



---

Dissertations

Theses and Dissertations


---

2018

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

Hiral Priyank Patel

Follow this and additional works at: [https://ecommons.luc.edu/luc\\_diss](https://ecommons.luc.edu/luc_diss)

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

---

### Recommended Citation

Patel, Hiral Priyank, "Regulation of Bacterial Glycogen Synthesis: Structure-Function Relationship of Adp-Glucose Pyrophosphorylase" (2018). *Dissertations*. 2978.

[https://ecommons.luc.edu/luc\\_diss/2978](https://ecommons.luc.edu/luc_diss/2978)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).  
Copyright © 2017 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

Copyright by Hiral P. Patel, 2018  
All rights reserved

## 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

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBRIVIATION	x
ABSTRACT	xi
CHAPTER ONE: STARCH AND GLYCOGEN SYNTHESIS AND ADP-GLUCOSE PYROPHOSPHORYLASE	
Introduction	1
Sugar-nucleotides in biosynthesis reaction	3
Glycogen in bacteria	3
Starch in higher plants	4
The reaction involved in glycogen and starch synthesis	4
ADP-Glucose pyrophosphorylase	5
Glycogen synthase in bacteria or starch synthase in plants	7
Branching enzymes	9
Structure of the ADP-Glc PPase	9
Allosterically regulated ADP-glucose pyrophosphorylase	13
<i>Agrobacterium tumefaciens</i> used for plant transformation	14
ADP-Glc PPase from <i>A. tumefaciens</i>	15
Residues involved in activator binding	17
N-terminus residues of the <i>A. tumefaciens</i> ADP-Glc PPase	17
C-terminus residues of the <i>A. tumefaciens</i> ADP-Glc PPase	17
<i>E. coli</i> ADP-Glc PPase regions involved in regulation	18
Plant ADP-Glc PPase residues in regulation	18
Allosteric mutants (pre-activated)	19
Inter-subunit interaction in ADP-Glc PPase	19
Potato tuber ADP-Glc PPase small subunit inter-subunit interaction	19
Subunit interaction in allosteric regulation	21
Hybrid ADP-Glc PPase	22
Objectives and outline of thesis	23
CHAPTER TWO: A CRITICAL AND SPECIFIC REVIEW OF RESIDUE FOR THE PYRUVATE REGULATORY SITE OF THE ADP-GLUCOSE PYROPHOSPHORYLASE FROM <i>A. TUMEFACIENS</i>	24
Abstract	24
Introduction	25
Results	27

The structural comparison of wild-type and the mutant K43A ADP-Glc PPases of <i>A. tumefaciens</i>	27
Effect of activators (Fru6P, Pyruvate) on the saturation curves of ATP	29
Effect on the activation by Fru6P and Pyruvate	29
Binding analysis by Thermal shift assay	34
Positive charge is crucial for binding pyruvate and signaling	34
Discussion	39
Positive charge is crucial for binding pyruvate	42
The Aspartate mutation of G329 makes the enzyme pre-activated	42
Materials and Methods	45
Site-Directed Mutagenesis	45
Expression and Purification	45
Enzyme assay	46
Substrate saturation assay	46
Activator saturation assay	48
Kinetics	48
Thermal shift assay	48
Heat stability assay	49
Crystallization and data collection	49
CHAPTER THREE: IDENTIFICATION AND INVOLVEMENT OF INTER-SUBUNIT SURFACE INTERACTION IN REGULATION OF THE ADP-GLUCOSE PYROPHOSPHORYLASE FROM <i>AGROBACTERIUM TUMEFACIENS</i>	50
Abstract	50
Introduction	51
Results	54
Sequence analysis	54
Gel-filtration analysis	55
Effect of activators (Fru6P, Pyruvate) on the saturation curves of ATP	55
The catalytic efficiency	58
Activation at saturating concentrations of substrate (ATP)	60
Activation at sub-saturating concentrations of substrate (ATP)	60
Activation rescue by a secondary mutation (R11D/D141R)	64
Discussion	67
Structural comparison between plant, and bacterial ADP-Glc PPase	67
The effect of mutations at interface NN-1	70
Materials and Methods	71
Site-Directed Mutagenesis	71
Expression and Purification	72
Enzyme assay	73
Substrate saturation assay	73
Activator saturation assay	75
Kinetics	75



CHAPTER FOUR: CONCLUSION AND FUTURE IMPLICATIONS	76
Understanding the allosteric regulation of ADP-Glc PPase	76
Interface inter-subunit interaction in regulation	78
APPENDIX A: SEQUENCE ALIGNMENT OF PLANTS AND BACTERIAL ADP-GLC PPASES	80
APPENDIX B: THE LIST OF SEQUENCES USED FOR ALIGNMENT	98
REFERENCE LIST	107
VITA	118

## LIST OF TABLES

Table 1. Classification of the ADP-Glc PPases based on the structure, regulatory properties, and carbon metabolism	11
Table 2: Structural comparison between - wild-type and the K43A mutant ADP-Glc PPase	30
Table 3: Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants	31
Table 4. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.	33
Table 5. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.	36
Table 6. Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants	57
Table 7. Kinetic parameters for the substrate (ATP) of ADP-Glc PPase wild type, and mutants	59
Table 8. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants	61

## LIST OF FIGURES

Figure 1. The structure of glycogen and starch	2
Figure 2. The model of the glycogen and starch biosynthesis pathway	6
Figure 3. The crystal structure of the ADP-Glc PPase from <i>A. tumefaciens</i>	16
Figure 4. Inter-subunit interactions of the homotetrameric potato tuber crystal structure (PDB: 1YP3)	20
Figure 5. Crystallographic structure of <i>A. tumefaciens</i> ADP-Glc PPase with pyruvate bound (PDB: 5W5R)	26
Figure 6. The aspartate at 329 mimics the oxygens of the pyruvate	28
Figure 7. Activator saturation curves of wild type and mutant ADP-Glc PPase from <i>A. tumefaciens</i> at saturating concentrations of ATP	32
Figure 8. Thermal shift analysis of wild-type and mutant ADP-Glc PPase.	35
Figure 9. Heat stability assay for K43A mutant ADP-Glc PPase	37
Figure 10. Effect of Li <sub>2</sub> SO <sub>4</sub> on the activation by pyruvate of the <i>A. tumefaciens</i> ADP-Glc PPase	40
Figure 11. Kinetic characterization of ADP-Glc PPase activation by pyruvate analogs	41
Figure 12: The alignment of K43A mutant with the wild type	43
Figure 13. The model for ADP-Glc PPase from <i>A. tumefaciens</i> wild-type, and mutants K43A, G329D, and K43A/G329D	44
Figure 14. The colorimetric malachite green enzymatic assay	47
Figure 15. The interface Inter-subunit interaction in <i>Agrobacterium tumefaciens</i> and Potato tuber.	53
Figure 16. Analytical gel filtration chromatography of ADP-Glc PPases wild type (WT), and mutant (D141A, D141E, D141N, D141R, R11D, R11K, and R11D/D141R)	56

Figure 17. Activator saturation curves of wild type and mutant ADP-Glc PPase from <i>A. tumefaciens</i> at saturating concentrations of ATP.	62
Figure 18. Relative activation curves of wild type and mutant ADP-Glc PPase from <i>A. tumefaciens</i> at sub-saturating concentrations of ATP	63
Figure 19. Substrate saturation curve of <i>A. tumefaciens</i> ADP-Glc PPase double mutant R11D/D141R	65
Figure 20. Effect of Fru6P and Pyr on the apparent affinities for ATP in the wild type (WT), and mutant (R11D, D141R, and R11D/D141R), <i>A. tumefaciens</i> ADP-Glc PPases.	66
Figure 21. The homotetramer and the dimer interface interaction in ADP-Glc PPase <i>A. tumefaciens</i>	68
Figure 22. Gel filtration chromatography of wild-type (WT) and the D141R mutant ADP-Glc PPase	74

## 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- $\beta$ -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

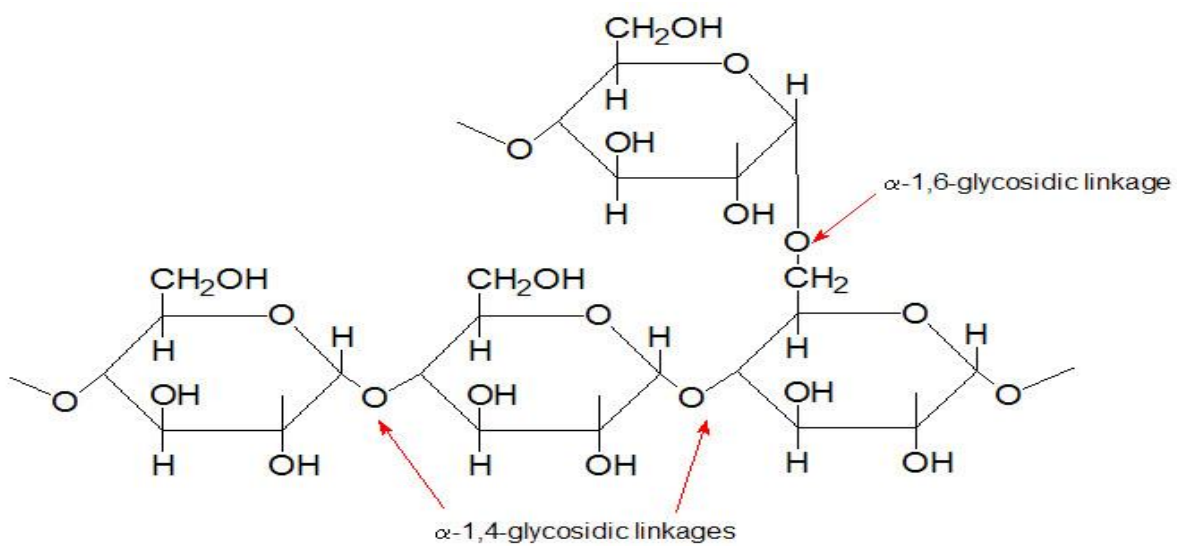
Starch is an important source for energy, and it has become a significant resource for bio- fuel production. ADP-glucose pyrophosphorylase is the enzyme that controls the synthesis of starch in plants, and glycogen in bacteria. This regulation is mainly driven by allosteric activators (Fru6P, FBP, and 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  $\alpha$ -D-glucose. These polymers are linked by  $\alpha$ -1,4-linkages and branched by  $\alpha$ -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  $\alpha$ -1,4-linkages with higher numbers of  $\alpha$ -1,6-linkage than starch. Starch is composed of amylose and amylopectin. Amylose is made up of fewer  $\alpha$ -1,4-linkages while the amylopectin is a more complex structure with higher numbers of  $\alpha$ -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  $\alpha$ -1,4-glycosidic linkages, and the branch with the  $\alpha$ -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, archaeobacteria and photosynthetic bacteria [11]. Glycogen is the polymer of glucose with  $\alpha$ -1,4-linkages, and few branches with  $\alpha$ -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

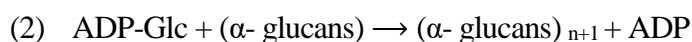
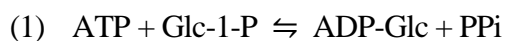
stored source of energy and carbon surplus [13]. Bacteria usually accumulate glycogen as a result of an excess source of carbon present at the stationary growth phase) [11]. Some exceptional bacterial species like *Rhodospseudomonas capsulata* and *Streptococcus mitis* synthesize glycogen during exponential 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: amylose and amylopectin. Amylose is mainly linear  $\alpha$ -1,4-polyglucans, while amylopectin is also branched with  $\alpha$ -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.

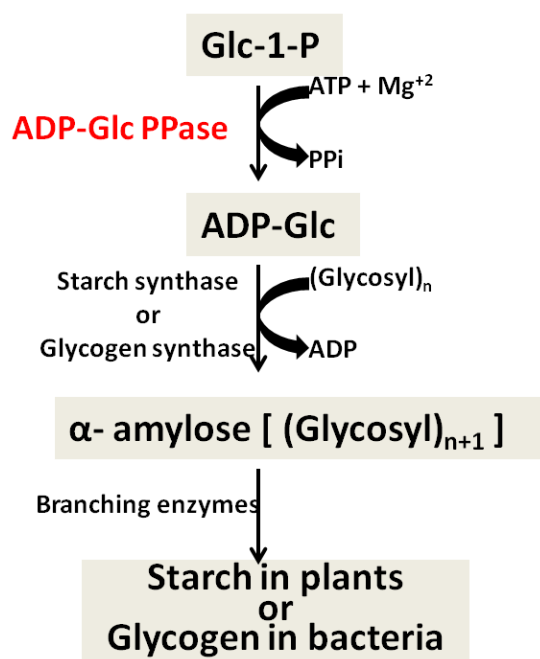


ADP-Glc PPase catalyzes the first committed reaction in starch and glycogen synthetic pathways. This enzyme catalyzes the conversion of ATP and glucose-1-phosphate (Glc-1-P) to ADP-glucose (ADP-Glc) and inorganic pyrophosphate (PPi) in the presence of  $Mg^{2+}$ . This reaction was first described in 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  $\alpha$ -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  $\alpha$ -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  $\alpha$  1,6 linkage.

ADP-Glc PPase is 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 *glgA* 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  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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 (*glgB*) 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  $\alpha$ -1,6-linkages. On the other hand, the amylopectin is formed by longer branches of 20-24 glucose units, with only 5% of  $\alpha$ -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 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 ( $\alpha_4$ ), or a heterotetramer ( $\alpha_2\beta_2$ ). ADP-Glc PPases in bacteria is a homotetramer ( $\alpha_4$ ) 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  $\alpha$  or S (small) subunits and two  $\beta$  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
<b>Accumulating Glycogen Prokaryotes</b>					
<i>Escherichia coli</i> <i>Salmonella enterica serovar</i> <i>Enterobacter aerogenes</i> <i>Typhimurium</i>	The Embden-Meyerhof pathway (glycolysis)	I	FBP, Pyruvate	AMP	Homotetramer ( $\alpha_4$ )
<i>Aeromonas formicans</i> <i>Micrococcus luteus</i> <i>Mycobacterium smegmatis</i>		II	Fru6P	AMP, ADP	
<i>Serratia Marcescens</i> <i>Enterobacter hafniae</i> <i>Clostridium pasteurianum</i>		III	none	AMP	
<i>Agrobacterium tumefaciens</i> <i>Arthrobacter viscosus</i>	Entner-Doudoroff Pathway	IV	Pyruvate, Fru6P	AMP, ADP, Pi	
<i>Chromatium vinosum</i> <i>Rhodobacter capsulata</i> <i>Rhodomicrobium vannielii</i>	Entner-Doudoroff Pathway	IV	Pyruvate, Fru6P	AMP, ADP	Homotetramer ( $\alpha_4$ )
<i>Rhodobacter gelatinosa</i> <i>Rhodobacter globiformis</i> <i>Rhodobacter sphaeroides</i>	Entner-Doudoroff Pathway & Glycolysis	V	Pyruvate, Fru6P, FBP	AMP, Pi	Homotetramer ( $\alpha_4$ )
<i>Rhodocyclus purpureus</i> <i>Rhodospirillum rubrum</i> <i>Rhodospirillum tenue</i>	TCA cycle Reductive carboxylate cycle	VI	Pyruvate	none	
<i>Bacillus subtilis</i> <i>Geobacillus stearothermophilus</i>	TCA cycle during sporulation	VII	None	3-PGA	Heterotetramer ( $\alpha_2 \beta_2$ )
<i>Streptococcus mutans</i>	TCA cycle	VIII	FBP	Pi	Heterotetramer ( $\alpha_2 \beta_2$ )

Table 1 continued					
Organism	Carbon metabolism	Class	Activators	Inhibitors	Quaternary structure
<i>Ruminococcus albus</i>	TCA cycle	IX	Pyruvate PEP	FBP	Heterotetramer ( $\alpha_2 \beta_2$ )
<i>Streptomyces coelicolor</i>	TCA cycle	X	Glc6P Man6P	NADPH Pi	Heterotetramer ( $\alpha_2 \beta_2$ )
<i>Rhodococcus jostii</i>	TCA cycle	XI	Glc6P Man6P Fru6P PEP	NADPH 6P-gluconate	Heterotetramer ( $\alpha_2 \beta_2$ )
<b>Cyanobacteria</b> <i>Synechococcus</i> PCC 6301 <i>Synechocystis</i> PCC 6803 <i>Anabaena</i> PCC 7120	Oxygenic photosynthesis Fixing CO <sub>2</sub> through Calvin cycle	XII	3-PGA	Pi	Homotetramer ( $\alpha_4$ )
<b>Accumulating starch Eukaryotes</b>					
<b>Green algae</b> <i>Chlorella fusca</i> <i>Chlorella vulgaris</i> <i>Chlamydomonas reinhardtii</i>	Fixing CO <sub>2</sub> through Calvin cycle	XII	3-PGA	Pi	Heterotetramer ( $\alpha_2 \beta_2$ )
<b>Higher Plants Photosynthetic tissue</b> Leaves of spinach, wheat <i>Arabidopsis</i> , maize, rice	Fixing CO <sub>2</sub> through Calvin cycle	XII	3-PGA	Pi	
<b>Non-photosynthetic tissues</b> 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<sub>2</sub> 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<sub>2</sub>. 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.

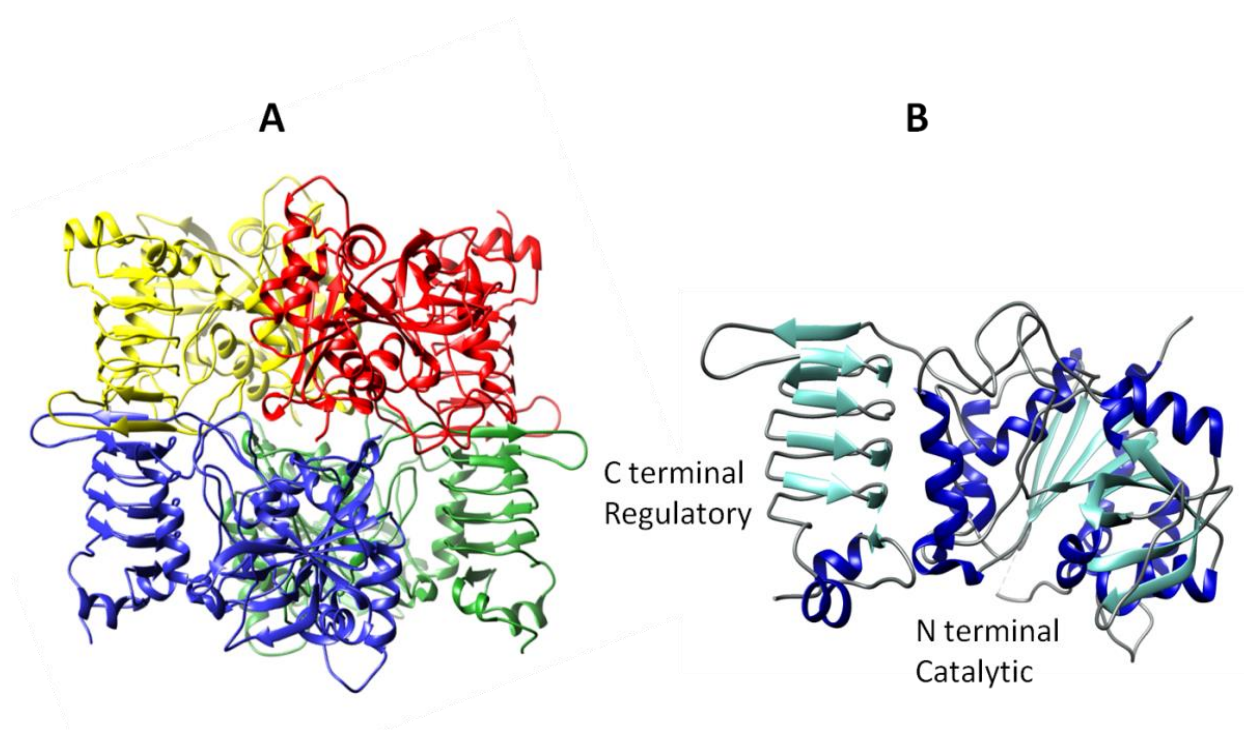
### ***Agrobacterium tumefaciens* used for generating transgenic plants**

*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 ( $\alpha_4$ ) of the ADP-Glc PPase while a single subunit is shown in (B). In the single subunit, the  $\beta$ -sheets are colored in cyan and the  $\alpha$ -helices 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 60 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,

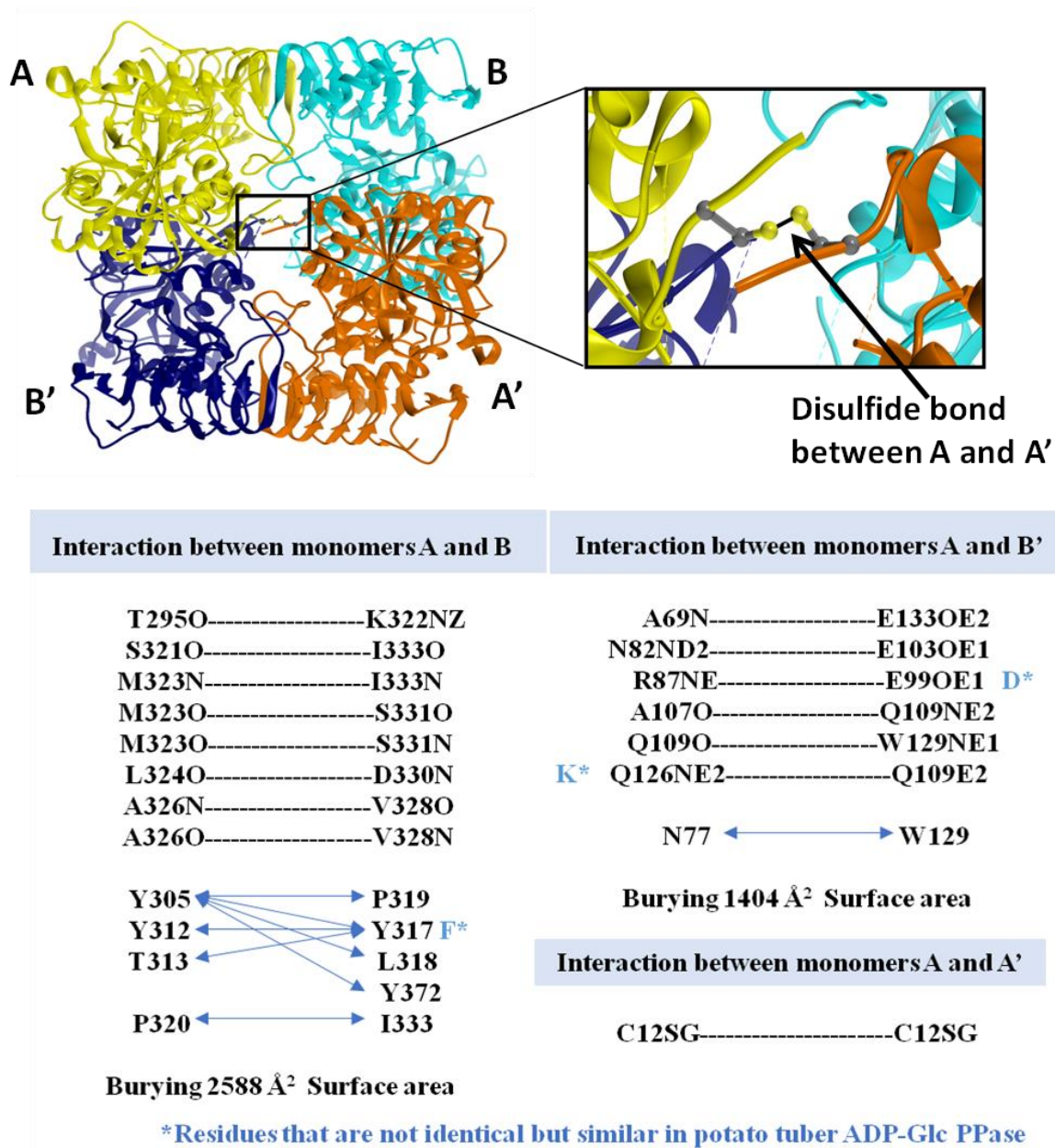


both the mutant enzymes showed decreased affinity for the 3-PGA and increased binding of Fru6P. This suggested that a single mutation can switch the activator selectivity.

