADP-glucose pyrophosphorylase*

109. Gomez-Casati, D.F., et al., *Identification of functionally important amino-terminal arginines of Agrobacterium tumefaciens ADP-glucose pyrophosphorylase by alanine scanning mutagenesis.* Biochemistry, 2001. **40**(34): p. 10169-78.

110. Gardiol, A. and J. Preiss, *Escherichia coli E-39 ADPglucose synthetase has different activation kinetics from the wild-type allosteric enzyme.* Arch Biochem Biophys, 1990. **280**(1): p. 175-80.

111. Sheng, J., Y.Y. Charng, and J. Preiss, *Site-directed mutagenesis of lysine382, the activator-binding site, of ADP-glucose pyrophosphorylase from Anabaena PCC 7120.* Biochemistry, 1996. **35**(9): p. 3115-21.

112. Ballicora, M.A., et al., *ADP-Glucose Pyrophosphorylase from Potato Tubers.* Plant Physiol, 1998. **118**(1): p. 265-74.

113. Boehlein, S.K., et al., *Probing Allosteric Binding Sites of the Maize Endosperm ADP-Glucose Pyrophosphorylase.* Plant Physiol, 2010. **152**(1): p. 85-95.

114. Meyer, C.R., et al., *Site-directed mutagenesis of a regulatory site of Escherichia coli ADP-glucose pyrophosphorylase: the role of residue 336 in allosteric behavior.* Arch Biochem Biophys, 1998. **353**(1): p. 152-9.

115. Meyer, C.R., et al., *A kinetic study of site-directed mutants of Escherichia coli ADP-glucose pyrophosphorylase: the role of residue 295 in allosteric regulation.* Arch Biochem Biophys, 1998. **352**(2): p. 247-54.

116. Ballicora, M.A., et al., *Heat Stability of the Potato Tuber ADP-Glucose Pyrophosphorylase: Role of Cys Residue 12 in the Small Subunit.* Biochemical and Biophysical Research Communications, 1999. **257**(3): p. 782-786.

117. Greene, T.W. and L.C. Hannah, *Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions.* Proc Natl Acad Sci U S A, 1998. **95**(22): p. 13342-7.

118. Lyerly Linebarger, C.R., et al., *Heat Stability of Maize Endosperm ADP-Glucose Pyrophosphorylase Is Enhanced by Insertion of a Cysteine in the N Terminus of the Small Subunit.* Plant Physiol, 2005. **139**(4): p. 1625-34.

119. Greene, T.W., et al., *Mutagenesis of the potato ADPglucose pyrophosphorylase and characterization of an allosteric mutant defective in 3-phosphoglycerate activation.* Proceedings of the National Academy of Sciences, 1996. **93**(4): p. 1509-1513.

120. Kim, D., S.K. Hwang, and T.W. Okita, *Subunit interactions specify the allosteric regulatory properties of the potato tuber ADP-glucose pyrophosphorylase.* Biochem Biophys Res Commun, 2007. **362**(2): p. 301-6.

121. Ventriglia, T., et al., *Regulatory Properties of Potato–Arabidopsis Hybrid ADP-Glucose Pyrophosphorylase.* Plant and Cell Physiology, 2007. **48**(6): p. 875-880.

122. Iglesias, A.A., et al., *Domain Swapping between a Cyanobacterial and a Plant Subunit ADP-Glucose Pyrophosphorylase.* Plant and Cell Physiology, 2006. **47**(4): p. 523-530.

123. Hwang, S.K., P.R. Salamone, and T.W. Okita, *Allosteric regulation of the higher plant ADP-glucose pyrophosphorylase is a product of synergy between the two subunits.* FEBS Lett, 2005. **579**(5): p. 983-90.

124. Salamone, P.R., et al., *Directed molecular evolution of ADP-glucose pyrophosphorylase.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 1070-5.

125. Georgelis, N., J.R. Shaw, and L.C. Hannah, *Phylogenetic analysis of ADP-glucose pyrophosphorylase subunits reveals a role of subunit interfaces in the allosteric properties of the enzyme.* Plant Physiol, 2009. **151**(1): p. 67-77.

126. Hill, B.L., et al., *The Pyruvate Regulatory Site of the Agrobacterium tumefaciens ADP-glucose pyrophosphorylase*

127. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

128. Tuncel, A., I.H. Kavakli, and O. Keskin, *Insights into Subunit Interactions in the Heterotetrameric Structure of Potato ADP-Glucose Pyrophosphorylase.* Biophys J, 2008. **95**(8): p. 3628-39.

129. Greene, T.W. and L.C. Hannah, *Maize endosperm ADP-glucose pyrophosphorylase SHRUNKEN2 and BRITTLE2 subunit interactions.* Plant Cell, 1998. **10**(8): p. 1295-306.

VITA

Hiral P. Patel was born in Gujarat, India on June 29, 1987. Before attending Loyola University Chicago, she attended the S.P. University, Gujarat, India, where she earned a Bachelor of Science in Biochemistry. After college, from the same university she received her Master of Science in Biochemistry in 2010.