**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 of the 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.

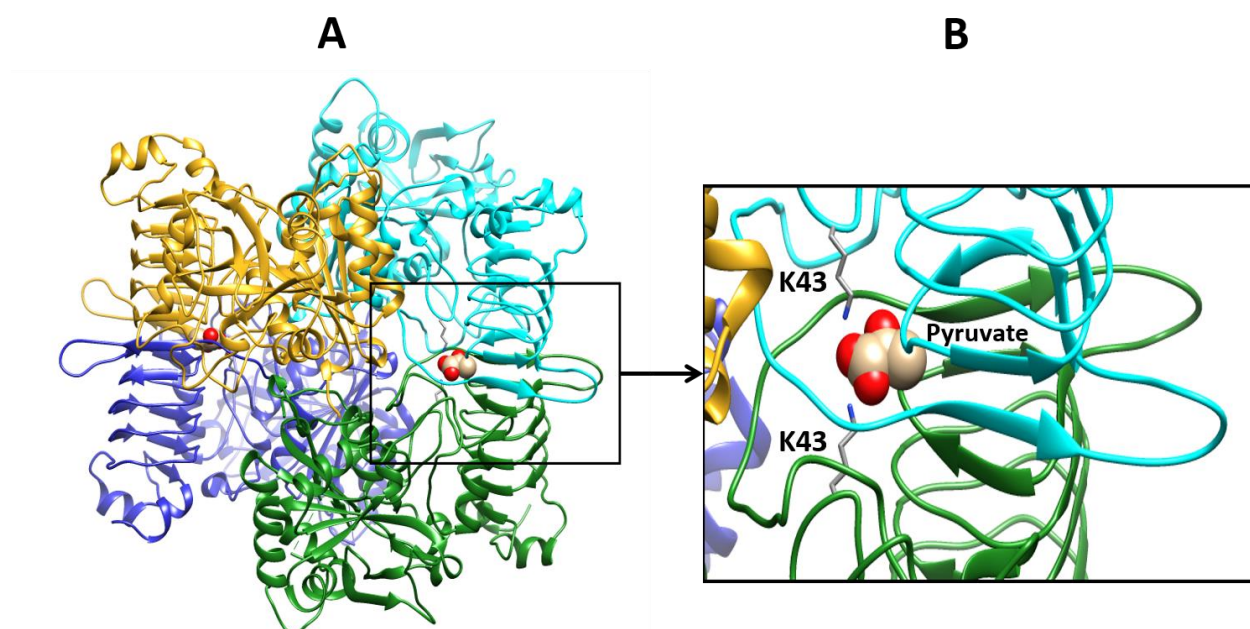
## Introduction

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-phosphate (Glc-1-P) to ADP-glucose (ADP-Glc) and inorganic pyrophosphate (PP<sub>i</sub>) in the presence of Mg<sup>2+</sup>; The reaction catalyzed by ADP-Glc PPase is shown below.



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 ( $\alpha_4$ ), 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).



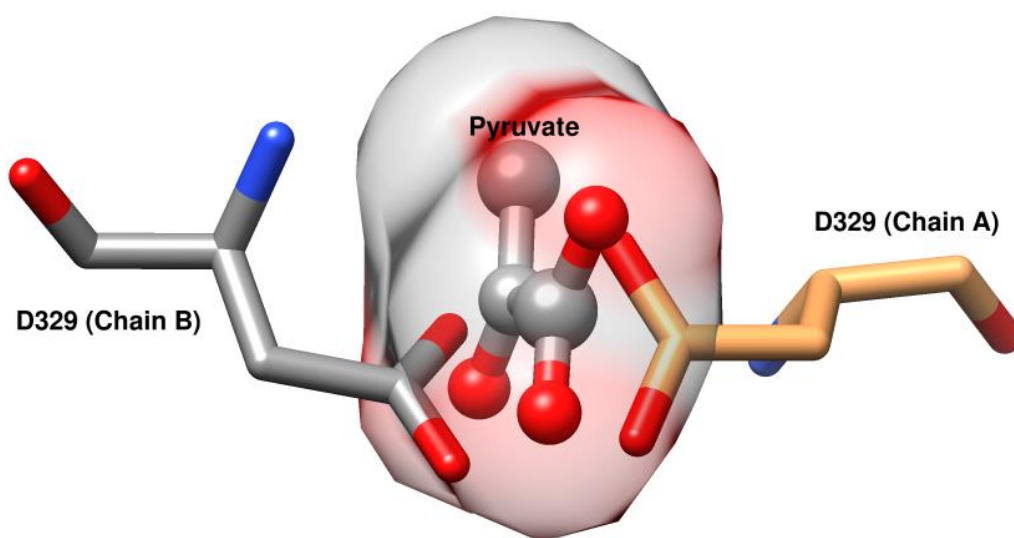
As we know, this enzyme is a homotetramer, but we can discuss it as a dimer of dimers for convenience. In one dimer, two K43 residues from two different subunits are making close contact with the pyruvate. K43 of the pyruvate bound from one subunit is making a hydrogen bond with the carboxylate oxygen, 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  $\pi$ - $\pi$  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 <sup>a</sup>	RMSD (Å) <sup>b</sup>
K43A vs. WT	0.47
WT vs. WT_Ethyl Pyruvate <sup>c</sup>	0.53
WT_Ethyl Pyruvate <sup>c</sup> vs. K43A	0.49

<sup>a</sup>The crystal structures of wild-type and K43A ADP-Glc PPase was compared using PDB Viewer.

<sup>b</sup>The RMSD values for the backbone were calculated using the PDB Viewer.

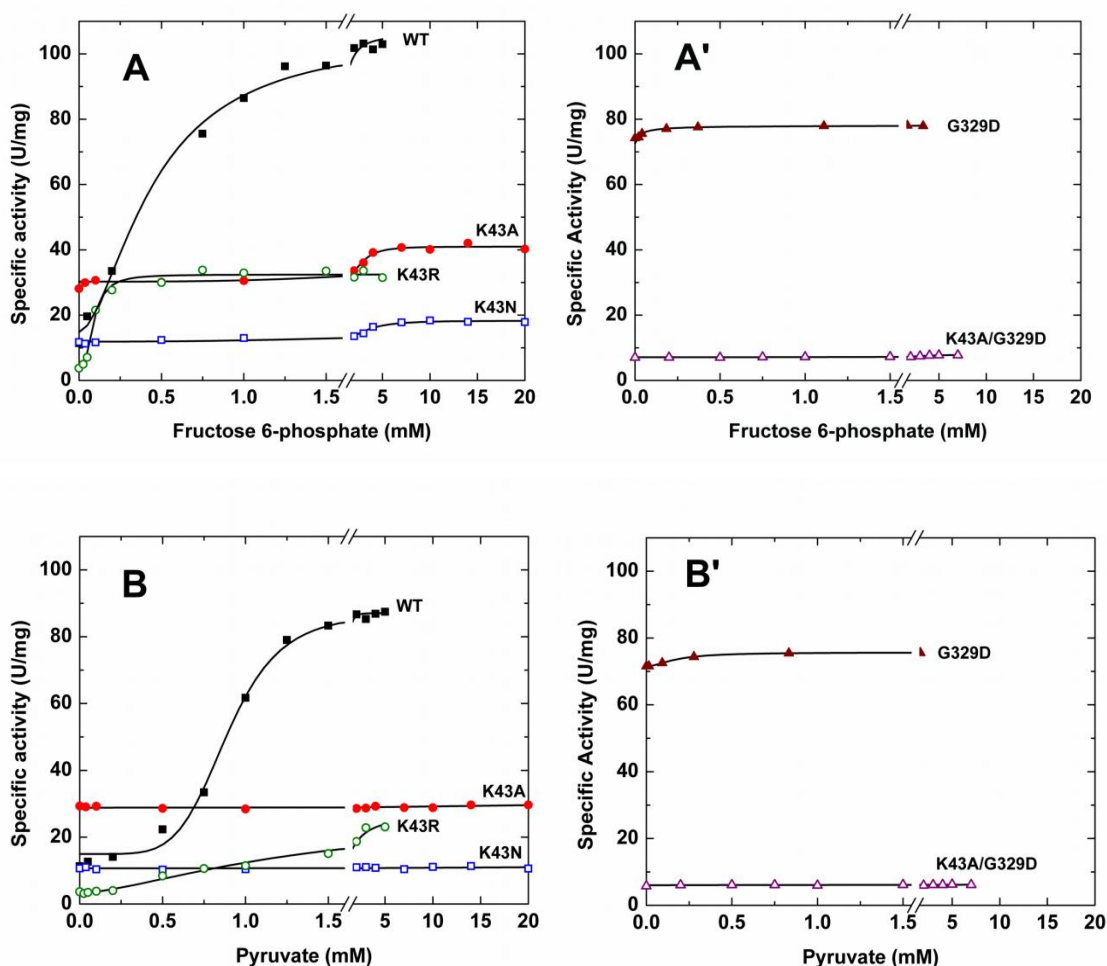
<sup>c</sup> WT\_Ethyl Pyruvate represents the wild-type crystal structure of ADP-Glc PPase with Ethyl Pyruvate-bound to it.

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

Enzyme <sup>a</sup>	Control			+ 1.5mM Pyruvate			+ 1.5mM Fru6P		
	$S_{0.5}$ <sup>b</sup> (ATP)	$n_H$ <sup>b</sup>	$V_{max}$	$S_{0.5}$ <sup>b</sup> (ATP)	$n_H$ <sup>b</sup>	$V_{max}$	$S_{0.5}$ <sup>b</sup> (ATP)	$n_H$ <sup>b</sup>	$V_{max}$
	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 <sup>b</sup>	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

<sup>a</sup> Assays were performed as described in substrate saturation assay, as stated under Materials and Methods.

<sup>b</sup> The  $S_{0.5}$  and  $n_H$  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 the G329D, 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 <sup>a</sup>	Fru6P				Pyruvate			
	$A_{0.5}$	$n_H$	$V_{max}$	Activation <sup>b</sup> ( $V_{max}/V_0$ )	$A_{0.5}$	$n_H$	$V_{max}$	Activation <sup>b</sup> ( $V_{max}/V_0$ )
	mM		U/mg	(-fold)	mM		U/mg	(-fold)
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 <sup>c</sup>	2.43 ± 0.23	3.5 ± 1.1	40.45 ± 0.51	1.34	N/A <sup>c</sup>	N/A <sup>c</sup>	28.84 ± 1.21	0.99
K43A/G329D	3.11 ± 0.70	2.3 ± 0.9	7.90 ± 0.18	1.22	N/A <sup>c</sup>	N/A <sup>c</sup>	6.37 ± 1.06	1.10
K43N	3.95 ± 1.52	2.6 ± 1.0	21.28 ± 4.64	1.88	N/A <sup>c</sup>	N/A <sup>c</sup>	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

<sup>a</sup> Assays were performed as described in activator saturation assay, as stated under Materials and Methods.

<sup>b</sup> Activation fold is calculated by dividing the maximum velocity ( $V_{max}$ ) by the velocity in the absence of activator ( $V_0$ ) ( $V_{max}/V_0$ ).

<sup>c</sup> 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.

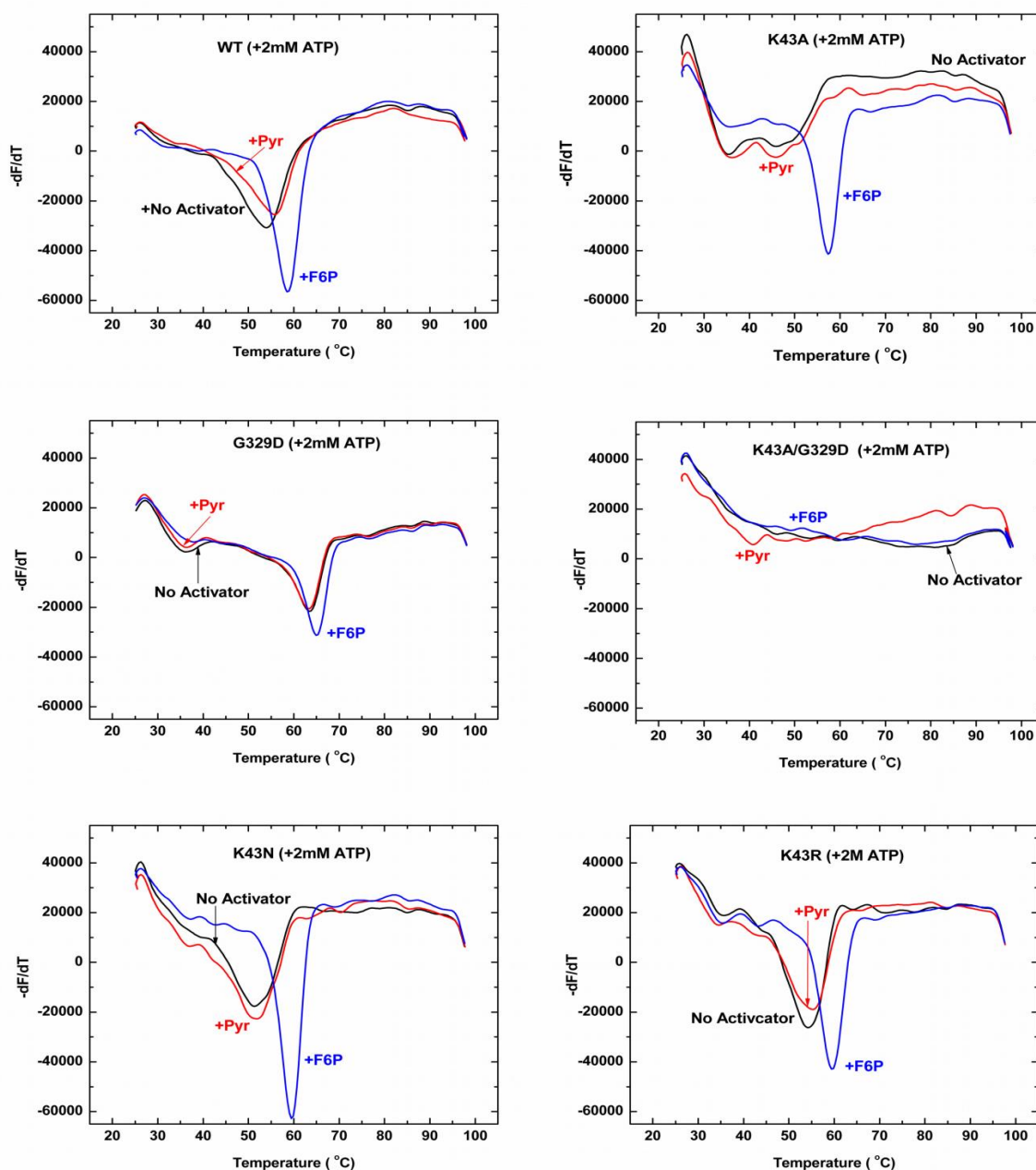
### **Binding analysis by Thermal shift assay**

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

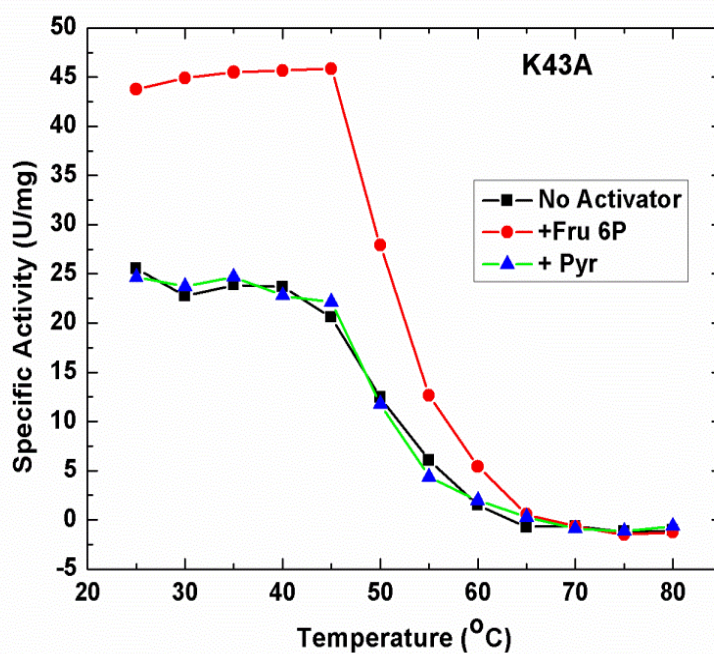
**Table 5. Kinetic parameters for the activators saturation curve of ADP-Glc PPase wild type, and mutants.**

Enzyme <sup>a</sup>	Substrate (ATP) (0mM)			Substrate (ATP) (2mM)		
	$T_m$ (°C) <sup>b</sup>			$T_m$ (°C) <sup>b</sup>		
	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 <sup>c</sup>	N/A <sup>c</sup>	N/A <sup>c</sup>	N/A <sup>c</sup>	N/A <sup>c</sup>	N/A <sup>c</sup>
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

<sup>a</sup> Thermal shift assays were performed for wild-type and mutant ADP-Glc PPases as described in Materials and Methods.

<sup>b</sup> The melting temperature were determined using the Step-One software.

<sup>c</sup> 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.

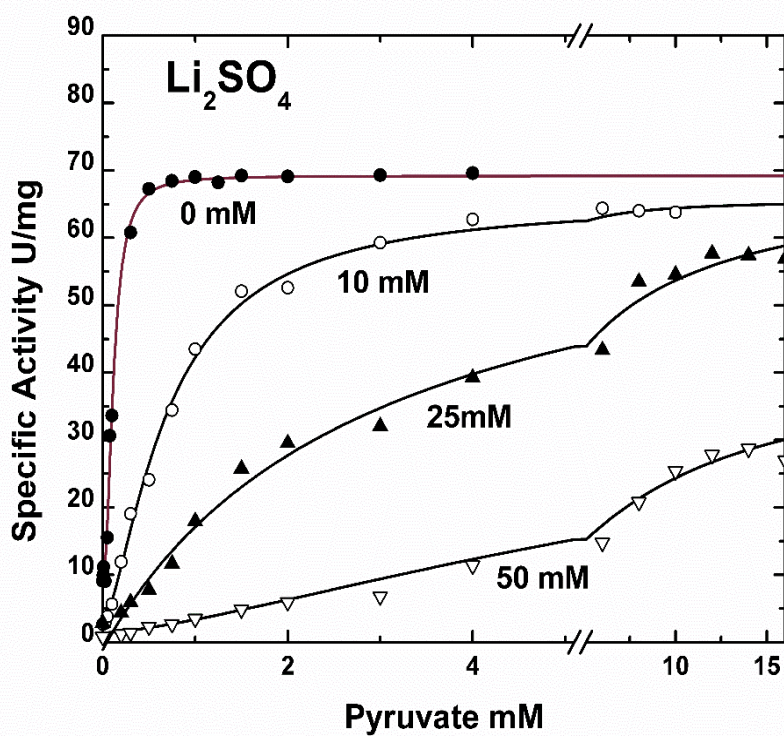
The G329D mutation mimics the presence of the pyruvate, by placing a carboxylate group at the position where the oxygens of pyruvate bind (Figure 5) [126]. The presence of aspartate at 329 positions seems to hinder the binding of pyruvate, since it is not expected to see any change in the thermal shift assay. In the presence of Fru6P, the melting temperature does increase 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.

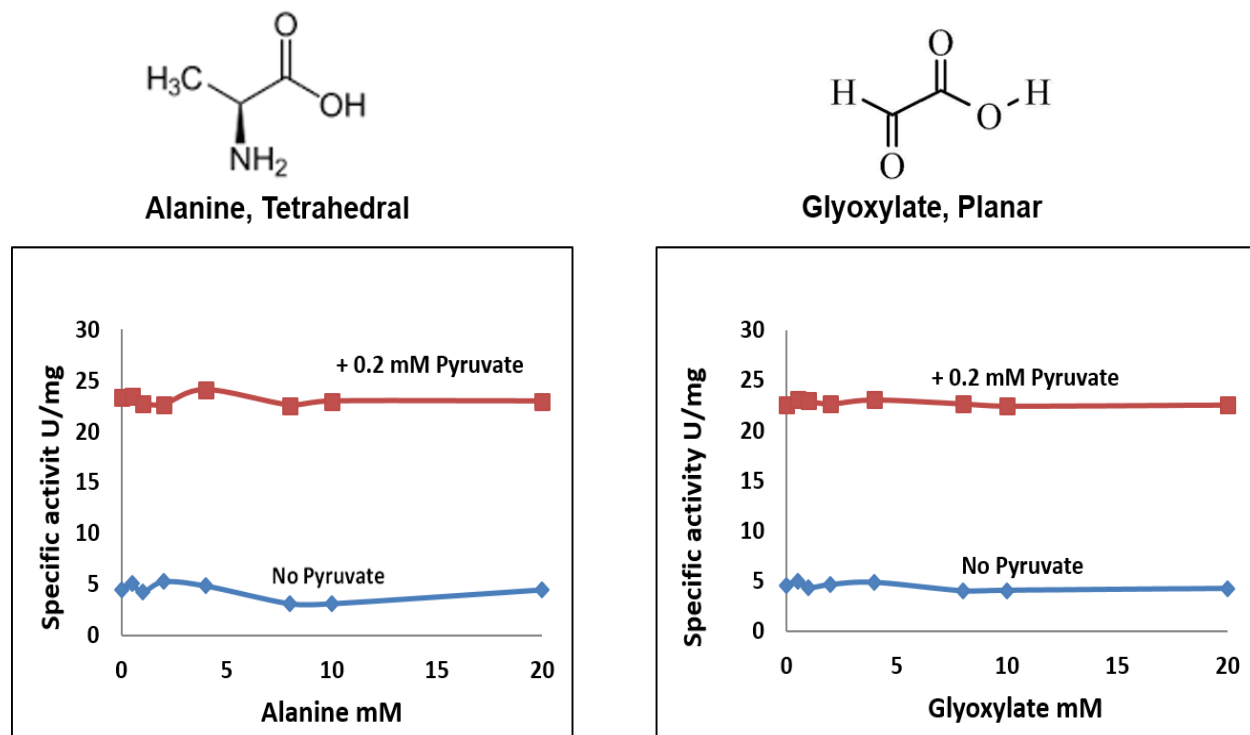
The double mutant is completely insensitive to pyruvate, while in the presence of Fru6P, it shows an increase in affinity for ATP without changing the  $V_{max}$  (Table 4). At a saturating concentration of ATP, the double mutant enzyme gets slightly activated by Fru6P with 1.2 activation fold, while 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  $\text{Li}_2\text{SO}_4$  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  $\text{Li}_2\text{SO}_4$  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**

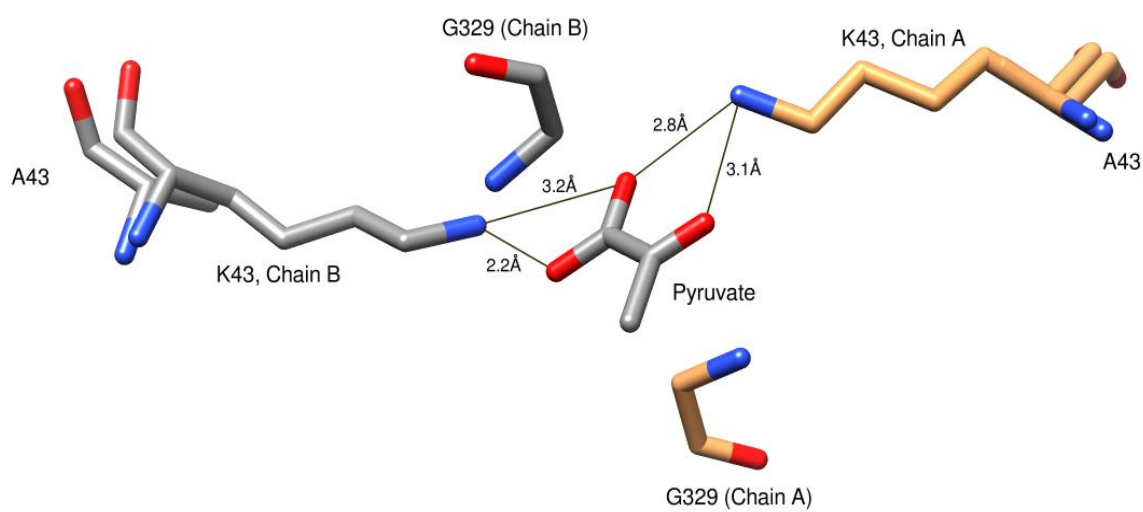
In that structure, the K43 residues from subunits A, and B form hydrogen bonds with the pyruvate oxygens [126]. The K43 is critical for binding the pyruvate, as was demonstrated with the K43 mutations (Figure 7). As we know, the mutant K43A was completely insensitive to pyruvate, and was partially activated 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.

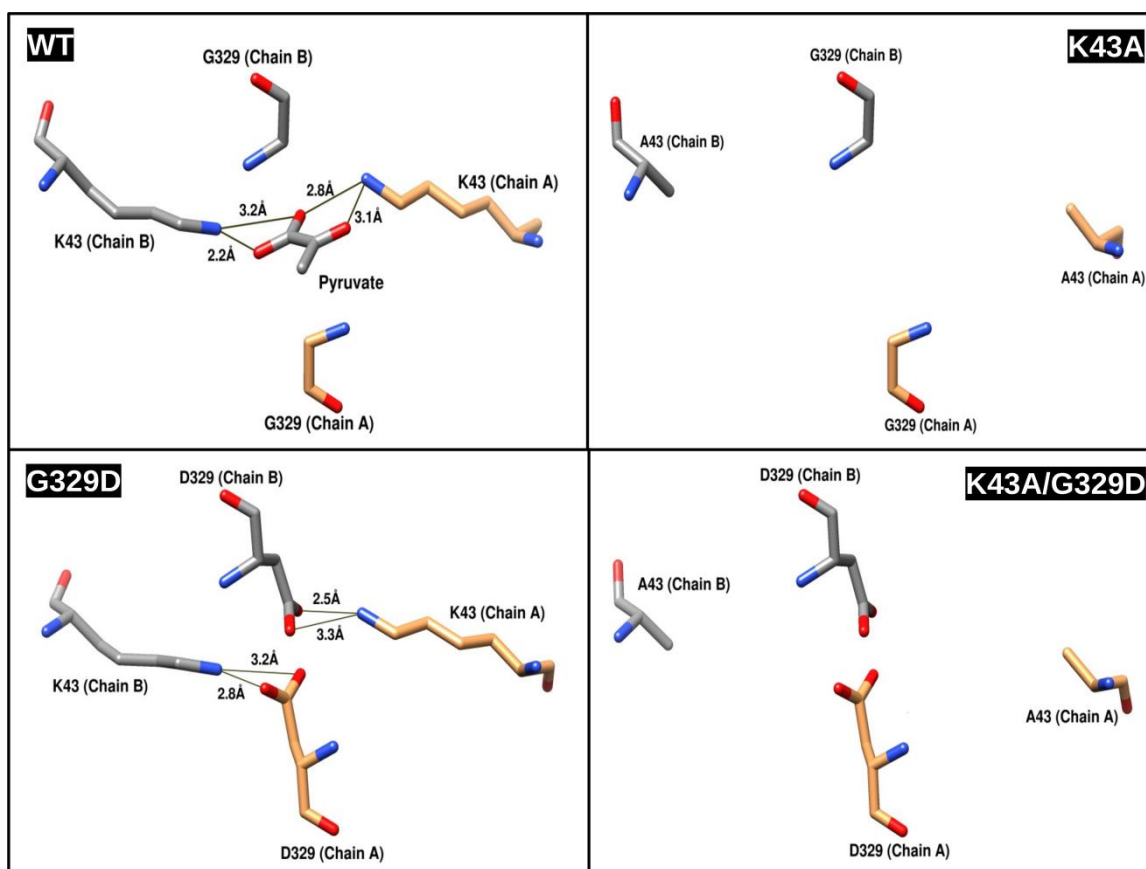
### **The Aspartate mutation of G329 makes the enzyme pre-activated**

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'-CGCGGTTTATTTTGGCGGCAACGCGCGC-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

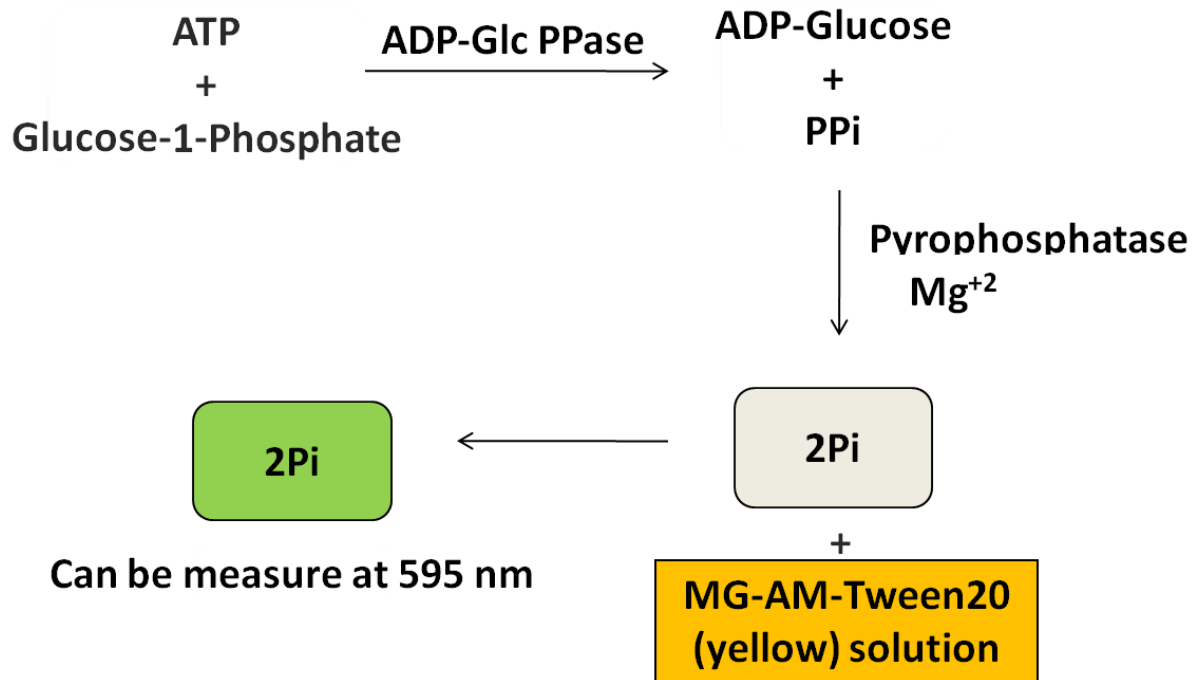
The wild-type and the mutants were expressed in *Escherichia coli* BL21 DE3 cells. The transformations were performed using the wild type, and each different mutant genes. The transformation mixes were plated on selection plates treated with X-gal for blue-white colony selection. White colonies were 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<sub>600</sub> 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<sup>2+</sup>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™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<sub>2</sub>, 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<sub>2</sub>, 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  $V = V_0 + [(V_{max} * S^n) / (K_m + S^n)]$  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.

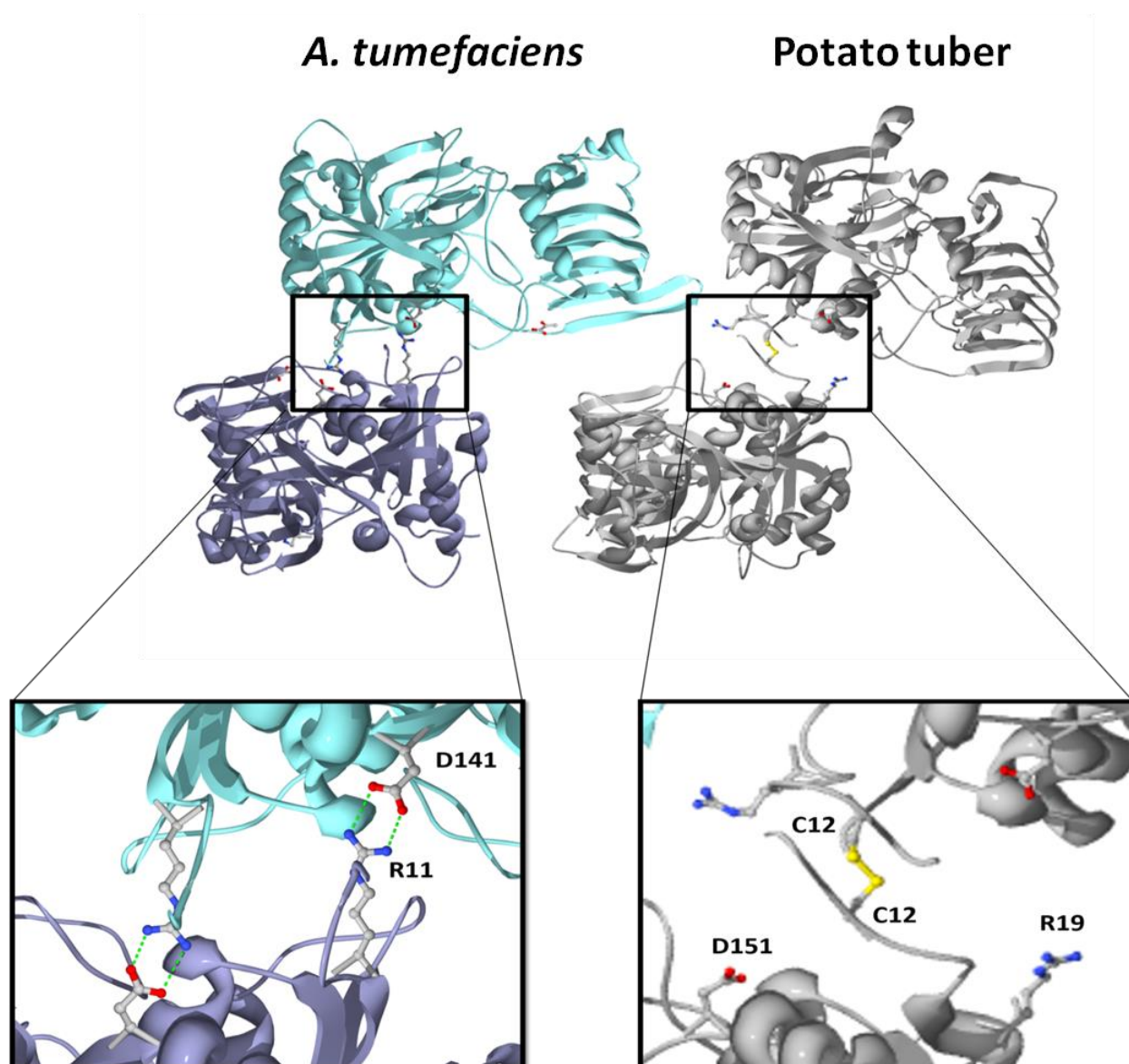
### Introduction

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 ( $\alpha_4$ ) of a ~50 kDa subunit, whereas in the plant this enzyme functions as a heterotetramer ( $\alpha_2 \beta_2$ ) [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  $\alpha$  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).

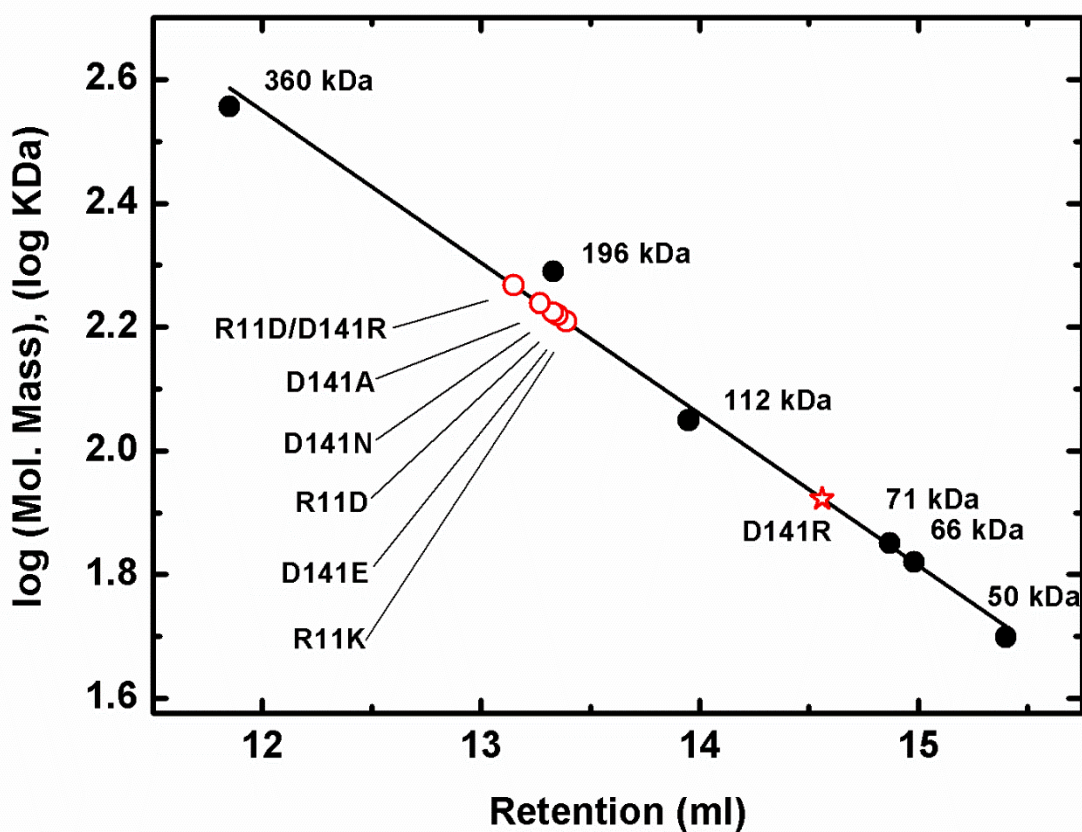
### **Gel filtration analysis**

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<sup>TM</sup>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 <sup>a</sup>	Control			+ 1.5mM Fru6P			+ 1.5mMPyruvate		
	$S_{0.5}$ (ATP)	$n_H$	$V_{max}$	$S_{0.5}$ (ATP)	$n_H$	$V_{max}$	$S_{0.5}$ (ATP)	$n_H$	$V_{max}$
	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 <sup>b</sup>	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

<sup>a</sup> Assays were performed as described in Assay A (substrate saturation), as stated under Materials and Methods.

<sup>b</sup> 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.**

Enzyme <sup>a</sup>	Control	+ 1.5mM Fru 6P	+1.5mM Pyruvate
	$k_{cat} / S_{0.5}^b$ (ATP) $s^{-1} mM^{-1}$	$k_{cat} / S_{0.5}^b$ (ATP) $s^{-1} mM^{-1}$	$k_{cat} / S_{0.5}^b$ (ATP) $s^{-1} mM^{-1}$
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 <sup>c</sup>	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

<sup>a</sup> Assays were performed as described in Assay A (substrate saturation), as stated under Materials and Methods.

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

<sup>c</sup> 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.**

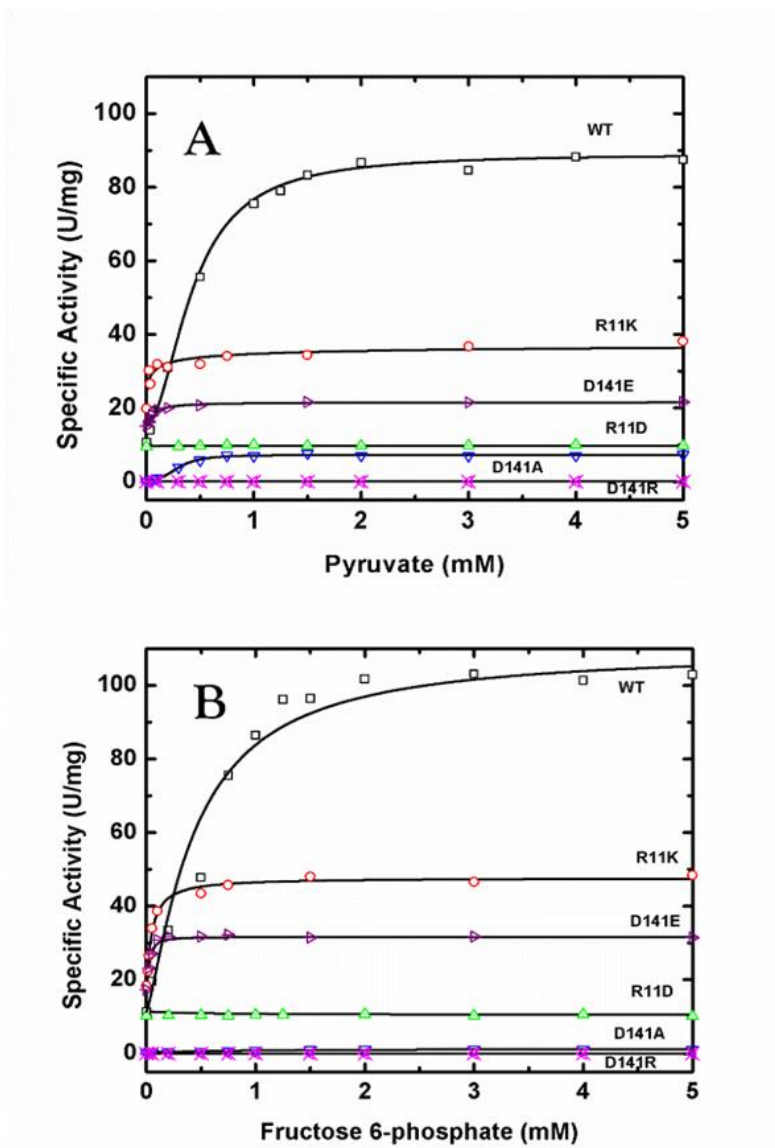
Enzyme <sup>a</sup>	Fru6P				Pyruvate			
	$A_{0.5}$	$n_H$	$V_{max}$	Activation <sup>b</sup> ( $V_{max}/V_0$ )	$A_{0.5}$	$n_H$	$V_{max}$	Activation <sup>b</sup> ( $V_{max}/V_0$ )
	mM		U/mg	(-fold)	mM		U/mg	(-fold)
WT	0.34 ± 0.07	2.1 ± 0.5	105.2 ± 3.9	9.01	0.45 ± 0.04	2.1 ± 0.4	87.4 ± 2.0	8.53
D141A	0.84 ± 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 <sup>d</sup>	N/A <sup>d</sup>	0.02 ± 0.01	1.00
D141R	N/A <sup>d</sup>	N/A <sup>d</sup>	0.02 ± 0.01	0.96	N/A <sup>d</sup>	N/A <sup>d</sup>	1.01 ± 0.17	0.95
R11A <sup>c</sup>	0.03 ± 0.008	1.3 ± 0.4	7.2 ± 0.2	1.8	N/A <sup>d</sup>	N/A <sup>d</sup>	5.0 ± 0.5	1
R11D	N/A <sup>d</sup>	N/A <sup>d</sup>	9.96 ± 0.16	0.98	N/A <sup>d</sup>	N/A <sup>d</sup>	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

<sup>a</sup> Assays were performed as described in Assay B (activator saturation), as stated under Materials and Methods.

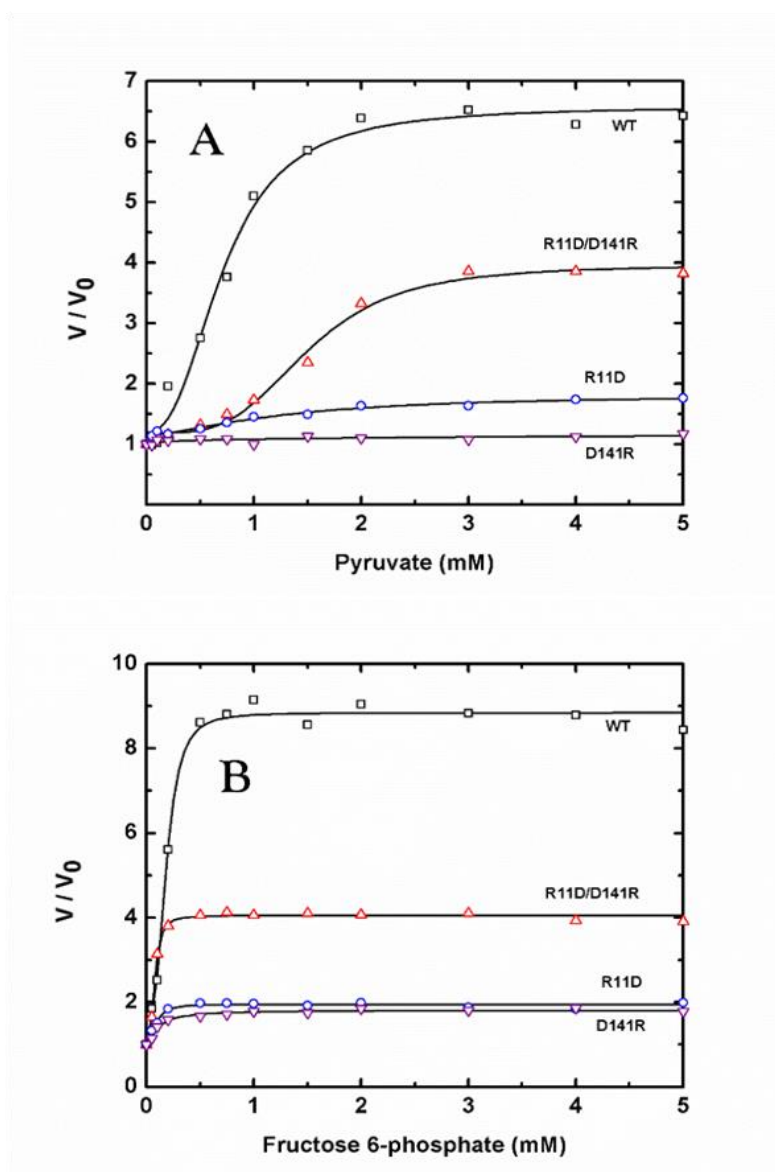
<sup>b</sup> Activation fold is calculated by dividing the maximum velocity ( $V_{max}$ ) by the velocity in the absence of activator ( $V_0$ ) ( $V_{max}/V_0$ ).

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

<sup>d</sup> 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), and R11D/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.

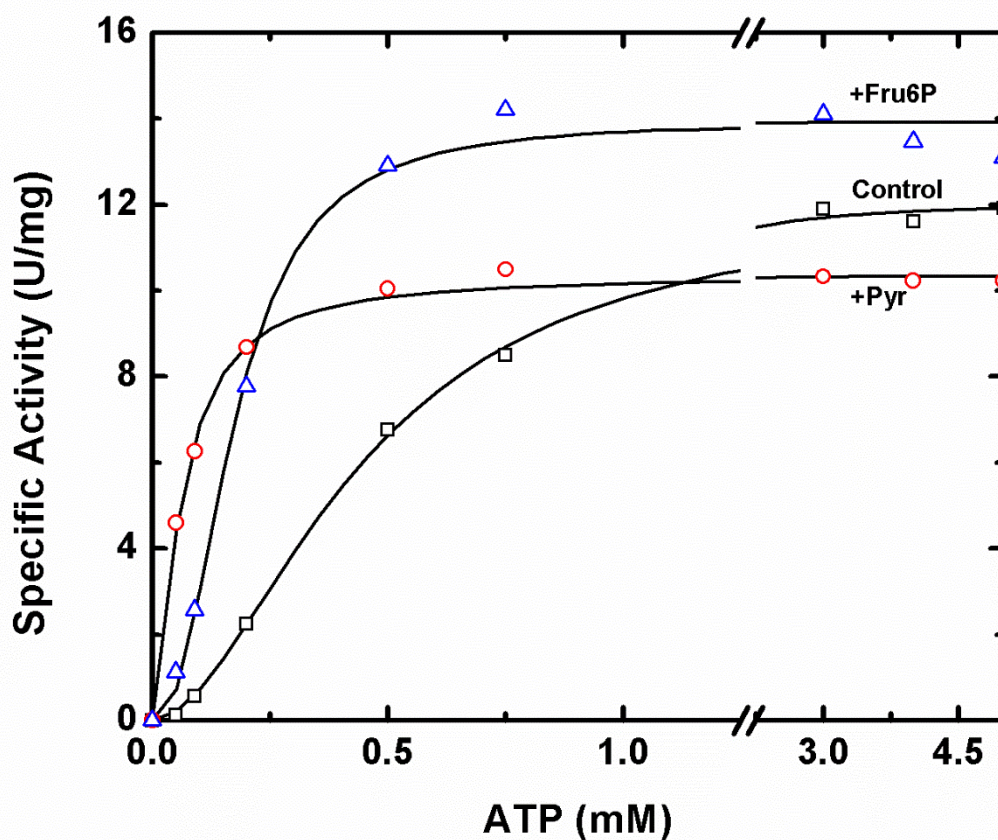
only 21%, and 24% activation, respectively, by Fru6P compared to wild-type. However, the activation by pyruvate for R11D and D141R were 18%, and 27%, respectively.

### **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).

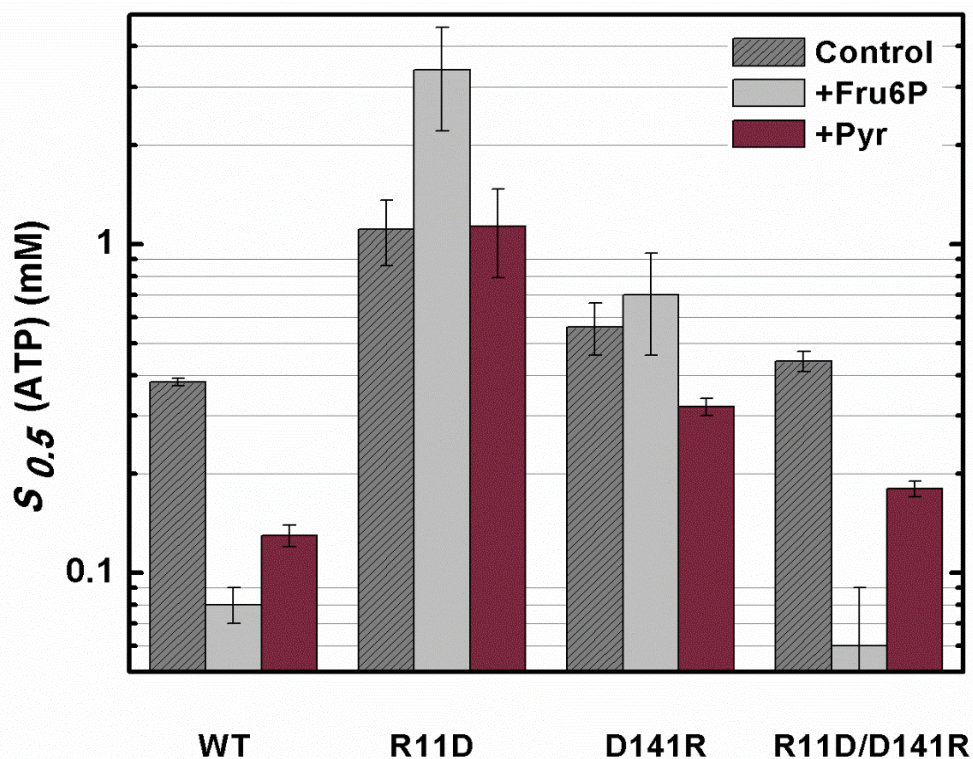
In the absence of any activator, the double mutant R11D/D141R had an identical  $V_{max}$  compared to wild-type, and the ATP  $S_{0.5}$  increased by only ~2-fold, whereas, in presence of activators (Fru6P, Pyr), the ATP  $S_{0.5}$  was similar compared to wild-type (Figure 19, and Figure 20). The ability 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.



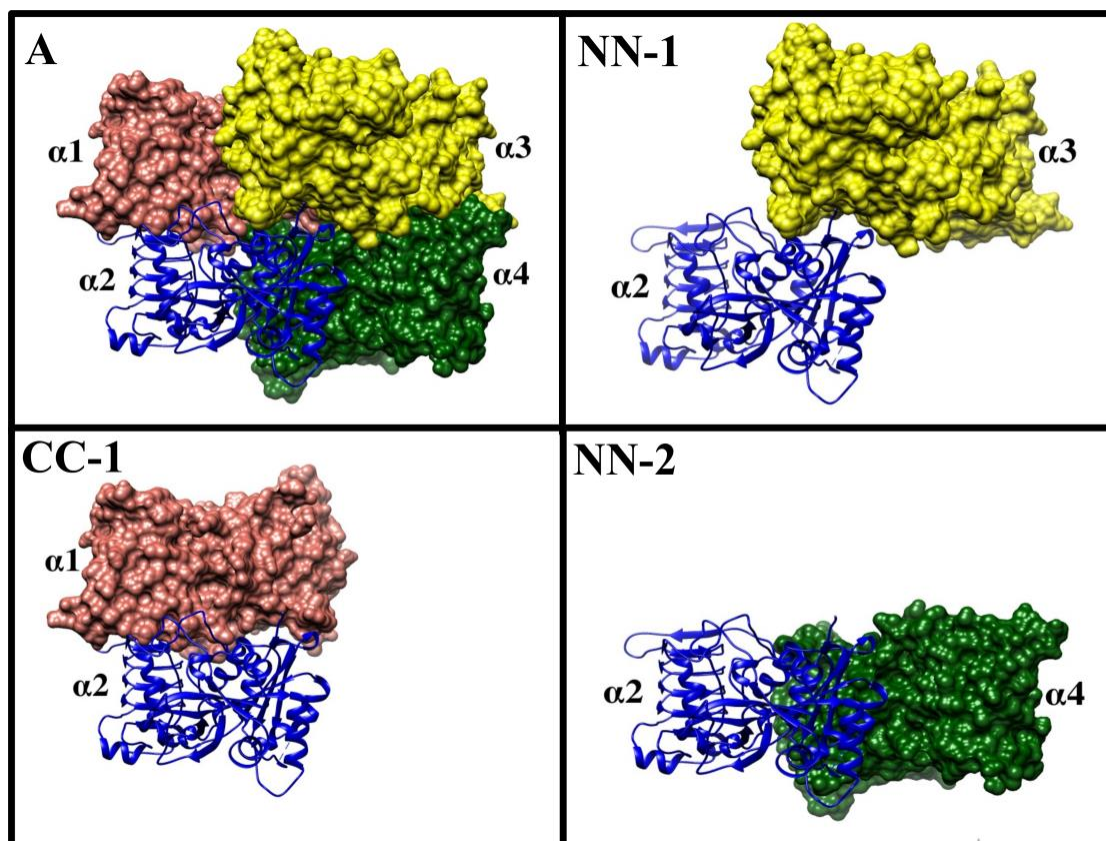
## Discussion

### 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 ( $\alpha 2$ -  $\alpha 3$  or  $\alpha 1$ -  $\alpha 4$ ), NN-2 ( $\alpha 2$ -  $\alpha 4$  or  $\alpha 1$ -  $\alpha 3$ ), and CC ( $\alpha 2$ -  $\alpha 1$  or  $\alpha 3$ -  $\alpha 4$ ) (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  $\alpha$  subunits, the  $\beta$  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  $\alpha$  subunits is an important part of the allosteric mechanism [70]. Overall the disulfide bond between  $\alpha$  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  $\alpha 2$ -  $\alpha 3$  of NN-1 group, R11 of  $\alpha 2$  makes a salt bridge with D141 of a neighboring subunit  $\alpha 3$ . In the same interface, D141 of the  $\alpha 2$  subunit makes another salt bridge with R11 of the  $\alpha 3$  subunit (Figure 15). Similarly, the other dimer  $\alpha 1$ - $\alpha 4$  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 the 11 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<sub>600</sub> 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

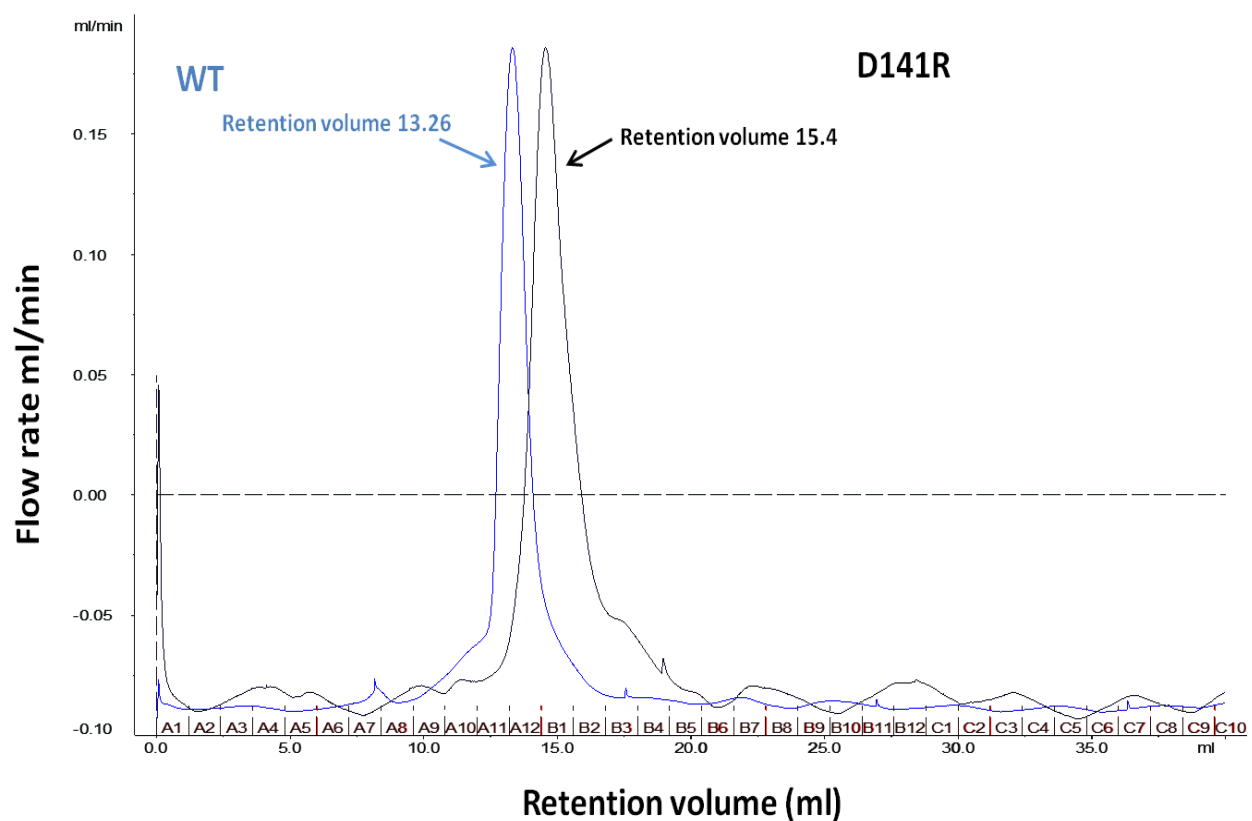
extracts were loaded onto a pre-equilibrated 5 ml His-Trap FF column (Ni<sup>2+</sup> Sepharose column) and eluted with a linear gradient 0-50% of buffer E (50 mM HEPES pH 7.5, 10% glycerol, 200 mM NaCl, and 750 mM imidazole). Active fractions were pooled, and concentrated after SDS-PAGE and enzyme assay, which was stored at -80 °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™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<sub>2</sub>, 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 P<sub>Pi</sub> 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<sub>2</sub>, 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 ( $n_H$ ) was calculated from fitting data in Origin™ 6.0 to the non-linear least square formula, by using the Hill equation  $V = V_0 + [(V_{max} * S^n) / (K_m + S^n)]$  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 IDFALSINALN SGIRRIGVAT QYKAHSLIRH LQRG--WDF 84  
 2 11 -----LA RQLPLKSVLAL ILAGGRGTRL KDLTNKRAKP AVHFGGKARI IDFALSNCIN SGIRRMGVIT QYQSHTLVQH IQRG--WSFF 91  
 3 5 DSQNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASASLNHR LSRAYASNMG 94  
 4 26 P-----RLERR RANPKDVAAY ILGGGEGTKL FPLTSRTATP AVFVGGCYRL IDIPMSNCIN SAINKIFVLT QYNSAPLNHR IARTY-FGNG 110  
 5 7 -----RFV SRLTRDTLAL IMAGGRGRL SNLTDWRTPK AVPFGGKFR L IDFPPLSNCIN SGIRRIEVL T QYKAHSLIQH IQRG--WGFL 88  
 6 7 -----RFV SRLTRNTLV L ILAGGRGSRL MDLTTWRAKP AVPFGGKFR I IDFTLSNCIN SGIRRIGVLT QYKAHSLIRH LRLG--WGS L 88  
 7 1 -----MIAGKSVLAF VMAGGEGSRL HPLTAERCKP SVFPNGRHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH TRRA--WVMS 78  
 8 37 SQVVETSQSV SVDMRQVASL ILSGGEGTRL HPLTLARCKP AINFGGKYRL IDVPISNSLH AGCKKVFLLT QFLSSSLHQH VFQTYMQ--G 124  
 9 7 QGYPPNYQAS HFYRDKVGI VLCGGEGKRL SPLTCWRCKP TVSPGGYK L IDVPISHAIA SGFSKIFVIG QYLTYTLQOH IVKTYFY--H 94  
 10 7 P-EASNFESS HFYRDKVGI ILCGGEGKRL SPLTNCRCCKP TVSPGGYK L IDIPISHAIS AGFSKIFVIG QYLTYTLQOH LFKTYFY--H 93  
 11 15 MREQTDMRFI SHLTRNTFAI ILAGGRGTRL KQLTDFRSKP AVPFAGKFR I LDFTLSCNVN SGIRKIGVAT QYKAHSLIRH IQRG--WSFL 102  
 12 5 ---LSSSRFV STLTKNTVAL ILAGGKGSRL RDLTNWTAKP AVPFGGKFR I IDFPPLSNCIN SGVRRIGVVT QYKAHTLIQH IQRG--WGFL 90  
 13 7 VQINDNPRFV STLTRNTLAL ILAGGRGTRL KNLTDWRAKP AVPFGGKFR I IDFTLSNCVN SGVRRIGVVT QYKAQSLIRH IQRG--WSFL 94  
 14 1 --MNNNPRFV STLTRNTLAL ILAGGRGTRL KNLTDWRAKP AVPFGGKFR I IDFTLSNCVN SGVRRIGVVT QYKAQSLIRH IQRG--WSFL 86  
 15 1 MKLPADLRSY SQLTRNSIAM ILAGGRGTRL RQLTDWRAKP AVPFGGKFR I IDFPPLSNCVN SGIRRIGVAT QYKAHSLIRH IQQG--WGFL 88  
 16 1 -----MKKVLAI ILGGGAGTRL YPLTKLRAKP AVPVAGKYRL IDIPVSNCLN SEIFKIYVLT QFNASASLNHR IARTYN--FS 75  
 17 5 -----RL NDLQRTTLAI VLAGGRGTRL GPLTNKRKVP AVHFGGKYRI IDFALSNCIN SGIRRIAVVT QYKAHSLIRH LQRG--WSFL 85  
 18 4 -----AH HRPVRRITISL VLAGGRGSRL QDLTENCAKP AVHFGGKFR I IDFVLSNCVN SGLHRIGVLT QYKSHSLIRH LQHG--WSFL 84  
 19 1 ---MCCWQSR GLLVKKVLA I ILGGGAGTRL YPLTKLRAKP AVPLAGKYRL IDIPVSNCLN SEIVKIYVLT QFNASASLNHR ISRAYN--FS 85  
 20 12 YRYYTEPTLV TELRKTAL VLAGGEGSRL KDLTAWRAKP AVFPGGKYRI IDFPPLSNCVN SGIRRIGVLT QYKSHSLIRH LQRA--WGLM 99  
 21 1 -----MAGGEGSRL QPLTADRCKP AVPFGRSRYRI VDFVLSNLTN SDIRSIYVLV QYRPSQLIEH VRKA--WTVT 67  
 22 13 AEYKSSSRPH SNISHETLAL ILAGGRGSRL HQLTDWRAKP AVFVGGKFR I IDFPPLSNCVN SGFRRIGVVT QYKAQSLIRH IQHG--WSFL 100  
 23 1 -----MNL RSEMKKTLCI LMAGGKERL YPLTKARAKP SVRFGGIYRI VDFTLSCNCLN SDIRRVYVLT QYRSVSLDRH IRLG--WNIF 81  
 24 1 -----MKNEMLAL ILAGGQGR L GKLTQSTAKP AVQFGGKYRI IDFALSNCIN SGINNVGIT QYQPLALNSH IGGNSWGLD 78  
 25 1 -----MRRRV DPGSNDVLSI ILGGGKGSRL YPLTKDRAKP AVPFGGKYRL VDIPI SNAN SDFKKIYILT QFNASASLNHR LSTYLFDT- 84  
 26 11 -----LA RQLPNKTVAL ILAGGRGSRL KDLTATRAKP AVHFGGKFR I IDFALSNCIN SGVRRIGVIT QYQSHTLVQH IQRG--WSFL 91  
 27 7 -----YI SNLTRDTYAL ILAGGRGSRL HELTTWRAKP ALYFGGKFR I IDFPPLSNCIN SGIRRVGVVT QYKSHSLIRH LVRG--WGHF 87  
 28 14 AECTTSTRFH NSVTYNTLAL ILAGGRGSRL HQLTDWRAKP AVPFGGKFR I IDFPPLSNCVN SGVRRIGVVT QYKSQSLIRH IQHG--WSFL 101  
 29 12 TPLTQNTINLH THRDVRSI ILGGGEGVRL FPLTDRCKP AVFVGGYRL IDFSISNSLN SGYQKIFILT QFLSSSLHQH IFRTYQF--D 99  
 30 7 -----SS FDLPRRSIAL VLAGGRGTRL KNLTDNRAKP AVYFGGKFR I VDFALSNCIN SGIRRIGVIT QYKSHSLRH LQRG--WAF L 87  
 31 6 -----L QOIINTVMV L VLAGGQGSRL KALTETRAKP VLEFGSHCRI IDFPPLSNCVN SGLNRIAVLT QYKSQSLIRH LMQH--WGLT 85  
 32 1 -----MRKREVVAM ILAGGQGSRL GVLTQNLAKP AVHFGGKFR I IDFTLSNCVN SGIVTVGVLT QYMPBSLHSH IGVGSPWDL D 79  
 33 1 -----MKQ-- ---KNRAIAI VLAGGKGR L YPLTMDRSKP AVPFAGKYRL VDIPI SNAN SGIRQIYILT QFNASASLNHR IANTYVFDN- 79  
 34 1 -----MAKVLAI ILGGGKGR L YPLTKERSKP AVPFGGNHRI VDIPI SNAN SGFRQIYILT QFNASASLNHR ISNAYNFD- 76  
 35 6 QNHLGGKTS YRYRDRVGI VLCGGEGKRL SPLTDRCKP TVSPGGYK L IDVPISHAIS SGFSKIFVIG QYLTYTLQOH LFKTYFY--H 93  
 36 1 MDLRLKTLTL IMAGGRGERL FPLTDRCKP AVTFGGKYRI IDFTLSNCIN SGIRIYVLT QYGSFSLDH LRMA--WEV V 78  
 37 1 MSSTKNVLA I VMAGGEGSRL HPLTAERCKP AVPFNGKHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH IRQS--WTMT 78  
 38 1 MIAGKSVLAF VMAGGEGSRL HPLTAERCKP SVFPNGRHRI VDFVLSNLVN SEIYSIYLLV QYKSQSLIEH IRRS--WVLT 78  
 39 7 -----RFV SRLTRDTLAL ILAGGRGTRL HELTQWRAKP AVPFGGKFR I IDFPPLSNCIN SGVRRIGVLT QYKAHSLIRH IQRG--WSSL 88  
 40 7 -----RFV SRLTRDTLAL IMAGGRGRL SSDLTDRTPK AIPFGGKFR L IDFPPLSNCIN SGIRRVGILT QYKAHSLIQH VQRG--WGFL 88  
 41 7 -----QG LDLPKRAIAL VLAGGRGSRL MNLTDRAKP AVYFGGKFR I IDFALSNCIN SGIRRIGVIT QYKSHSLIRH LQRG--WAF L 87  
 42 7 -----YI SNLTKDTYAL VLAGGRGSRL FELTDRCKP AVYFGGKFR I IDFALSNCIN SGINRIGVAT QYKSHSLIRH INRG--WGNF 87  
 43 1 -----MAGILAM ILAGGEGSRL FPLTQRTKP AVPFGGNYRL VDFALNNEVN ADLMKIYVLT QYKSQSLNH LRKA--WRLS 75  
 44 1 -----MAKVLSM ILAGGEGSRL YPLTQSRTPK SVPFGGSYRL VDFALNNEVN SDDLRIYVLT QYKSQSLNH LRKA--WELN 75  
 45 4 -----IE PPFARHAMAY VLAGGRGSRL MELTDRRAKP AVYFGGKSRI IDFALSINALN SGIRRIAVAT QYKAHSLIRH LQHG--WNFF 84  
 46 8 -----RFV SRLTRDTLAV ILAGGRGSRL HQLTDWRAKP AVHFGGKFR I IDFPPLSNCVN SGIRRISVLT QYKSHSLDRH IQRG--WGFL 89  
 47 1 -----MQNQTLTF LLAGGVGSRL HPLTSSRSKP SVPFGGKYRI IDFTLANCLH SGLRRILVLT QYKSHSLNH LRDG--WSIF 76  
 48 5 -----SL NDLQHTTLAI VLAGGRGTRL GPLTNKRKVP AVHFGGKYRI IDFALSNCIN SGIRRIAVVT QYKAHSLIRH VQRG--WGFL 85  
 49 1 -----M ILAGGEGRL GPLTHDRAKP AVPFGGYRI IDIVLSNEVN SGLHRIKILT QYKASLDEH IARA--WRLS 69  
 50 7 -----RR RILTRRTLAL VLAGGRGSRL RDLTNVRAKP AVHFGGKFR I IDFALSNCMN SGLRRIGVIT QYKSHSLRH LQRG--WSFL 87  
 51 1 -----MKRVLAI ILGGGKGSRL YPLTKMRAKP AVPLAGKYRL IDIPISNCIN SGIEKMYVLT QFNASASLNHR IGRTYN--LN 75  
 52 4 -----PP LRLTSQAMAF VLAGGRGSRL KELTDRRAKP AVYFGGKARI IDFALSNCIN SGIRKMAIAT QYKAHSLIRH IQRG--WNFF 84  
 53 8 -----FV SRLTRDTLAL ILAGGRGSRL KHLTAWRAKP AVPIGGKFR I IDFPPLSNCVN SGIRRIGVLT QYKAHSLIRH IQQG--WGFM 88  
 54 1 -----MRK KRYGGRIIAF VMAGGEGKRL HPLTAERCKP AVPFGAR YRL VDFVLSNLIN SEIRSIYLLV QYKQALIEH IRRRA--WVIS 81  
 55 4 -----PN QRLSSQAMAF VLAGGRGSRL KELTDRRAKP AVYFGGKTRI IDFALSINALN SGIRKMAIAT QYKAHSLIRH MQRG--WNFF 84  
 56 17 -----QP HMLVRSIAL VLAGGRGSRL KQLTDKRAKP AVYFGGKFR I IDFALSNCVN SGIRRIGVIT QYKSHSLRH LQRG--WSFL 97  
 57 4 -----KT FQLPRKAIAL VLAGGRGSRL HELTDRRAKP AVHFGGKFR I IDFALSNCVN SGIRRVGVVT QYKSHSLRH LQRG--WNFL 84  
 58 1 MEGDYAQH IH VRQPRVLAV VLAGGEGTRL YPLSAHRAKP AVPFGGSYRI IDFVLSNEVN SGFHRIKVL T QYKSDSLVNH ISRG--WRLS 88  
 59 4 ---LSTREI SALTKNTVAL ILAGGKGSRL KDLTNWRAKP AVPFGGKFR I IDFPPLSNCMN SGVRRIGVVT QYKSHSLMQH IQRG--WGFL 89  
 60 4 ---MRSPRFI SALTKNTVAL VLAGGRGSRL HNLTDWNAKP AVQFGGKFR I IDFPPLSNCMN SGIRRIGVVT QYKAHTLIEH IQQG--WGFL 89  
 61 1 -----MRAFNTVAL ILGGGRTLR L YPLTKARSKP AVSILGGQYRM IDIPVSNCLN SGFRNIYVIT QFNASASLNHR IYNAYRFDN- 78  
 62 1 -----MGGILSM ILAGGEGTRL PFPFSLRAKP AVPFGGNYRI IDFVLSNFIN SDDLQIFVLT QYKSHSLMKH LSQA--WRIS 75  
 63 1 -----MEIT AMIMKDVGLG IMGGGRTLR L YPLTKARSKP AVPLAGKYRL IDVPISNCLH SGIDKISILT QFNASASLNHR IFQTYR--R 81  
 64 16 -----QA HQLVRRITIAL VLAGGRGSRL KQLTDKRAKP AVYFGGKFR I IDFALSNCIN SGIRRMVAVT QYKSHSLRH LQRG--WSFL 96  
 65 1 -----METLAM VLAGGKGSRL DILSADRRAKP SVPFAGKFR I IDFALSNCVN SGIYDVGILT QYLPRSLNHR IGIGKPWDL D 76  
 66 15 ASSARPS--V DQLKNTYAM VLAGGQDCRL QQLTECCAKA AVFVGGKFR I IDFVLSNCMN SGIRHIGVAT QYRSQNLQH IKRG--WSFL 100  
 67 15 ASGDPEPRFV SRLTKNTYAM VLAGGRGSRL HELTNWRAKP AVPFGGKFR I IDFVLSNCVN SGIRRIGVAT QYKSHSLIQH IQRG--WSFL 102

68 4 ---QHSDFI STLTKNAAI ILAGGRSRL KNLTDWRAKP AVQFGGKFR I DFPLSNCLN SGIRRVGVL QYKAQSLIQH LQRG--WGFL 89

69 8 -----FV SRLTRDTLAL ILAGGRSRL KHLTAWRAKP AVPFGGKFR I DFPLSNCLN SGVRRIGVLT QYKAHSLVRH IQQG--WGFM 88

70 7 TGVANSRPFV SNLTKNLAL VLAGGRSRL GPLTDWRAKP AVPFGGKFR I DFCLSNCLN SGIRRVGVL QYKAHSLIRH LQIG--WGFL 94

71 7 -----YI SSSLRETYAL ILAGGRSRL HELTDWRAKP AVYFGGKHRI I DFPLSNCLN SGVRRVGIAT QYKSHSLIRH VNRG--WGHF 87

72 1 -----MKK-QCVAM LLAGGKGSRL SELTKNMAKP AVSFGGKYRI IDFTLSNCSN SGIDTVGILT QYQPLELSNY IGIGSAWDL 78

73 5 -----QS TQLVRRITIAL VLAGGRSRL KALTDRAKP AVYFGGKFR I VDFAMSNCLN SGIRRVGVL QYKSHSLIRH LQRG--WSFL 85

74 7 -----LPST EEMANETLVL ILAGGRSRL YELTDQRAKP AVYFGGGHRI IDFALSNCIN SGLLKIGVIT QYEAHSLIRH LQHG--WSFL 89

75 7 KEEQINRKRK HFYRDNVGVI VLCGGEGKRL SPLTCWRCKP TVSFGGKYRL IDVPISHAFA SEFSKIFVIG QYLTYTLQOH LFKTYFY--H 94

76 6 -----V SELTRNTLAL VLAGGEGSRL KELTQWRAKP AVPFGGKYRI IDFVLSNCLN SDIRRVGVL QYKSHSLIRH IQRA--WSFM 85

77 24 -----RFV SRLTKSTLAL VMAGGRSRL GPMTQWRAKP AVPIAGKFR I DFVLSNCLN SGIRRVGVL QYKSHSLIQH VQKA--WNFL 105

78 8 -----LE RRLPKRAMAL ILAGGRSRL KQLTDTRCKP AVYFGGHFR I IDFVLSNCLN SGLRRIGVLT QYKSHSLIRH LQRG--WNFL 88

79 1 -----MLDKTLTI ILAGGVGSRL HPLTDARAKP AVPFGGNYRI IDFTLSNCLH SGLRRMLVLT QYKSHSLQKH LRDG--WSIF 76

80 9 -----DHNI SLLTRNTIAL ILAGGRSRL NMMTDRAKP AVPFGGKFR I DFPLSNCLN SGIRRVGVL QYKSDSLIRH IQQG--WGFL 91

81 1 -----MQDTLTV ILAGGMGARL APLTDNRAKP AVPFGGKYRI IDFTLNCNH SGLRRILVLT QYKSHSLQKH LRDG--WSIF 75

82 7 -----YI SNLTKDTYAL ILAGGRSRL HELTDWRAKP AVYFGGKFR I DFPLSNCLN SGIRRVGIAT QYKSHSLIRH VNRG--WGHF 87

83 1 -----MREVPVHGLI VLAGGEGKRL YPLTDARAKP AVPFGGKYRI IDFVLSNCLN ARYLKICVLT QYKSHSLDRH ISQN--WSIF 78

84 7 -----VSSK YTLAKDTLVL ILAGGRSRL HELTDKRAKP ALYFGGNRRI IDFALSNCIN SGLNHIGVIT QYAAHSLLRH LQKG--WSFL 89

85 7 -----FV SRLTRDTLAL VLAGGRSRL YELTDSRAKP AVYFGGKFR I DFPLSNCLN SGIRRVGVL QYKAHSLIRH LVNG--WGSF 87

86 7 -----RFV SRLTRDTLAI ILAGGRGRL ANLTDWRKAP ALPFGGKFR I DFPLSNCLN SGVRRIQVLT QYKAHSLIQH VQRG--WGFL 88

87 2 -----VA SQLPKRTVAL VLAGGRGTRL HQLTDNRAKP AVYFGGKYRI IDFALSNCIN SGIRRVGVL QYCPHSLIRH LQRG--WSFL 82

88 11 -----LPNK YELVKDTLVL ILAGGRSRL YELTDKRAKP ALYFGGNRRI IDFALSNCIN SGLNRIGVLT QYAAHSLLRH LQNG--WSFL 93

89 11 -----LA QQLPKETIAL VLAGGRGTRL KALTSKRAKP AVYFGGKFR I DFVLSNCLN SGIRRVGVL QYQSHSLVQH IQRG--WSFF 91

90 1 -----MKTTECLAM ILAGGQGSRL GALTQWRAKP AVPFGGKYRI IDFVLSNCLN SGIEKVGILT QYRPLELNQY LGSGSAWDL 79

91 8 -----FLHN TDIADKTLVL ILAGGRSRL HEMTDERAKP AVYFGGNRRI IDFTLSNCLN SNLLRIGVIT QYEAHSLLRH LQHA--WSFL 90

92 3 -----EYPRFV SLLTKNTVAL ILAGGRSRL KNLTEWRAKP AVPFAGKFR I DFPLSNCLN SGIRRVGVL QYKAHSLIQH LHRG--WSFL 104

93 17 YRAELAPRFI GALTKKTYAM VLAGGRSRL QQMTDWRAP AVPFGGKLR I DFPLSNCLN SGIRRVGVL QYMAHSLIHH IQRG--WSFL 87

94 1 -----MKVLAM VLAGGQGSRL HPLTAERSKP AVPFGGKYRI ADFVLSNCLN SNIRGVYLLV QYKQSLIEH VNKA--WGYA 74

95 6 -----RFV SRLTRDTLAL ILAGGRSRL KHLTLWRAKP AVPFGGKFR I DFPLSNCLN SGIRRVGVL QYKAHSLIQH VQRG--WSFL 87

96 1 -----MQT ---DRIAF VMAGGQGSRL QPLTARSKP SVPFSGKYRI VDFVLSNCLN SQIRTYLLV QYKQSLIEH VRKS--WTS 77

97 1 -----MAGVLGM ILAGGEGSRL KPLTERTKPK AVPFGGKYRI IDFALNFCN ADLMRIYVLT QYKQSLIYH MKKG--WNLS 75

98 1 -----MQDTLAV ILAGGMGSRL SPLTDDRKP AVPFGGKYRI IDFTLNCNH SGLRRILVLT QYKSHSLKH LRNG--WSIF 75

99 1 -----MAGVLGM ILAGGEGSRL RPLTESRKP SVPFGGKYRI IDFALNFCN ADLMKIYVLT QYKQSLFHH MKKG--WNLS 75

100 1 --MHNHYIE VLTMQDTLTI VLAGGVGSRL SPLTDNRAKP AVPFGGKYRI IDFTLANCLH SGLRQILVLT QYKSHSLQKH LRDG--WSVL 86

101 1 -----MLRKEMIAM ILAGGQGSRL GVLTKNVAKP AVYFGGKYRI IDFALSNCIN SGIDTVGILT QYQPLKNAH IGIGKPWDMD 79

102 4 -----RV QPLARDAMAY VLAGGRSRL KELTDRAKP AVYFGGKARI IDFALSNCIN SGIRRVGVL QYKAHSLIRH LQRG--WDF 84

103 1 -----MVVT GNLAGNTIAM VLAGGGERL APLTLRRPKP GVAFGGKYRI IDFVLSNCLN SGIRRVGVL QYRAYSLMKH IRES--WGKW 82

104 1 -----MRQVTAI ILGGGRGTRL YPLTKRRAKP AVPIGGKYRI IDIPVSNCLN SGIQHIYILT QFNASLNRH VSQTYQ--FS 75

105 8 -----LDI NRALKETLAL VLAGGRSRL RDLTNRKPK AVPFGGKYRI IDFVLSNCLN SGIRRMVLT QYRAHSLIHH IQRG--WGFL 89

106 74 DSQNSQTCLD PDASSSVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASLNRH LSRAYASNMG 163

107 1 ----MAVQM EFSMKRVLAI ILGGGAGTRL YPLTKRRAKP AVPLAGKYRL IDIPVSNCLN SEIHNIVVLT QFNASLNRH IARTYT--FP 83

108 1 ----MKRVLAI ILGGGKGSRL YPLTKRRAKP AVPLAGKYRL IDIPVSNCLN SNITKMYVLT QFNASLNRH LAQTYN--LS 75

109 1 ----MRQVTAI ILGGGRGTRL YPLTKRRAKP AVPIGGKYRL IDIPVSNCLN SGIQHIYILT QFNASLNRH VSQTYQ--FS 75

110 1 ----MDNVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SDINKVYCLT QFNASLNRH LSQAYNTNIG 77

111 37 S-----KTVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKIYVLT QFNASLNRH LARTYFNPG 113

112 66 AAGTGQNDPA GDISKTVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNVTKIYCLT QFNASLNRH LSQAYNSVG 155

113 60 R----EQDR QPRGEVCSI ILGGGAGTRL FPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKIYVLT QFNSTSLNRH LGRAYNMGSG 144

114 56 ASE--DQALE ARNSKTVVAV ILGGGAGTRL FPLTRRAKP AVPIGGAYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LSRAYDFNSG 145

115 33 DS----TYLN PQAHDVSLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASLNRH LSRAYNSNIG 118

116 64 DSRSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNVSKIYVLT QFNASLNRH LSRAYGNIA 153

117 66 AAAA-AAARR DVSPDVTASI ILGGGAGTRL FPLTRRAKP AVPEGGCYRL IDIPMSNCLN SKINKIYVLT QFNASLNRH IARTYFNFGG 154

118 72 Q----SSRKN YADANRVAI ILGGGTGSQL FPLTSTRATP AVPEGGCYRL IDIPMSNCLN SGINKIFVMS QFNSTSLNRH IHRTY--LEGG 156

119 82 AP----VEETP QADPSNVAI ILGGGAGTRL FPLTSRAKP AVPIGGCYRL IDIPMSNCLN SGIKKIFILT QFNFSLNRH LARTYFNFGG 168

120 81 RLR--DLEME KRDPRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGCYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LARAYNFGG 168

121 83 P----RFERR KADPKNVAI ILGGGAGTQL FPLTRRAATP AVPLGGCYRL IDIPMSNCLN SGINKIFVLT QFNSTSLNRH LARTY--FGNG 167

122 76 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASLNRH LSRAYASNMG 165

123 83 P----IFERR RADPKNVAI ILGGGAGTQL FPLTIRQATP AVPEGGCYRL IDIPMSNCLN SNINKIFILT QFNASLNRH IARTY--FGNG 167

124 73 KLR--DLEME KRDPRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGCYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LARAYNFGG 160

125 63 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASLNRH LSRAYASNMG 152

126 78 GP----VEERQ HADPSSVAI ILGGGAGTRL FPLTSRAKP AVPIGGCYRL IDVPMSNCLN SGIRKIFILT QFNASLNRH IARTYFNFGG 164

127 60 ASE--DQALE ARNSRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LSRAYDCTNG 147

128 64 DSRSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKRRAKP AVPLRANRYL IDIPVSNCLN SNVSKIYVLT QFNASLNRH LSRAYGNIA 153

129 73 Q----SSRKS YADANRVAI ILGGGTGSQL FPLTSTRATP AVPEGGCYRL IDIPMSNCLN SGINKIFVMT QFNSTSLNRH IHRTY--LGGE 157

130 75 ----SFRRN YADPNEVAIV ILGGGTGSQL FPLTSTRATP AVPIGGCYRL IDIPMSNCLN SGINKIFVMT QFNASLNRH IHRTY--LGGG 159

131 28 S-----KSVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCLN SGISKIYVLT QFNASLNRH LARTYFNFGG 104

132 55 IATEETATEVN DN--TDNVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SDINKMYCLT QFNASLNRH LSQAYNNVNG 145

133 77 TP----VEETP RADPKKVAI ILGGGAGTRL FPLTSRAKP AVPIGGCYRL IDIPMSNCLN SGIRKIFILT QFNFSLNRH IARTYFNFGG 163

134 83 P-----RFERR KADPKNVAI ILGGGAGTQL FPLTRRAATP AVPEGGCYRL IDIPMSNCLN SGINKIFVLT QFNASLNRH LARTY--FGNG 171

135 80 FQAT--VLRQ EADPKTVASI ILGGGAGTRL FPLTRRAKP AVPEGGCYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH IARTYFNFGG 168

136 86 KLR--DLEME KRDPRTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LARAYNFGG 173

137 75 DSRNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKRRAKP AVPLGANRYL IDIPVSNCLN SNISKIYVLT QFNASLNRH LSRAYASNMG 164

138 76 KVH--ELETE KRDSRTVASI ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCLN SGINKIYVLT QFNASLNRH LARAYN--SNG 162



139 76 S----MFERR KADPQNVAAI ILGGNGAKL FPLTMRRAATP AVFVGGCYRL IDIPMSNCIN SCINKIFVLT QFNASASLNHR LARTY-FGNG 160  
140 19 -----IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNASASLNHR LARTYFNGG 103  
141 16 -----IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNASASLNHR LARTYFNGG 100  
142 88 LPFS-VFETP RVDPKSVVSI ILGGGVGTRL FPLTKQRAKP AVPIGGYRL IDVPMSNCIN SGINRVFVLT QFNASASLNHR LARTYF--- 173  
143 53 LPFS-VFETP RVDPKSVVSI ILGGGVGTRL FPLTKQRAKP AVPIGGYRL IDVPMSNCIN SGINRVFVLT QFNASASLNHR LARTYF--- 138  
144 7 ARDESSPYLE PDARSSVLAV ILGGGAGTRL HPLTKRRAKP AVPLGANRYL IDIPVSNICIN SNIPRIYVLT QYNSTSLNSH LYRAYAGNMG 96  
145 13 DSMEEICLCK PDAGVSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SNIRKIYVLT QFNASASLNHR LSRAYSSNMG 102  
146 1 -----MQSVLAV ILGGGAGTRL HPLTKRRAKP AVPLGANRYL IDIPVSNICIN SNIPRIYVLT QYNSTSLNSH LYRAYAGNMG 77  
147 3 -----IGKP RADPRTVVSL ILGGGAGTRL FPLTNRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIFILT QFNASASLNHR LARTYFNGG 87  
148 63 APQLRYEPAT KARTNTVLSI ILGGGAGTRL FPLTKQRAKP AVPIGGAYRL IDVPMSNCIN SGISKIYILT QFNSTSLNHR LARAYNMGSS 152  
149 64 TAAAPASYDYA GDISKTVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SNVTKIYCLT QFNASASLNHR LSQAYNSSVG 153  
150 58 TTTTACKVVV DSDTDNLVLI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SDINKMYCLT QFNASASLNHR LSQAYNSVG 147  
151 64 S-----KSAVAV ILGGGAGTRL YPLTKKRAKP AVPIGGAYRL IDVPMSNCIN SGISKIYILT QFNSTSLNHR LARTYFNGG 140  
152 1 -----MDNVLSI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SDINKIYVLT QFNASASLNHR LAQAYNTNIG 77  
153 19 T-----KTVAAV ILGGGAGTRL YPLTKSRAKP AVPIGGAYRL IDVPMSNCIN SGISKIYILT QFNASASLNHR LARTYFNGG 95  
154 75 P----SFLRR RADPKNVISI ILGGGPGTQL FPLTKRAATP AVFVGGCYRL IDIPMSNCIN SGINKIFVLT QFNASASLNHR LARTY-FGNG 159  
155 84 VP---TFEKP EVDPKSVASI ILGGGAGTRL FPLTKRRAKP AVPIGGCYRL IDIPMSNCIN SGIRKIFILT QFNSTSLNHR LSRAYSFGNG 170  
156 75 P----SFLRR RADPKNVISI ILGGGPGTQL FPLTKRAATP AVFVGGCYRL IDIPMSNCIN SGINKIFVLT QFNASASLNHR LARTY-FGNG 159  
157 85 VP---TFEKP EVDPKSVASI ILGGGAGTRL FPLTKRRAKP AVPIGGCYRL IDIPMSNCIN SGIRKIFILT QFNSTSLNHR LSRAYSFGNG 171  
158 86 P----SFLRR RADPKNVISI ILGGGPGIQL FPLTKRAATP AVFVGGCYRL IDIPMSNCIN SGINKIFVLT QFNASASLNHR LARTY-FGNG 170  
159 84 P----SFLRR KADPKNVISI ILGGGPGIQL FPLTKRAATP AVFVGGCYRL IDIPMSNCIN SGINKIFVLT QFNASASLNHR ISRTY-FGNG 168  
160 77 GP---IFQSP KANPENVAI ILGGGAGTRL FPLTSTRAKQ AVPIAGCYRL IDIPMSNCIN SGIRKIVVLT QFNSTSLNHR LSRAYFNGG 163  
161 78 GP---IFQNP KANPENVAI ILGGGAGTRL FPLTSTRAKQ AVPIAGCYRL IDIPMSNCIN SGIRKIVVLT QFNSTSLNHR LSRAYFNGG 164  
162 76 KLR---DLDME RRNPRTVLAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIYILT QFNASASLNHR IARAYNSGNG 163  
163 57 ASE---DADTE TRNARTVVAV ILGGGAGTRL FPLTKRRAKP AVPIGGAYRL IDVPMSNCIN SGINKIYVLT QFNASASLNHR LSRAYFNGG 144  
164 78 -----SFRRN YADPNEVAV ILGGGTGTQL FPLTSTRATP AVPIGGCYRL IDIPMSNCIN SGINKIFVMT QFNASASLNHR IHRTY-LGGG 162  
165 71 DSKSSQTCLD PDASTSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SNISKIYVLT QFNASASLNHR LSRAYSGNIG 160  
166 68 DSKNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SNISKIYVLT QFNASASLNHR LSRAYSNMG 157  
167 84 P----SFIRR KADPKNVASI VLGGGPGVQL FPLTKRAATP AVFVGGCYRL IDIPMSNCIN SGINKIFVLT QFNASASLNHR LARTY-FGNG 168  
168 69 DSQNSQTCLD PDASRSVLGI ILGGGAGTRL YPLTKKRAKP AVPLGANRYL IDIPVSNICIN SNISKIYVLT QFNASASLNHR LSRAYSNLG 158  
  
1 85 RPERN-ESFD ILPA--SQRV SETQWYEGTA DAVYQNDIIE EPY---APE YMVILAGDHI YKMDYEMLQ QHVDGADV T IGCLEVPME 167  
2 92 NEEMN-EFVD LLPA--QORM KGENWYRGTA DAVTQNDLII RRY---KAE YVVLVLAGDHI YKQDYSRMLI DHVEKGARCT VACMPVPIE 174  
3 95 GYKNE-GFVE VLAAQQSP-- ENPDWFQGT DAVRQYLWLF EEHT---VL EYLILAGDHL YRMDYKFKFI AHRETADIT VAALFMDPKR 177  
4 111 VSFGD-GFVE VLAATQTPGE AGKKWFQGT DAVRKFIVWF EDA-KKNKIE NIVVLSGDHL YRMDYMLVQ NHIDRNADIT LSCAPAEISR 198  
5 89 RGEFG-EFVE LVPA--QORM DKPLWYAGTA DAVYQNDIIE KAH---NPS YVLVLAGDHI YKMDYGGMLA RHAESGAAMT VGCEVPRKR 171  
6 89 RCDFG-EFVE ILPA--QORT EG-SWYRGTA DAVYQSLDIV RMH---DPD YVLVLAGDHI YKMDYGGMLA RHVETGADVT VGCEVPEE 170  
7 79 PLLPH-HFVT VVPP--QMQR GP-EWFQGT DSVYQNLHLV ELV---KPD LVVVFGADHV YRMDLRQMTIE FHAIATQADAT VAALFPIEH 160  
8 125 PGAG--SIE ILTAEQKP-- SKKNWFQGT DAVRQNDIYL LES----PFE YFLILSGDQL YNIDFQEMVH FAKKNSDSDVV VATIPVNTQD 205  
9 95 GVLQD--QIH LLAPEGRD-- GSQVYKGT DAIQNLILYL EDT---GIE YFLVLSGDQL YNMDFRKIVD YALSMQSDMV IVAQPIQEK 176  
10 94 GVLQD--QIH LLAPEARQ-- GDQIYWGTA DAIRQNLILYL EDT---EIE YFLVLSGDQL YNMDFRSIVD TAIKTHYDVM LVAQPIQEK 175  
11 103 DGRFD-EFIQ LLPA--QQQI DETQWYQGT DAVYQNLHFL RRY---QPD HILVVGADHI YKMDYGRMLA HHVKHHDAMT VACIDVPLDE 185  
12 91 RGEFN-EFVE LLPA--QORI Q-EWYKGT DAVFQNLIDIL RQT---NIE FVLILAGDHI YKMDYQMLA AHVRNKADMT VACINVLPE 172  
13 95 DGRFQ-EFIE LLPA--QORT EBTWYQGT DAVFQNLIDIL RTH---NPG YVLILGGDHI YKMDYGRILA EHVERQADLT IACLEVPVED 177  
14 87 DGRFH-EFIE LLPA--QORT EGTWYQGT DAVFQNLIDII RTH---NPS YVLVLAGDHI YKMDYQMLA EHVEKQADLT IACLEVPED 169  
15 89 DGRFQ-EFIE LLPA--QORT EES-WYQGT DAVYQNDIIE RSH---NPD YVLVLAGDHI YKMDYAKLLA DHIAKSAEMT IACIDLPLEE 170  
16 76 G-FSE-GFVE VLAAQQTP-- ENPNWFQGT DAVRQYLWML QSD---RFD EFLVLSGDHL YRMDYRLFIQ RHRETADIT LSVIPIDDR 157  
17 86 RGEFG-EFID LWPA--QORV EGAWHYRGTA DAVFQNLIDII RSH---RFD YVVLVLAGDHI YKMDYTRMLA DHESGADCT VACIEPDMR 168  
18 85 RNEVN-EFID LLPA--QORV DEASWYRGTA DAVYQNDIIE REH---DPK YILVLAGDHI YKMDYASLIE DHVALGAPCT VACIEVPLAE 167  
19 86 G-FQE-GFVE VLAAQQTK-- DNPWFQGT DAVRQYLWLF REWD---VD EYLILSGDHL YRMDYQFVK RHRETADIT LSVVPEDDR 167  
20 100 RTEVG-EFVE ILPA--QORT HKKEWYQGT DALFQNLIDIM QRH---HPE YVLVLAGDHI YTMDDYQMLL YHVQTAGDVT VGSVEVPAE 182  
21 68 SLFAD-QFLT VVPP--QMTK TS-TVFGGT DAVYQSLDLM NMH---RPD LVAIFGADHI YRMDVROMVR FHCEHDAEAT VAALFVSLNQ 149  
22 101 DGRFK-EFVE LLPA--QORT VEETWYQGT DAIFQNLIDIL LRH---DAK YVLVLAGDHI YKMDYSKLLA EHIEKSADMT VACLEIPLQE 183  
23 82 NHELG-EFIE CIPP--QORN VD-RWYRGTA DSIYQNIHIL QRE---RPE RVLVLAGDHI YKMDYNDMLA FHIEKNAQLT VAGVEVDRSE 163  
24 79 GINSG---AT ILQP--YSAT EGNRWFGT HAIYQNDIYI DSI---NPE YVLVLAGDHI YKMDYDMLQ THKDNMASLT VAVIDVPLKE 159  
25 84 --FSR-GFVE ILAAEQTF-- DHSGWYEGTA DAVRKNFQHF RTQ---NPS HYLILSGDQL YRMDLAEMYR RHLESGAQT IAGTLVTRQE 165  
26 92 NEEMN-EFVD LLPA--QORL STEQWYKGT DAVQNLIDII RRY---DAE YVILVLAGDHI YKMDYSRMLL DHVEKGAECT VACIPVPIE 174  
27 88 KKELG-ESVE ILPA--SQRF SD-SWYEGTA DAVFQNDIIE RDE---LPK YVMILSGDHI YRMDYGTMLA RHVESGAKMT VSCMSVPIE 169  
28 102 DGRFK-EFVE LLPA--QORT AEBTWYQGT DAVFQNVDDL QRH---DAK YVLVLAGDHI YKMDYSKLLD EHIEKAADMT VACLEVPVEE 184  
29 100 PFSGG--FIE LLAPEQKP-- HKKTWYQGT DAVRQSLIEF IET---PVD YFLVLSGDQL YNMDFRPMLQ FAHENDADLV VASHVFNKAD 181  
30 86 KSEMN-EFVD LLPA--QORV DEESWYRGTA DAVYQNDIIE AAYK---AD YVVLVLAGDHI YKMDYALMLA DHVAQAGRECT VGCEVPRQD 170  
31 86 NNSFG-CRDL LLPA--SQQQ SE-SWYQGT DACEFNIDYI KSR---APK YVMILSGDHI YQMDYRKLIA EHVKNAGRECT VSCIEVPTKN 167  
32 80 RLNGG---VF ILPP--HQKA SGANWYQGT DAVYQNIPII DAY---EPN LVVLVLAGDHI YKMDYNDMLA FHKDKGARCT IAVTDVPLSE 160  
33 79 --FSN-GFVE ILAAEQTF-- HSEWYQGT DAVRKNLKHFI RQD---AAD YVILVLAGDHI YRMDYQMLL KHIESGAEIT IAAKPIREQ 160  
34 76 --FSH-GFVE ILAAEQTL-- EHSGWYEGTA DAVRKNFHF KDQ---NPT HYIILSGDQL YRMDLKKFLD KHIESGADIT IATTSVTRED 157  
35 94 GVLQD--HIH LLVPEGRQ-- GNQIYWGTA DAIRQNLILYL KTL---DLI YFLVLSGDQL YNMDFHVIVE SMISSQADM LVAQFVSEKD 175  
36 79 NPEMG-EYII SIPP--QOVT VN-RWYRGTA DSIYQNISIL QSE---RPD YVLVLAGDHI YKMDYMEMLN YHIDKRADMT AASVEFPRLE 160  
37 79 RFIQF-HFIT VVPP--QMRN GP-EWFQGT DSVYQNLHLI ESF---KPD IVAVFGADHI YRMDVROMD FHVNDADHVS VATLFPVLAD 160  
38 79 FLIPH-HFIT VVPP--QMQR GP-EWFQGT DSVYQNLHLI DLL---KPD LVVVFGADHV YRMDIRQMVQ YHVKTHADAT VAALFPIEQ 160  
39 89 SSDFG-EFIE LLPA--QORI AD-SWYLGTA DAVYQSLDIV RLH---DPD YVLVLAGDHI YKMDYGPLLA YHVERGADV T VSCLEVAIEE 170  
40 89 RGEFG-EFVE LIPA--QORM DKPLWYSGTA DSVYQNDIIE QAH---DPS YVLVLAGDHI YKMDYGAMIA RHVESGADV T VGCVQVTELEQ 171

41 88 KTEMN-EFVD LLPA--QQRN DNESWYRGTA DAVHQNYDIL ESYG----AD YIVVLGADHI YKMNYALMLA DHVAKGRDCT VGCIAVPRHE 170

42 88 KANLS-EFVE VLPA--SQGN NN-DWYLGTA DAVYQNDIID CAE----RPK YVLILSGDHV YRMDYGPLIA EHVANNADMT VCCLKATTEE 169

43 76 GIGKANRFIE AIPA--QQRV NK-NWYSGTA DAIYQNRFI EKS----AAE HVCIFGSDHI YKMDVQOMVE HHERKGGALT VSAIRIVKEQ 158

44 76 NITDT--FID AIPA--QMRK GK-HWYSGTA DAIYQNLQFI ESD----EAE LVCIFGSDHI YKMDIRQKIA YHOEKEAVLT VSAIRLPEKE 156

45 85 RPERN-ESFD ILPA--SQRV SEELWYLGTA DAVYQNDIID ESY----DPQ FIVLLAGDHI YKMDYKMLQ QHVQGGAHVT VGCIEVPRKE 167

46 90 GEMG-EFVE LLPA--QQRD DE-SWYAGTA DAVYQNDIID RRH----NPE YVLILGADHI YKMDYGTMLA AHVERGADIT VGCIEVPLDI 171

47 77 NPELG-EYIT VPVA--QMNS GE-HWYQGT ADFQNLNLL ERS----NAE YTLILSGDHI YRMDYAAML AHOEQGADVT IACMEVPEE 158

48 86 RGEFN-EFID LWPA--QQRV EGAHWYRGTA DAVFQNDIID RSI----RPK YVVVLGADHI YKMDYTRMVM DHVESKADCT VGCIEVPRME 168

49 70 PMLDS--FIE TVPA--QQR T GK-SWFKGSA DAVYQTOHVI TDE----SPE HLCIFGGDHV YKMDVROMLH DHLSDAEVT VAAIPVTKEE 150

50 88 RNEMG-EFVD LLPA--QQR I DEEQWYQGT ADFQNDIID RNS-T--PPD YIVVLGADHV YKMDYSIMLE DHAASGRGVT VGCIEVPREE 171

51 76 GPFQ-GFVE VLAQQQTP-- DSPKWFEGTA DAVRKYQWLF QEWD----VD EYLILSGDL YRMDYSLFVQ HHRDNGSDDLT VAALFVDEAQ 158

52 85 REERN-EYLD ILPA--SQRV DEHKWYLGTA DAVQNDIV DSY----DIK YVILGADHV YKMDYIEMLR QHCETGADVT IGCLETPRME 167

53 89 RGYLG-EFVE LLPA--SQRI ED-SWYAGTA DAVYQNDIID RTH----NPD YVLVLGADHV YKMDYGDMLA YHVESADMT VGCIHVPLKE 170

54 82 PLLPD-QFVT AVPP--QMHE DT-LTFKGT ADFYQSLRLL EPH----NPD LVAVFGADHV YRMDVROMAW FHREKQADVT VAALFVPEMEQ 167

55 85 RAERN-EYLD ILPA--SQRI DESKWYLGTA DAVQNDIID EY----DVK YVILGADHI YKMDYEVMLL QHVLKADVT IGCLETPRAE 163

56 98 RAELN-EMVD LLPA--QQRV DEEHWYRGTA DAVYQNDIID QSS----KPE YVVILGADHV YKMDYSIMLQ DHATSGAQT VGCIEVPRSE 180

57 85 HGEVN-EFVD LLPA--QQR I DEESWYRGTA DAVYQNDIID EYTP--APE FVVILGADHV YKMDYAVMLV DHVESGAECT VACIEVPRER 169

58 89 AMLDH--YVE VPVA--QQR M GK-HWFLGSA DALYQSFNVV TDE----NPE YVCVFGDHI YRMDVROML S FHIACHADAT VAALFVPEASE 169

59 90 RGEFN-EFVS LLPA--QQR I Q-EEWYKGT ADFQNDIID RNT--T--PPD YIVVLGADHV YKMDYGMMLA YHVKKADMT VACVNEVED 171

60 90 RGEFN-EFVS LLPA--QQR I Q-EEWYKGT ADFQNDIID RGY----NPE YIVILGADHV YKMDYGMMLA YHVESADMT VGCIEVPEAGQ 171

61 78 --FSG-GHVS ILAAEQTD-- TNIDWYQGT ADRKLNDFI DNE----FVN NVVILSGDQV YRMDYVMLQ HMETGADIV VGTVPVVED 159

62 76 GLTDH--FID PIPA--QMR M GK-HWYKGT ADRYQNLPLI DTY----DPE VCVFGGDHI YKMEIRQMDI FHRNKRAALT VAAIPVSEK 156

63 82 DMFTN-GWVQ IWAEEQTP-- DSTGWYQGT ADRVQOMVEI KNS----GIK YVLVLGADHL YRMDYRKFVQ YHVDTKADIT LAVQFVNGLE 164

64 97 RAELN-EMVD VLPA--QQR T GDEHWYRGTA DAVYQNDIID QTRST--KHD YVVVLGADHI YKMDYSIMVK DHAERGLGCT VGCIEVPRME 181

65 77 RQFGG--AT LLQP--YT G KGG-WYQGT AHAIQNYNLI KDI----DPE YVILSSDHV YKMDYSKMN YHKEKADLT IAVKPVSMKE 156

66 101 NGHFS-EFVD LLPA--QQR V SAGHG YRGTA DAVYQNDIID RAH----TPE FVLILSGDHV YKMDYKLLA FHVTRADMT MACVEVEVSG 183

67 103 NGQFG-EYLD LLPA--QQR I SEDQWYQGT ADFQNDIID RAS----KCE FIVILGADHI YKMDYKLLA FHVVEKADMT VACIEVPIAE 185

68 90 RGEFN-EYVN IIPA--QQR I S-EEWYKGT ADFYQNDIID REG----GPE YLILGADHI YKMDYKMLA THVKSADMT VACINPLED 170

69 89 RGYLG-EFVE LMPA--SQRI ED-SWYAGTA DAVYQNDIIV RSH----NPE YVLILGADHV YKMDYGDMLA YHVERADMT VGCIHVPLKE 171

70 95 RGEFS-EFIE ILPA--QQR M N-TGWYKGT ADFQNDIID RAH----RPR YVVILGADHI YRMDYGMMLA EHVQTOADMT VACIEVGLLE 176

71 88 KKELG-ESVE ILPA--SQRH GD-EWYCGTA DAVFQNDIID RHE----LPK YVMILSGDHV YRMDYGALLA EHVQKADMT VCCIEVEVEE 169

72 79 RYNGG--VT VLPP--YAES SEVKWYKGT SAIVENLNYL NQY----DPE YVLILSGDHI YKMDYKMLD FHIKADVT ISVIEVSEE 159

73 86 RGEFN-EFVD LLPA--QQR I DEEHWYRGTA DAVFQNDIID ASG----KAE YVVVLGADHI YKMDYSVMLK DHVEHGAGCT VGCIEVPRME 168

74 90 PRERG-QFVD MLPA--RQQL NDQTYWYRGTA DAVWQNVHIM KDHY--KPK YVLILGADHI YKMDYQMR DHVTSGAKVT VGCIEVPREQ 173

75 95 GVMQD--QTH LLVEERRD-- GSQVWYQGT ADRQNLVLL QDS----RVE YFLILSGDL YNMDFRSIVD YADDAQADMV IASQVSDK 176

76 86 RYEVG-EFVE LLPA--QQR L G-KEWYQGT NALYQNDIIV RRH----NPE YVLVLGADHI YAMDYRDMIA THAASGADVT VGCIEVPRME 167

77 106 GGEFG-EFVE LLPA--QQR I DENSWYMGTA DAVYQNDIID RAH----EPS HVLILGADHV YKMDYGRMLA HHVEKGAQTS VGCIEVPEE 188

78 89 KSEMH-EFVD LIPA--QQR V DEEYWYRGTA DAVYQSDIIV KSNK----PE YVVILGADHI YKMDYARMLA DHALSAGVT VGCIEVDRQE 171

79 77 NPEIS-EYIT VPVP--QMRT DQ-SWYSGTA DAIRQNLVLL ERS----NAS HVLILSGDHI YRMDYAAMLQ FHRDQAGLT IACMPVSLVS 158

80 92 RGEFG-EYVD LMPA--QQR H DENSWYEGTA DAVYQNDIID RSR----HPE HLLVLGADHI YKMDYGMLA DHVEKNADLT IGCIEVSLQD 174

81 76 NPELG-EYIT VVPP--QMR K GD-KWYSGTA DAIYQNLWLL SRS----DAK YVVVLGADHI YRMDYAPMLE RHKETGADLS IACMEVPAE 157

82 88 KKELS-ESVE ILPA--SQRY GN-DWYSGTA DAVYQNDIID RAE----MPK YVMILSGDHV YRMDYDMLA KHVENGADMT VCCIEVTEE 169

83 79 G-LAG-EYIT VPVA--QQR L GP-RWYTGSA DAIYQSLNLI YDE----DPD YIVVFGADHV YRMDPEQMV RPHIDSGAGT VAGIRVPREN 159

84 90 PBERG-EFID MLPA--RQOI DDSTWYRGTA DAVYQNMII RDHY--KPK YLILGADHI YKQDYGMML DHVNSGART VGCIEVRE 173

85 88 HTVLG-EFVE ILPA--SQRT TG-EWYAGTA DAIYQNDIID RIM--KPK YVLVLGADHI YKMDYGALLA YHVKKADMT VACVDVLED 169

86 89 RGEFG-EFVE IIPA--QQR L DKPLWFACT ADRQNDIID KAH----RPR YVLILGADHV YKMDYGMMLA LHVEHADMT VGCIEVPRER 171

87 83 RGEQN-EFVD LMPA--GQQ M EEWYRGTA DAVQNKGLI RSY--EPE YLVLGADHI YKMDYSMLL DHVESKSLCT VACIEVARE 165

88 94 PAERG-EFID MLPA--RQOI DDSTWYRGTA DAVYQNMII RDHY--CPK YLILGADHI YKQDYSQMLL DHVNSGART VGCIEVEREK 177

89 92 NEDMN-EFVD LLPA--QQR R NTEHWYMGTA DAIYQNDIID RNY----QAK YVVILGADHI YKMDYARMLL DHVEHKSFT VACIRVPEKN 174

90 80 KRQGG--LF VLPP--YAE KGAEWYRGTA DAIYQNLNFI DMA----DPA YVLILSGDHI YTMDYAWMLE HHKKNKAQT IGVFEVPEWDE 160

91 91 PRERG-QFVD MLPA--RQSV NEEMWYRGTA DAVQNMNIM KNHY--KPK YVLILGADHI YKMDYMNVM RSHIQSGART VGCIEVKKED 174

92 88 RGEFN-EFVE LMPA--QQR I DETMWYRGTA DAVFQNDIID RNY----DST YVLILGADHV YKMDYGMMLA FHAASAADMT VACIEVPIED 170

93 105 DDQFN-EFID VLPA--QQR V KEG-WYEGTA NAVFQNDIIV RSC----APE YVLILCGDHV YKMDYSRILA DHVAKAADVT VACIEVPAE 186

94 75 NSCPG-QFIT VVPP--QMR E GP-EWFQGT ADFQNLNLI KHH----APD MVAVFGADHI YRMDVROMLD FHOQKSAHV VAALFVPLAE 156

95 88 RGEFG-EFIE LLPA--QQR I ET-SWYAGTA DAVYQNDIID RQH----APE YVLILGADHI YKMDYGMMLA YHVESGADMT VGCIEVDRDH 169

96 78 PLLQD-QFVT VVPP--QMIG GE-SWYQGT ADFYQNLGLI ETH----APD LVVVFADHI YRMDIRQMV FHKEREADVT IAAIFVPLD 159

97 76 GITDR--FID IIPA--QMR D GK-RWYEGTA DAIYQNLRFV EIV----APD QVCIFGSDHI YKMDIRQMLD FHRMEALT VSAIRMPISQ 156

98 76 NPELG-EFIT VVPP--QMR K GG-KWYEGTA DALFHNWLL ARS----DAK YVVVLGADHI YRMDYAAML EHSKNAITL IACMVERE 157

99 76 GITDR--FID PIPA--QMR T GK-RWYEGTA DAIYQNLRFM QLE----EPD QVCIFGSDHI YKMDIRQMLD FHKKKAALT VSAIRMPLE 156

100 87 NPELG-EYIT NVPP--QMRT GD-SWYSGTA DAIYQNLVLL SRS----EAK HVVVLGADHI YRMDYAPMLK QHKQNEADLT VACMEVSI 168

101 80 RIDGG--VT ILSF--YLKA EMGEWFKGTA NAVYQNIQYI DRY----SPH YVILSGDHI YKMDYSKMLD FHKENHADAT ISVINVPEE 160

102 85 RPERN-ESFD ILPA--SQRV SETQWYEGTA DAVYQNDIID EPY----APE YMVILGADHI YKMDYEMLQ QHVDGADVT IGCLEVPRME 167

103 83 AGLGE-FFVA ISPE--TSSE SE-EWFKGT DAINHLYRFI ESS----DAD YVAIFGGDHI YRMDYSQMLG YHRRNADIT IAAIEVEE 164

104 76 R-FSD-GFCE ILAAEQTD-- ENPNWFQGT ADRVQYLWLL EPGS----ST EYLILSGDL YRMDYSKFRV RHRETADVT IAVLPDLE 157

105 90 RAETG-EFVE LWPA--QQR T DKESWYLGTA DAVYQNDIID RMH----DPR FVLILGADHI YKQDYSKLLA FHIARSGDCT VACVDVPEE 172

106 164 GYKNE-GFVE VLAQQQSP-- ENPNWFQGT ADRVQYLWLF EEHN----VL EYLILGADHL YRMDYKFTI AHRETADIT VAALFVDEQR 246

107 84 G-LTG-GFVE VLAQQQTP-- ENPNWFQGT ADFYQNLWLL ADWD----VD EYLILSGDL YRMDYRFLVQ RHRDGTADVT LSVLPVEKA 165

108 76 SPFAQ-EFVE VLAQQQTP-- ESPSWFEGTA DAVRKYQWLF QEWD----VD EYLILSGDL YRMDYSLFVE HHRDGTADVT VAALFVDAQ 158

109 76 R-FSD-GFCE ILAAEQTD-- ENPNWFQGT ADRVQYLWLF EPGS----ST EYLILSGDL YRMDYSKFRV RHRETADVT IAVLPDLE 157

110 78 TYTRQ-GFVE VLAQQQSP-- INKAWFQGT ADRVQYLWLF AESG----CE EYLILSGDL YRMDYRPFIR DHRAKNADIT VAALPTDEK 160

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	IAALPCAKE	241
113	145	VRFGDGFVE	VLAATQTP--	TDKEWFQGTA	DAVRQYSWLL	EDT-KNRAIE	DVLILSGDHL	YRMDYMKFVN	YHRETNADIT	IGCIAYGSDR	231
114	146	VATGD-GFVE	VLAATQRPQT	EGKRWFQGTA	DAVRQFDWLF	DDA-KSKDIE	DVLILSGDHL	YRMDYMDLVQ	SHRQRGAGIS	ICCLPIDGSR	233
115	119	GYKNE-GFVE	VLAAQQSP--	DNPWFQGTA	DAVRQYLWLF	EEHN----VM	EFLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEKR	201
116	154	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYMWLF	EEHN----IM	EFLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEQR	236
117	155	VGFSG-GSVE	VLAATQTAGE	SGKWFQGTA	DAVRQFLWLF	EDA-RLKICIE	NLILSGDHL	YRMDYMDLVQ	KHVDGADIS	VACVPMDES	242
118	157	INFAD-GSVQ	VLAATQMPPEE	PA-GWFQGTA	DSIRKFIWLV	EDYYSKSID	NIVILSGDQL	YRMDYMDLVQ	KHVEDDADIT	ISCAVPMDES	244
119	169	VSPGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFIWVF	EDA-RTKNVE	HVLILSGDHL	YRMDYMEFVQ	KHIDTNADIT	VSCVPMDDSR	256
120	169	VSPGD-GFVE	ALAATQTPGE	AGKWFQGTA	DAVRQFHWLF	EGP-RSKEIE	DVLILSGDHL	YRMDYMDLVQ	NHRQGGADIT	LSCVPMDDSR	256
121	168	IIFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	HHIDSNADIT	ISCAVPMDES	248
122	166	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYLWLF	EEHN----VL	EFLVLAGDHL	YRMDYERFIQ	AHRETNADIT	VAALPMDEKR	255
123	168	VNFGD-GSVE	VLAATQTPGE	AGKWFQGTA	DAVRKFIWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	KHVDGADIS	VSCVPMDES	242
124	161	VNFGD-GYVE	ALAATQTPGE	AGKWFQGTA	DAVRQFHWLF	EDQ-RSKEIE	DVLILSGDHL	YRMDYMDLVQ	NHRQGGADIT	ISCLPMDDSR	248
125	153	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYLWLF	EEHN----VL	EFLVLAGDHL	YRMDYERFIQ	AHRETNADIT	VAALPMDEKR	235
126	165	VNFGD-GSVE	VLAATQTPGE	AGKWFQGTA	DAVRQFIWVF	EDA-KNKNIE	HLLILSGDHL	YRMDYMDLVQ	KHIDSNADIT	VSCVPMDDSR	252
127	148	VAFGD-GFVE	VLAATQRPQS	EGKRWFQGTA	DAVRQFDWLF	DDA-KSKDID	DVLILSGDHL	YRMDYMDLVQ	SHRQRGAGIS	ICCLPIDDSR	235
128	154	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYMWLF	EEHN----IM	EFLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEQR	236
129	158	INFAD-GSVQ	VLAATQMPPEE	PD-GWFQGTA	DSVRKFIWLV	EDYYNHKSIE	HIVILSGDQL	YRMDYMDLVQ	KHVEDNADIT	VSCVPMDES	245
130	160	VNFGD-GSVE	VLAATQMPGE	AA-GWFQGTA	DAVRKFIWLV	EDA-KNKNIE	HLLILSGDHL	YRMDYMDLVQ	KHIDSNADIT	LSCVPMDES	247
131	105	IMYGGNGFVE	VLAATQTPGL	GGKEWFQGTA	DAVRQYSWLF	EDI-KNKDVQ	DIVILSGDHL	YRMDYMAFVA	RHREVNADIT	IGCLPMDDKR	193
132	146	SYNRQ-GFVE	VLAAQQSP--	KNKDFWQGTA	DAVRQYIWFV	NESE----CD	EYIILSGDHL	YRMDYKPFIL	KHRQTKADIT	VSAVPMDDSR	228
133	164	VNFGD-GFVE	VLAATKTPGE	AGKWFQGTA	DAVRQFIWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMEFVQ	KHIDGADIS	VSCVPMDDSR	251
134	172	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	HHVDSNADIT	ISCAVPMDES	259
135	169	VNFGD-GFVE	VLAATQTPGE	SGKWFQGTA	DAVRQFLWLF	EDA-KHSHIE	NLILSGDHL	YRMDYMDLVQ	KHIDGADIS	VSCVPMDES	256
136	174	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFHWLF	EDA-RSKDIE	DVLILSGDHL	YRMDYMDLVQ	NHRQGGADIT	ISCLPMDDSR	261
137	165	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYLWLF	EEHN----VL	EFLILAGDHL	YRMDYERFIQ	AHRETNADIT	VAALPMDEKR	247
138	163	VSPGD-GYVE	VLAATQTPGE	SGKWFQGTA	DAVRQFHWLF	EDA-RSKDIE	DVLILSGDHL	YRMDYMDLVQ	DHRQGGADIS	ISCIPIIDRR	250
139	161	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRKFIWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	LSCVPMDES	248
140	104	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFIWVF	EDT-RSKEIE	NLILSGDHL	YRMDYMEFVQ	KHIDGADIS	IGCVPMDDSR	191
141	101	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDT-RSKEIE	NLILSGDHL	YRMDYMEFVQ	KHIDGADIS	IGCVPMDDSR	188
142	174	VNFGD-GFVE	VLAATQTPGE	SGKWFQGTA	DAVRQFTWVF	EDV-RNKDID	YVLVLSGDHL	YRMDYMDLVQ	KHIDGADIS	ISCVPMDES	261
143	139	VNFGD-GFVE	VLAATQTPGE	SGKWFQGTA	DAVRQFTWVF	EDV-RNKDID	YVLVLSGDHL	YRMDYMDLVQ	KHIDGADIS	ISCVPMDES	226
144	97	GFRND-GFVE	VLAAEQSL--	DNPWFQGTA	DAVRQYIWFV	EDQD----VM	EFLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAAVPEEKR	179
145	103	NYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYMWLF	EEQP----VM	EYLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEKR	185
146	78	GFRND-GFVE	VLAAEQSL--	DNPWFQGTA	DAVRQYIWFV	EDQD----VM	EFLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAAVPEEKR	160
147	88	VNFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDT-RSKEIE	NLILSGDHL	YRMDYMEFVQ	KHIDGADIS	IGCVPMDDSR	175
148	153	VRFGDGFVE	VLAATQTP--	TDKEWFQGTA	DAVRQYSWLL	EDT-KNRAIE	DVLILSGDHL	YRMDYMKFVN	YHRETNADIT	IGCIAYGSDR	239
149	154	GYNTR-GFVE	VLAASQSS--	ANKSWFQGTA	DAVRQYMWLF	EEAVREG-VE	DFLILSGDHL	YRMDYRDFVR	KHRESGAAIT	IAALPCAKE	230
150	148	SGLRQ-GFVE	VLAAQQSP--	KSKVWFQGTA	DAVRQYMWLF	NESE----CE	EYIILSGDHL	YRMDYKPFIL	EHRKTADIT	VSAVPMDDSR	230
151	141	IMYGGNGFVE	VLAATQTPGL	GGKEWFQGTA	DAVRQYSWLF	EDV-KNKDVE	DVILSGDHL	YRMDYMAFVA	RHREVNADIT	IGCLPMDDSR	229
152	78	THTRQ-GFVE	VLAAQQSP--	VNKAWFQGTA	DAVRQYLWLF	EESE----CE	EYLILSGDHL	YRMDYRPFIM	KHRETEAAT	VAALPCDEKR	160
153	96	IMYGGNGFVE	VLAATQTPGQ	GGKEWFQGTA	DAVRQYSWLF	NDV-KNKDVE	DIVILAGDHL	YRMDYMKFVE	AHRESNADIT	VGTLPIDEAR	184
154	160	VNFGD-GIVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	VSCVPMDDSR	247
155	171	ITFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	HLLILSGDHL	YRMDYMDLVQ	RHVDNADIT	VSCVPMDDSR	258
156	160	VNFGD-GIVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	VSCVPMDDSR	247
157	172	MTFGD-GFVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	HLLILSGDHL	YRMDYMDLVQ	RHVDNADIT	VSCVPMDDSR	259
158	171	VNFGD-GCVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KHANIIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	VSCVPMDDSR	258
159	169	VNFGD-GCVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KHANIIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	VSCVPMDDSR	256
160	164	VNFGD-GFVE	VLAATLTPGE	AGKWFQGTA	DAVRRFWVF	EDA-KNKNIE	HLLILSGDHL	CRMDYMKLVE	KHIDGADIS	VSCVPMDES	251
161	165	VNFGD-GFVE	VLAATKTPGE	SGKWFQGTA	DAVRRFWVF	EDA-KNKNIE	NLILSGDHL	CRMDYMKLVE	KHIDGADIS	VSCVPMDES	252
162	164	VTFGD-GYVE	VLAATQTPGE	AGKWFQGTA	DAVRQFHWLF	EDP-RSKDIE	DVLILSGDHL	YRMDYMDLVQ	NHRESGADIT	LSCVPMDDSR	251
163	145	VSPGD-GFVE	VLAATQRPGL	EGKRWFQGTA	DAVRQFDWLF	DDA-KAKDIE	DVLILSGDHL	YRMDYMDLVQ	SHRQRGAGIS	ICCLPIDDSR	232
164	163	INFAD-GSVE	VLAATQMPPEE	AA-GWFQGTA	DAVRKFIWLV	EDYYSKSID	HLLILSGDQL	YRMDYMDLVQ	KHVEDNADIT	LSCVPMDES	240
165	161	GYKNE-GFVE	VLAAQQSP--	DNPWFQGTA	DAVRQYLWLF	EEHN----VM	EYLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEER	243
166	158	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYLWLF	EEHN----VL	EYLILAGDHL	YRMDYERFIQ	AHRESADIT	VAALPMDEAR	240
167	169	VNFGD-GYVE	VLAATQTPGE	AGKWFQGTA	DAVRQFTWVF	EDA-KNKNIE	NLILSGDHL	YRMDYMDLVQ	SHVDNADIT	VSCVPMDDSR	256
168	159	GYKNE-GFVE	VLAAQQSP--	ENPNWFQGTA	DAVRQYLWLF	EEHN----VL	EYLILAGDHL	YRMDYKFKFIQ	AHRETNADIT	VAALPMDEKR	241
1	168	ATG-FGVMHV	NEKDEIIDFI	EKPAD-----	-----	--PPGIPGNE	GFALASMGIV	VFHTKFLMEA	LRRDADPTS	SRDFGKDIIP	239
2	175	ASA-FGVMHV	DENDKIIIEFV	EKPAN-----	-----	--PPSMNRAE	SKSLASMGIV	VFDADYLYEL	LEEDDRDENS	SHDFGKDLIP	246
3	178	ATA-FGLMKI	DEEGRIIEFA	EKPQ-GEQLQ	AMKVDTTILG	LDD---KRAE	MPFIASMGIV	VISKDVLMLN	LRRKF---PG	ANDFGKEVIP	260
4	199	ASD-FGLVKI	DSRGRVVFQA	EKPK-GFDLK	AMQVDTTILG	LSP--QDAK	SPYIASMGIV	VFKTDVLLKL	LKWSY---PT	SNDFGSEIIP	281
5	172	ASA-FGVMHV	NEERQVLAFN	EKPKD-----	-----	--PTMFGNPN	DRALVSMGIV	VFDRLYLFQL	LREDAENFDS	SRDFGKDVIP	243
6	171	ASA-FGVMHV	DGDNRVRFQ	EKPAD-----	-----	--PPSIPGNS	DRALVSMGIV	IFNRAFVFNQ	LIADA-RKES	DHDFGKDIIP	241
7	161	TSS-FGIIRT	DSAGRIRDFR	EKPES-----	-----	--AEMFGRSP	GHALASMGIV	VFSADLILVA	LRRHS--RG	GHDGKGLLIP	230
8	206	AKR-MGILKV	DEQNSITSFY	EKPQDNDLLQ	QLRSPSNILE	KAGV-APTGE	RVYLGSMGIV	LFRKRALVEL	LSE----DI	REDFGKHLIP	288
9	177	ASR-MGVLQI	DEANLLDFY	EKPEEEIILN	RFRSSQECR	KIKL-DEQYG	N-FLGNMGIV	LFRRESLFL	LQE----EQ	GDDFGKHLIQ	258
10	176	AYR-MGVLDI	DSEGLLIDFY	EKPEEKEVLK	RFQLSSEDRR	IKHL-TDQSG	D-FLGSMGIV	LFRRESLFL	LRE----EE	GDDFGKHLIQ	257
11	186	ARE-FGVMGV	DEQDRVIDFV	EKPN-----	-----	--PPAIPGQP	DRALVSMGIV	IFNTRKFLFQ	LERDAMTKGS	NRDFGKDIIP	257
12	173	ASA-FGVMGV	DENDRVVDFE	EKPAH-----	-----	--PSSLPDDP	DHALASMGIV	VFNAAFVLEQ	LIRDADDPKS	SHDFGHDIIIP	244

13 178 ASA-FGVMAV DDSWRITTSFA EKPEH----- --PAPIPKP GHALISMGY VFNAKFLYEQ LIQDHDMDQS SHDFGKDVIP 249  
14 170 ASA-FGVMAV DNDSRITNFT EKPAH----- --PAPIPKP GHALISMGY VFNAKFLYEQ LIIDHDMNQS SHDFGKDVIP 241  
15 171 ASA-FGVMSV TKDGRVTDFT EKPSV----- --PTAVPGRP GYALVSMGY VFNADFLFDQ LIRDHDDPNS SHDFGKDLIP 242  
16 158 ASD-FGLMKI DNSGRVIDFS EKPK-GEALT KMRVDTTVLG LTP--EQAAS QPYIASMGY VFVKDVLKIL LKE-A---LE RTDFGKEIIP 239  
17 169 AVA-FGVMTV DENRRVTGFV EKPAD----- --PPAIPGRP DTALASMGY VFSADYLYSL LEENISSVDT DHDGKDLIP 240  
18 168 ASA-FGVMTV DAMRHITRFD EKPAH----- --PQMLDQP EQALVSMGY VFDADYLFPA LQTDIEDAAS HHDGKDLIP 239  
19 168 APE-LGLMKI DAQGRITDFS EKPO-GEALR AMQVDTSVLG LSA--EKAKL NPYIASMGY VFKEVHLNL LEK-Y---EG ATDFGKEIIP 249  
20 183 AAA-FGVMSV DESLRITFEN EKPRE----- --PDSMPGKP GTALVSMGY VFSKDFLYKA LIEDAGATRS SHDFGKDIIP 254  
21 150 ASS-FGIJET DAQNRIIGFD EKPKA----- --AKFMPDDP DHAYASMGY LFNADVLEA LEEAHR--RG ETDGFRHVLIP 219  
22 184 ASA-FGVMSV SDQWQVSSFA EKPAH----- --FVPIPGQP EKALVSMGY VFNADFLYEQ LIRDHQDHDS SHDFGKDIIP 255  
23 164 ASA-FGVIGS DDRFRIVDWE EKPKN----- --PKFVPGNP DKAFVSMGY IFNTDMLVKS VIADAKNSAS SHDFGKDVIP 235  
24 160 ASR-FGIMNT DSDRIVEFE EKPEQ----- -----P KSTKASMGY IFNWDRLRTRM LVDAEKNNDI MSDFGKNVIP 224  
25 166 ATG-LGVIRT DRRGFIDDFV EKPPLRQNI EYMRVHPDLLP SNH--LQNER RYVLSMGY FFNADALETA LDN-----S FTDGKRIIP 246  
26 175 GSE-FGIMEV TADYQITAFY EKPAN----- --PPPPIGDP SNALASMGY IFNADYLFKL LEEDNNTPGS SHDFGKDIIP 246  
27 170 AAGSFGVMSV DENFRINGFA EKPEH----- --PAPLPDGD TRCLASMGY VFDFTEFLFEQ LRRDAETSQS QRDGKDIIP 242  
28 185 ASS-FGVMSV NDAWQVTSFA EKPDN----- --PVEIPGQP EKALVSMGY VFNAAFVLDQ LVRDHNADHS SHDFGKDIIP 256  
29 182 ASR-MGILKV BQDFQIKDFC EKPKTQEELD PFYLPN----- --AEG KNYLGSMGY LFKREVLFDL LLT-----DS REDFGKHLIP 254  
30 171 ATA-FGVMTV DENRRILDV EKPAD----- --PPCLPGKP DRALASMGY VFNARVLYRE LEEDMADPNS SHDFGKDIIP 242  
31 168 AANQLGVLSA DSKGRVTAFD EKPK----- --PHPLLDAP EYCLASMGY VVNAEVLRYL LKADAKALE DHDGKKNVIP 240  
32 161 ASR-FGIMNT DNDLSIYFE EKPRH----- -----P KSTKASMGY VFDWQLLKEA LIEDANNPES VHDGKKNVIP 225  
33 161 ATG-LGIIGC DKGQVYKFF EKPAIDEDIS DYRVEQVMM QGLGTVNAS NEYLSMGY IFNTKSMEEV LKN-----D KTDGFRVIVIP 243  
34 158 ASG-FGIMKI DKKYRITAFM EKPAPELAID DWKIPADAHA D-----IPEG KDYLASMGY IFNAEAMESA LDN-----D FTDGKRIIP 235  
35 176 ARR-MGLLRI NIEGKVIFDY EKPDQDELLN RRLTPDVRK QHNL-LESEG E-FLGSMGY MFKEKSLRFL LAE---EE GEDFGKHLI 257  
36 161 STG-FGILHV DEDNRIINFL EKPKD----- --PPGLPGNP DVSLANMGY IFKTEVLVQE VIRDARLPES DHDGKNIIP 232  
37 161 CNO-FGIVET DDNHRIVDFV EKPO----- --PRMPGSS THALASMGY LFNADILLDA LREAHE--TG HSDFGKHILP 231  
38 161 VSS-FGIIRT DSAGRIRDFQ EKPO----- --AEPMPNRP GVALASMGY VFSTDVLDVA LTRAHS--KG GHDGKNLIP 230  
39 171 ATA-FGVMAI DENRRVTRFD EKPAQ----- --PAPIPGRA DRALASMGY VFNDRFLFR LGADA-RTSS EHDGKDIIP 241  
40 172 ARA-FGVMSV QEDGRVTALT EKPO----- --PEMPGHD DVALVSMGY VFNDRYLLQV LREDAENFAS SRDFGRDVLP 243  
41 171 ASA-FGVMAV DKHSMVTEFL EKPAD----- --PPAMPGRD DMSLASMGY IFNAEVLRYE LARDMADPDS SHDFGKDIIP 242  
42 170 AADSFGVMTV NADNKVIAFD EKPAQ----- --PNEIPGQP GQCLASMGY LFNTEFLFPH LLNDYSSENS SRDFGKDIIP 242  
43 159 AYH-FGIIEV DDEGRMIGFA EKPAVED--- --AKTIPGDP DHVLSMGY IFESKVLLE LYEDAANSTS QHDGNDIIP 232  
44 157 AYH-FGIIEV DAEGRMIGFD EKPAVED--- --AKTIPGDP DHVLSMGY VFDSKALQAE LIRDAAVGDS SHDFGKDIIP 230  
45 168 ASG-FGVMTV DANDQILSFT EKPAH----- --PPAMPDKP DMALASMGY VFEKFLFEE LRRDAADKNS SHDFGKDIIP 239  
46 172 AHA-FGVMTV DKDHRIVKFT EKPAN----- --PEMPGKP DKALASMGY VFSTKVLYQ LMKDRDNPNS SHDFGKDIIP 243  
47 159 AKA-FGVVVT NADLKIIAFE EKPO----- --PTPLPENP EKALVSMGY VFSTKLLSRA LVEDQHNDAS SHDFGKDIIP 230  
48 169 AVA-FGVMTV DEERRVTGFV EKPAD----- --PPAMPGHP DIALASMGY VFNADYLYSL LEONITSVAT DHDGKDIIP 240  
49 151 ARA-FGIVIC DESGRIIAFH EKQV--D--- --PPSMPGRP GMCLASMGY IFKTKALLDV LEHDAATEDS AHDGRIIP 222  
50 172 AHA-FGVMAI DARRHITAFV EKPAD----- --PPALPGNP GLSLASMGY IFSANYLYRL LEDDAKNPDS SHDFGKDIIP 243  
51 159 AEG-FGLMRT DDVGNIKFES EKPS-GEKLK AMAVDTSKFG LSK--ESAAE KPYLASMGY VFSRNTLFDL LNK-F--PN YTDGKDIIP 240  
52 168 ATA-FGVMTV DASLRITDFL EKPAD----- --PPGIPGDE GNALASMGY VFDWAFRLDL LIRDAEDPNS SHDFGHDLIP 239  
53 171 AKA-FGVMSV DNDLNRVIEFI EKPEH----- --PKPSPGRS GETLASMGY IFNASFVLEQ LIKNADTSSS SHDFGKDIIP 242  
54 164 APN-FGILAT DADGRIRQFQ EKPO----- --PKSMPSDS NRAYASMGY LFDTRVLEA LMESH--LG ETDGHHVLP 233  
55 168 ASA-FGCMMA DKTGRITQFI EKPAN----- --PPGLPDDP THSLVSMGY VFDWVFLREL LIRDAEDPNS SHDFGHDLIP 239  
56 181 ASA-FGVMSI DASRKIVFTI EKPAD----- --PPAMPGND QMSLASMGY IFNASALYRM LDEDMADPAS SHDFGKDIIP 252  
57 170 ASG-FGVMAI DENGLITDFI EKPAD----- --PPAMPGND DMAMCSMGY VFNAKVLYE LARDITDPTS SHDFGKDIIP 241  
58 170 AHA-FGVIQV DENWRMVGQF EKPT--N--- --PVEIPGRP GWVLSMGY IFNPEVLHDA LGRDANDEGS AHDGKNIMP 241  
59 172 AKA-FGVMTV DDEDRVIDFS EKPDN----- --KPLPDNP DQVLSMGY VFNASFVLEQ LIRDADAPHS QHDGRIIP 243  
60 172 SAS-FGVMTV GDKDRIVKFT EKPPV----- --GDEIPGKP GRILASMGY VFNAKFMYEQ LIRDADTKKS SHDFGKDIIP 243  
61 160 AKG-FGVMLV NKRQITNFM EKPKAEELD SLKLSQK MFN--IEDPN KEYLASMGY VFRNRVLKEI LEDV----S MMDGKDIIP 241  
62 157 AQO-FGVIEV DENGKMGFE EKPK-QN--- --PKTIPNRP THVLSMGY VFDADTLVNY LQIDAEDDNS KHDGHSILP 229  
63 165 APE-LGILKR SPDGEITSFI EKPD----- --PESLHD LES--SPGSE KPFMASMGY VFSTDLLABEL LAT-----P GDDFGKDIIP 235  
64 182 ATA-FGVMAI DDGRQITAFI EKPAD----- --PPAMPGHP DVALASMGY VFDSEVLYQL LEEDAANPDS SHDFGKDIIP 253  
65 157 ASQ-FGILIT DEEMQINDFQ EKPDN----- -----P SSNLASMGY VFTKDVLVGK LEEFCQEN--SDFGHHIIP 219  
66 184 ASA-YGVLSV DENSRVVAFS EKPSR----- --PAELPDNP GSSLVNMGY IFNSRFLFDH LLRDGEDLHS CHDFGKNLIP 255  
67 186 AFA-FGVMTV DENSRVVEFV EKPAN----- --PPSIPDNP EKSLASMGY VFNTQFLIEQ LIRDADSPNS SHDFGKDIIP 257  
68 172 AKG-FGVLAV DGTDRVIEFA EKPAN----- --PKHMPGDT TKAFASMGY VFNAKFLYEQ LIRDAGDPKS THDFGGDIIP 243  
69 171 AKA-FGVMSV DEDFRVTEFM EKPEH----- --PQSPGRS DETLASMGY VFNAAFVLEQ LIKNADTSSS SHDFGKDIIP 242  
70 177 ARS-FGVMSV NHEDRVVAFT EKPAE----- --FVPIPGQS DRALASMGY VFNTEFLYEQ LIRDADDPQS SHDFGHDLIP 248  
71 170 AADTFGVTM DEESRVCRFD EKPA----- --PSSVPGKP GTCCLASMGY IFNTEFLFEQ LKKAENEGS GRDFGDIIP 242  
72 160 ASR-FGIMKT DADGTITHFD EKPKF----- -----P KSNLASMGY IFNWPLKQY LEMDRNPNYS SHDFGKDIIP 224  
73 169 ASA-FGVMAV DAVRQITSFV EKPV----- --PPMPDRP DQSLASMGY IFNAEVLRYL LNEEDADPNS SHDFGKDIIP 240  
74 174 ATA-FGVMAI NEKLVKAFV EKPSD----- --PPMPDRP GSSLASMGY VFDADYLYEV LEREATSVDT SHDFGNDIIP 245  
75 177 VSR-FGVLVV DDESKLIDFY EKPOSEELK HFRLSNTAMK KFGL-DPQHG N-FLGSMGY LFRKDCFLQL LLE----ET GDDFGKELI 258  
76 168 ATG-FGVMSV NNDLRVTRFT EKPAD----- --PEAIPGKP DKALASMGY IFSQFLFDK LIEDHDDPHS SKDFGKDIIP 239  
77 189 ATG-FGVMQV DSDSRVVKFA EKPKN----- --PEGMPGRP DTALASMGY IFDAAYLLEL LTRDAGATMS SHDFGHDIIP 260  
78 172 AKA-FGVMAI DENKVTSEFV EKPAD----- --PPAMPGKP DRSLASMGY IFTDYLYRM LDEIDALEGS SHDFGKDIIP 243  
79 159 ASS-FGIMSV DDTQRIRAFD EKPKH----- --PKFMPDDP HRALASMGY IFNMDLLIHE LQADHCLTAS NHDGKDIIP 230  
80 175 ATA-FGVMDV DSNRRVKAFF EKPEH----- --PMLMPGRD DTALASMGY IFNAEFLFEQ LLKDADTKGS TRDFGKDIIP 246  
81 158 ATN-FGVMAI DENQRIVEFT EKPAQ----- --PTPLNDP EKSLASMGY IFSTDALVDA LEQDANPDS NHDGNDIIP 229  
82 170 AAGQFGVMTV DQDNRVKRFD EKPAQ----- --PNEIPGKP GQCLASMGY VFNTEFLFDQ LEKDATRITS DRDFGNDIIP 242  
83 160 ATA-FGCIDA DDSGRIRSFV EKPLE----- --PPGTPDDP DTFVSMGY IFTKVLIDA IRADADDDHS DHDGMDIIP 231



84	174	ASE-FGVMAV	NENLKVQDFV	EKPKD	-----	-----	PPPMVQKP	DVSLASMGY	VFDADYLYEM	LNKEVNTPYT	SHDFGKDVLP	245
85	170	ARG-FGVMSV	DKDQRVIGFD	EKPAN	-----	-----	PSQPQGP	DKALASMGY	VFNTEFLYEQ	LEKDAGESSS	AHDFGHNIIIP	241
86	172	ARA-FGVMTV	DENGRVLRFT	EKPOE	-----	-----	PNFVQPK	DTALVSMGY	VFEREYLFQ	LRADAENIDS	SRDFGRDVIIP	243
87	166	ASG-FGVMDV	DENRKITRFL	EKPAD	-----	-----	PPGMPDN	DKSLASMGY	VFNADYLYRL	LDEDCGDNTS	SHDFGKDIIIP	237
88	178	ASE-FGVMAV	NENLKVKSFV	EKPKD	-----	-----	PPAMVQKP	NTSLASMGY	VFDADYLYDV	LEREVTSPYT	SHDFGKDVLP	249
89	175	AFQ-FGIMDI	DGNRRVLNLF	EKPSN	-----	-----	PPCIPND	EHSLASMGY	VVDRDYLFDL	LEEDSRDPNS	HHD FGQDIIIP	246
90	161	APR-FGIMNT	DESGRIVEFE	EKPAK	-----	-----	-----	KSNLASMGY	IFNRDYLAEY	LTADARSETS	SHDFGKDIIIP	225
91	175	AKE-FGIMSV	NDKWQVQAFV	EKPOD	-----	-----	PPTMREK	DTSMASMGY	VFDSNYLYDV	LEDEIQKQTK	NLDFGKHIMP	246
92	171	ARE-FGVMSV	DEGRVAVAFN	EKPEH	-----	-----	PQSTPQNP	DMALASMGY	VFNAEFLYEQ	LARDADDPNS	SHDFAKDIIIP	242
93	187	GRE-FGIIGV	DEEGRVITYFH	EKPEH	-----	-----	PAAMPGR	DSALASMGY	VFGARFLYEQ	LIRDHDDPES	GHD FGKDLIIP	258
94	157	ASA-FGIIDA	DAQQIKGFL	EKPKN	-----	-----	PPPISDP	SRAYGSMGY	LFNTDVLKKA	LYEAKE-RG	EHD FGKNILP	226
95	170	ARA-FGVMAV	DGDGRVTDPL	EKPPD	-----	-----	PEMPKGP	GTSLASMGY	VFNATFLFER	LIRDADDSRS	SHDFGKDIIP	241
96	160	ARG-FGIISA	APDGRVQAFQ	EKPAN	-----	-----	PTPIGDP	ERAFASMGY	VFRADVLMAA	LEEAHR--NG	ETD FGGHILP	229
97	157	ASQ-FGVIEV	DENKGMVGF	EKPS-N	-----	-----	PKSIPGP	EWALVSMGY	IFEAETLSKE	LRDAENNGS	SHDFGKDIIIP	228
98	158	ASA-FGVMAI	DDSRITCFV	EKPAD	-----	-----	PPCIPNRP	DHSLASMGY	IFNMDVLKKA	LTEDAEIEQS	SHDFGKDVIIIP	229
99	158	ASE-FGVIEV	DAEGRMIGFE	EKPA-N	-----	-----	PKPIGDP	DSALVSMGY	VFEANELFAE	LVEDADREGS	SHDFGKDIIP	228
100	169	AKE-FGVMEI	DESLQINNFT	EKPRY	-----	-----	PACVQGR	TRSMASMGY	IFDKVLTQA	LLADAEDPDS	SHDFGKDIIP	240
101	161	ASR-YGIMNC	HENKGIYEF	EKPKN	-----	-----	-----	KSTLASMGY	IFTWSTLREY	LIRDNECSDS	VND FGKNIIP	225
102	168	ATG-FGVMAV	NEKDEIIDFI	EKPAD	-----	-----	PPGIPGP	GFALASMGY	VFHTKFLMEA	LRRDADDPDS	SRDFGKDIIP	239
103	165	ARG-FGVFCV	DDDNRVITAFE	EKPAN	-----	-----	PVPIGR-	ETCFASMGY	IFSTRRLIEY	LOEGKK-LHA	LDL FGKVIIP	234
104	158	ASD-FGLIKT	DADGRVVQFT	EKPK-GAELE	RMRVDTTTLG	LTL--EEAER	RPFVSMGY	VFRHDVMLKL	LRD-D---PS	RTDFGKEIIP	239	
105	173	ATA-YGCVVE	DNDNIVVHFL	EKPAN	-----	-----	PPGIPGP	DRAFASMGY	IFNADFLYEI	LESDALNEAS	QHD FGDIIP	244
106	247	ATA-FGLMKA	DEEGRIIEFA	EKPK-GHEHL	AMQVDTTTLG	LDD--QRAKE	MPFIASMGY	VVSRDVMLDL	LRQF---PG	ANDFGSEIIP	329	
107	166	ASG-FGLLKV	DGTGRVTDPR	EKPT-GDALR	DMRVDTTRYG	LTI--EEAHR	KPYIASMGY	VFKRQVLIDL	LQQ-M---AD	ATDFGKEIIP	247	
108	159	AEG-FGLMRT	DNDGNIREFK	EKPS-GEALK	AMAVDTSRFG	LSP--DSAKE	RPYLASMGY	VFSRSTLFDL	LNK-Y---PS	YKDFGKEVIIP	240	
109	158	ASD-FGLIKT	DADGRVVQFT	EKPK-GAELE	RMRVDTTTLG	LTL--EEAER	RPFVSMGY	VFRHDVMLKL	LRD-D---PS	RTDFGKEIIP	239	
110	161	ASS-FGLMKA	NEHATILIEFS	EKPK-GDALK	AMQVDTTTLG	LDA--ERAKE	MPYIASMGY	VFNAKAMEQV	LQDDF---PE	ANDFGSEIIP	243	
111	203	ASD-FGLMKA	DSSGRIVEFT	EKPK-GDALQ	AMQVDTTTLG	LTA--AEAEA	KPFIASMGY	VFKKSMVLKF	LDDY---PE	DND FGGEIIP	285	
112	242	ASA-FGLMKA	DEEGRVIEFA	EKPK-GEALT	KMRVDTGILG	VDP--ATAAA	KPYIASMGY	VMSAKALREL	LLNRM---PG	ANDFGSEVIIP	324	
113	232	AKE-FGLMKA	DKRRVTSFAS	EKPKTQEALD	AMQVDTTTLG	LTP--EPADE	KPYIASMGY	VFKKSVLLQL	LNDSY---AK	ANDFGSEIIP	315	
114	234	ASD-FGLMKA	DDTGRVISFS	EKPK-GDELK	AMQVDTTTLG	LSK--EEAEN	KPYIASMGY	IFKDKILLNL	LRWRF---PT	ANDFGSEIIP	316	
115	202	ATA-FGLMKA	DEEGRVIEFA	EKPK-GEQLK	AMQVDTTTLG	LDD--VRAKE	MPYIASMGY	VFSKDVMLQL	LRQF---PE	ANDFGSEVIIP	284	
116	237	ATA-FGLMKA	DDEGRVIEFA	EKPK-GEKLR	SMQVDTTTLG	LDP--ERAKE	LPYIASMGY	VFSKDVMLRL	LRNF---PA	ANDFGSEVIIP	319	
117	243	ASD-FGLMKA	DRNGHITDFL	EKPK-GADLE	SMQVDMGLFG	LSP--EFAST	YKMASMGY	VFKADVLRKL	LRGHY---PT	ANDFGSEVIIP	325	
118	245	ASK-NGLVKI	DHTGRVLQFF	EKPK-GADLN	SMRVDTNFLS	YAI--DDAQK	YPYLASMGY	VFKKDALLDL	LKSKY---TQ	LHDFGSEIIP	327	
119	257	ASD-YGLMKA	DSTGRVIEFA	EKPK-GTDLK	AMQVDTTTLG	LSK--QEAMQ	FPYIASMGY	VFRDVLVLLK	LRCSY---PS	CND FGSEIIP	339	
120	257	ASD-FGLMKA	DNKGRVLSFS	EKPK-GVDLK	AMEVDTTTLG	LSK--EALAK	KPYIASMGY	VFKKEILLNL	LRWRF---PT	ANDFGSEIIP	339	
121	256	ASD-YGLVKI	DGRQVVFQFA	EKPK-GSELR	SMRVDTSTLG	LSP--QDAMK	SPYIASMGY	VFKTDILLKL	LRWRY---PT	ANDFGSEIIP	338	
122	249	ATA-FGLMKA	DEEGRVIEFA	EKPK-GEQLK	AMQVDTTTLG	LDD--ERAKE	MPYIASMGY	VVSKNVMLDL	LRQF---PG	ANDFGSEVIIP	331	
123	256	ASD-YGLLKM	DNRGRIIQFA	EKPK-GADLK	AMQVDTTTLG	LSP--QEAMK	SPYIASMGY	VFKTDILLNL	LRWRY---PT	SND FGSEIIP	338	
124	249	ASD-FGLMKA	DNRGRVLSFS	EKPK-GEDLK	AMEVDTTTLG	LSR--EEAEK	KPYIASMGY	VFKKEILLNL	LRWRF---PT	ANDFGSEIIP	331	
125	236	ATA-FGLMKA	DEEGRVIEFA	EKPK-GEQLK	AMQVDTTTLG	LDD--ERAKE	MPYIASMGY	VVSKDVMLDL	LRQF---PG	ANDFGSEVIIP	318	
126	253	ASD-YGLMKA	DNTGRVIEFA	EKPK-GPNLK	AMQVDTTTLG	LSE--KEAEK	CPYIASMGY	VFRDVLVLLK	LTRKY---LS	CND FGSEIIP	335	
127	236	ASD-FGLMKA	DDTGRVISFS	EKPK-GDELK	AMQVDTTTLG	LSK--EEAEK	KPYIASMGY	IFKDKILLNL	LRWRF---PT	ANDFGSEIIP	318	
128	237	ATA-FGLMKA	DDEGRVIEFA	EKPK-GEKLR	SMQVDTTTLG	LDP--ERAKE	LPYIASMGY	VFSKDVMLRL	LRNF---PA	ANDFGSEVIIP	319	
129	246	ASN-NGLVKC	DHTGRVLQFF	EKPK-GADLN	SMRVDTNFLS	YAI--GDAQK	YPYIASMGY	VFKKDALLDL	LKSKY---TQ	LHDFGSEIIP	328	
130	248	ASD-YGLVKF	DSSGRVIEFA	EKPK-GAAL	EMKVDTSFLN	FAI--DDPTK	YPYIASMGY	VFKRDVLLDL	LKSRY---AE	LHDFGSEIIP	330	
131	194	ASD-FGLMKA	DDTGRVIEFA	EKPN-GDALK	AMEVDTTTLG	LTA--EATAS	SPYIASMGY	VFKKALLNL	LNAEY---PK	NDFGSEIIP	276	
132	229	AAA-FGLMKA	DDTGRVIEFA	EKPT-GDALK	AMQVDTTTLG	LDA--ERAKE	MPYIASMGY	VFNARAMEKL	LMEDF---PT	CHDFGGEIIP	311	
133	252	ASD-YGLMKA	DNTGRVIEFA	EKPK-GDLK	AMQVDTTTLG	LSK--QDALQ	YPYIASMGY	VFRTEVLCKL	LRWSY---PS	CID FGSEVIIP	334	
134	260	ASD-YGLVKI	DSRGRVIEFA	EKPK-GAELK	SLKADTTQLG	LSP--QDALK	SPYIASMGY	VFRTEILLKL	LRWRF---PT	SND FGSEIIP	342	
135	257	ASD-FGLIKI	DETGRIQFL	EKPK-GESLK	SMRVDTSTLG	LSI--SDARK	LPYIASMGY	MFKTDVLLKL	LRWHY---PT	ANDFGSEIIP	339	
136	262	ASD-FGLMNI	DNKGRVLSFS	EKPK-GADLK	AMAVDTTTLG	LSK--EEAEK	KPYIASMGY	VFKKEILLNL	LRWRF---PT	ANDFGSEIIP	344	
137	248	ATA-FGLMKA	DEEGRVIEFA	EKPK-GEQLK	AMQVDTTTLG	LDD--ERAKE	MPYIASMGY	VVSKNVMLDL	LRQF---PG	ANDFGSEVIIP	330	
138	251	ASD-FGLMKA	DNKGRVLSFS	EKPK-GDELK	AMAVDTTTLG	LSK--EEAEK	KPYIASMGY	VFKKEILLNL	LRWRF---PT	ANDFGSEIIP	333	
139	249	ASN-FGLVKI	DRGGRVIEFA	EKPT-GVDLK	SMQVDTTTLG	LSH--QEATD	SPYIASMGY	CFKTEALLNL	LTRQY---PS	SND FGSEVIIP	331	
140	192	ASD-FGLMKA	DANGQILYFS	EKPK-GADLK	AMQVDTTTLG	LTP--EEAIE	KPYIASMGY	VFKKDILLKL	LRWRY---PT	ANDFGSEIIP	274	
141	189	ASD-FGLMKA	DANGQILYFS	EKPK-GADLK	AMQVDTTTLG	LTP--EEAIE	KPYIASMGY	VFKKDILLKL	LRWRY---PT	ANDFGSEIIP	271	
142	262	ASD-FGLVKT	DARGRIISFS	EKPK-GMDLK	AMQVDTTTLG	LSR--EEAKK	MPYIASMGY	VFRKDVLKLL	LRWRY---PT	SND FGSEIIP	344	
143	227	ASD-FGLVKT	DARGRIISFS	EKPK-GMDLK	AMQVDTTTLG	LSR--EEAKK	MPYIASMGY	VFRKDVLKLL	LRWRY---PT	SND FGSEIIP	309	
144	180	ATN-FGLMKA	DSEGRVIEFA	EKPK-GGILQ	GKQVDTTTLG	LDP--KRAEA	LPYIASMGY	VISKEAMYKL	LHEKF---PN	ANDFGSEIIP	262	
145	186	ATA-FGLMKA	DDEGRVIEFA	EKPK-GSALK	AMEVDTTTLG	LDP--ERAKE	MPYIASMGY	VVSKDVMSRL	LRDEF---PN	CND FGSEVIIP	268	
146	161	ATN-FGLMKA	DSEGRVIEFA	EKPK-GGILQ	AMQVDTTTLG	LDP--KRAEA	LPYIASMGY	VISKEAMYKL	LHEKF---PN	ANDFGSEIIP	243	
147	176	ASD-FGLMKA	DANGQILYFS	EKPK-GADLK	AMQVDTTTLG	LTP--EEAIE	KPYIASMGY	VFKKDILLKL	LRWRY---PT	ANDFGSEIIP	258	
148	240	AKE-FGLMKA	DDKRRVLSFA	EKPKTQEALD	AMQVDTTTLG	LTP--DEAAD	KPYIASMGY	VFKKSVLCKL	LNETY---AK	ANDFGGEIIP	323	
149	240	ASA-FGLMKA	DDAGRVVVEFA	EKPK-GEALQ	RKQVDTTTLG	VDP--ATAQS	KPFIASMGY	VMSAKALREL	LLNRM---PG	ANDFGSEVIIP	322	
150	231	AEA-FGLMKA	DDSGRIIDFA	EKPK-GKELE	AMAVDTTTLG	LDK--KLAKE	MPYIASMGY	VFKASAMDEL	LTEKF---PD	CHDFGGEIIP	313	
151	230	ASD-FGLMKA	DKTGRVIEFA	EKPE-GNDLL	AMQVDTTTLG	LSP--EESQA	SPYIASMGY	VFKKSAISFL	LNSEY---PK	DND FGGEIIP	312	
152	161	ASS-FGLMKA	DDEGRVIEFA	EKPK-GAELQ	AMQVDTTTLG	LDA--DKAQE	MPFIASMGY	VFDKAKMRE	LLNFN---KE	ADD FGGEIIP	243	
153	185	ASD-FGLMKA	DSTGRVIEFA	EKPK-GDALQ	AMQVDTTTLG	LTA--DEAKE	KPFIASMGY	VFKKSAVLFK	LEKDY---PE	DND FGGEIIP	267	
154	248	ASD-YGLVKV	DDRGRIIQFS	EKPN-GDDLK	AMQVDTTTLG	LSP--QDALK	SPYIASMGY	VFKTDVLLNL	LKRWY---PT	SND FGSEIIP	330	

155	259	ASD-YGLMKI	DKTGRIIQFA	EKPK-GSDLK	AMRVDTTLLG	LSP--QEAEK	YPYIASMGVY	VFRTEPLLQL	LRWNG---SS	CNDFGSEIIP	341
156	248	ASD-YGLVKV	DDRGRIIQFS	EKPK-GDDLK	AMQADTSLG	LSS--QDALE	SPYIASMGVY	VFKTDVLLNL	LKWRY---PT	SNDFGSEIIP	330
157	260	ASD-YGLMKI	DKTGRIIQFA	EKPK-GSDLK	AMRVDTTLLG	LFP--QEAEK	HPYIASMGVY	VFRTEPLLQL	LRWKC---SS	CNDFGSEIIP	342
158	259	ASD-YGLVKA	DARGRIIQFS	EKPN-GADLK	AMQVDTSVLG	LPL--HEAKR	SPYIASMGVY	VFKTDVLLRL	LKWRY---PT	SNDFGSEIIP	341
159	257	ASD-YGLVKA	DARGRIIQFS	EKPK-GADLK	AMQVDTSVLG	LPP--HEAKR	SPYIASMGVY	VFKTDVLLKL	LKWRY---PT	SNDFGSEIIP	339
160	252	ASD-YELMKI	DRKGITQFV	EKPE-GSDLK	AMHVDTTLLG	LTA--EEAQT	YPYIAPMGVS	VFRTEPLLKL	LRWSC---PS	CNDFGSEIIP	334
161	253	ASD-YELMKI	DRKGITQFV	EKPE-GSDLQ	AMHVDTTLLG	LTA--EEAQT	YPYIAPMGVS	VFRTEPLLKL	LRWSC---PS	CNDFGSEIIP	335
162	252	ASD-FGLMRI	DNKGRILSFS	EKPK-GEEKL	AMQVDTTVLG	LSK--DEAOK	KPYIASMGVY	VFKKEILLNL	LRWRF---PT	ANDFGSEVIP	334
163	233	ASD-FGLMKI	DDTGRVISFS	EKPK-GDDLK	AMQVDTTVLG	LSK--EEAEE	KPYIASMGVY	IFKKEILLNL	LRWRF---PT	ANDFGSEIIP	315
164	251	ASE-YGLVKF	DSSGRVIQFS	EKPK-GVDLE	AMKVDTSFLN	FAI--DDPAK	FPYIASMGVY	VFKRDVLLNL	LKSRY---AE	LHDFGSEILP	333
165	244	ATA-FGLMKI	DEEGRILEFA	EKPK-GEOLK	AMMVDTTILG	LDD--VRAKE	MPYIASMGYI	VISKHVMLQL	LREQF---PG	ANDFGSEVIP	326
166	241	ATA-FGLMKI	DEEGRILEFA	EKPK-GEOLK	AMKVDTTILG	LDD--ERAKE	MPYIASMGYI	VVSKHVMLDL	LRDKF---PG	ANDFGSEVIP	323
167	257	ASD-YGLVKV	DSGGRIIQFS	EKPK-GADLK	SMQVDTSLG	LSN--QDALR	SPYIASMGVY	VFKTDVLLKL	LKWRY---PT	SNDFGSEIIP	339
168	242	ATA-FGLMKI	DEEGRILEFS	EKPK-GDQLQ	AMKVDTTILG	LDD--ERAKE	MPYIASMGYI	VISKHVMLDL	LREQF---PG	ANDFGSEVIP	324
1	240	YIVE-HGKAV	AHRFADSCVR	SDFEHE---P	YWRDVGTTIDA	YWQANIDLTD	VV---PDL	IYDKSWPIWT	YAEITPP---	AKFVHDEDDR	318
2	247	KITE-AGLAH	AHPFPLSCVQ	SDPDAE---P	YWRDVGTTLEA	YWKANLDLAS	VV---PELD	MYDRNWPIRT	YNESLPP---	AKFVQDRSGS	325
3	261	GATSLGMRVQ	AYLYDG----	-----	YWEDIGTIEA	FYNANLGTIK	KPV---PDFS	FYDRSAPIYT	QPRYLP---	-----S	321
4	282	AAIDD-YNVQ	AYIFKD----	-----	YWEDIGTIKS	FYNASALTO	EF---PEFQ	FYDPKTPFYT	SPRFLP---	-----T	340
5	244	NATA-NHKVQ	AYPFSDFVSG	QQA-----	YWRDVGTVDA	FYQANMELIG	ED---PELN	LYDEEWPIWT	YQQLP---	AKFVQGRDGR	318
6	242	SLID-QARVI	AFPPRDAATG	QQA-----	YWRDVGTTIDA	FWRTNLELVG	VN---PQLN	LYDKEWPIWT	HQEQLP---	AKFVFDDDDR	316
7	230	AMAERNRVM	AYDFTTNRMR	GIRDYE-EPA	YWRDVGTTIDA	YDANFDTLG	EF---PKFC	MSNPHWPIYA	-NPDQTE---	AAQVH	305
8	289	TKVAS-GKIS	AFLYTG----	-----	YWEDIGTIET	FYQANLALTE	TN---PVFN	FHNARPIYT	YRVDLPP---	-----A	347
9	259	VQMKR-GSVK	TFLYDG----	-----	YWDTIGTIAS	YYEANIALTO	RPHQPVRGLN	CYDDGGMIYS	KNHHLPG---	-----T	321
10	258	AQMKR-GQVQ	TLLYNG----	-----	YWADIGTIES	YYEANIALTO	KPHAERKGLN	CYDDNGMIYS	KNHHLPG---	-----A	320
11	258	YIVPR-YRVF	AHRFADSCVG	SDN---HRP	YWRDVGTTIDA	YWEANMEMTK	VT---PELN	VYDRDWPIWT	YQEQIPP---	AKFVFDEDDR	335
12	245	YLTKK-YRVF	AHRETDSCVG	AADG----NY	YWRDVGTVDA	YWEANMELTK	VV---PELN	LYDRQWPIWT	YQEQLP---	AKFVFDNEER	322
13	250	RLVASNARVY	AHRFQNSCVN	MAS---GVP	YWRDVGTVDA	YWKANIDLTT	IT---PDLN	LYDEDWPIWT	HQEQLP---	AKFVFDDDDR	328
14	242	RLVASNIQVY	AHRFQNSCVN	MDS---GVP	YWRDVGTVDA	YWEANIDLTT	IT---PDLN	LYDEDWPIWT	HQEQLP---	AKFVFDDNDR	320
15	243	HLVPR-SRVF	THRFSDSCVN	MVS---GVP	YWRDVGTVDA	YWEANLDLVQ	VI---PDLN	LYDQDWPIWT	HQEQLP---	AKFVFDDNDR	320
16	240	DAAK-DHNQV	AYLEDD----	-----	YWEDIGTIEA	FYNANLALTO	QPM---PPFS	FYDEEAPIYT	RARYLPP---	-----T	299
17	241	RVVT-QGTAI	AHPFMSCVS	SDPNVE---P	YWRDVGTTIDA	YWAANLDLAS	TI---PTLD	LYDRSWPIWT	YQEQLP---	AKFVRDMKGL	318
18	240	AIVS-RGEAM	AHPFDLSCVK	SSPESP---S	YWRDVGTVDA	YWAANIDLTA	TI---PQLD	LYDKDWPIWT	YQTPSP---	AKFVFDDDEGR	319
19	250	DSAS-DHNQV	AYLEDD----	-----	YWEDIGTIEA	FYEANLALTK	QPS---PDFS	FYNEKAPIYT	RGRYLP---	-----T	309
20	255	SSIS-RARIM	AFPPRD-REG	KPG-----	YWRDVGALNC	YWQTNMCLCS	IE---PALN	LYDCEWPIWT	YQPQYPP---	AKFIFDDEGR	328
21	219	-RLAQSHGVY	AYNFATNEVP	GTKRYE-EQA	YWRDVGTLDT	YFDAHLVDLG	LE---PRFD	AFNPQWPIYS	-SHYQGP---	TARVL	294
22	256	YLVSR-YRVF	AHRFMSCVN	MAS---GIP	YWRDVGTVDA	YWEANVDLIS	VT---PQLN	LYDEDWPIWT	HQEQLP---	AKFVFDDDDR	333
23	236	SALTSG-VF	VHSFKEANN-	-EGR-	YWRDVGTLDA	YWEANMLDCK	KD---PPVN	LYDPEWPIRT	YQEQVPP---	AKTVSTDDSE	310
24	225	AYLESGERVY	TYNFNG----	-----	YWKDVGTTIES	LWEANMEYIG	ED---NDLH	SRDRSWKIYS	KNLIAFPNEI	TE	288
25	247	QLISR-GNVH	AYIFGG----	-----	FWEDIGTIRS	FYDTSNLNAS	IN---PDFN	FYDERMPIYT	HRRDLPA---	-----S	305
26	247	QLIA-RKVWV	AHPFDLSCVT	SNAELP---P	YWRDVGTTIDA	YWRDNLGLAS	VT---PELD	MYDRAWPIRT	HMEFLP---	AKFVQDRSGS	325
27	243	SIK-DHPVY	AFEFESTGG-	GDA-----	YWRDVGTTIDS	FWEANMEMVA	FV---PQLN	LYDQWPIWT	YQEQLP---	AKFVWEDHDR	316
28	257	YLVPR-YRVF	AHRFLNSCVN	MAS---GIP	YWRDVGTVDA	YWEANLDLIS	VT---PQLN	LYDEDWPIWT	HQEQLP---	AKFVFDEDDR	334
29	255	TKVKE-GGVY	TYIHGG----	-----	YWEDIGTIGS	FYEANIALTO	VN---PHFN	CYDETYPIYT	SRSYLPG---	-----A	313
30	242	KMYR-AGVAV	AHPFELSCVG	-TRA--GTGP	YWRDVGTTIDA	YWDANIDLTA	TD---PLLN	LYDTNWPIWT	YQQLP---	AKFVHNDDDR	321
31	241	NITE-QHKVF	AHRFRGADGC	QIP-----	YWRDVGTTIDS	YWRANMDDLN	NS---DMLN	LADPSWPIWG	SALSSAP---	TQIRGGTQKG	315
32	226	AILERGERLY	AYFPKG----	-----	YWKDVGTTVES	LWEANMDLIS	EN---DDLN	IGDPTWRVYS	LAQAMPAQYI	GN	289
33	244	DTTIS-CKVA	TYLEDD----	-----	FWEDIGTIKA	FYEANLDLAS	IT---PAFN	FYDEEMPIYT	HRRHLPA---	-----T	302
34	236	MAIKK-RKVN	SYVYNG----	-----	YWEDIGTIRS	FYDANLDLTR	IN---PKFN	FYDEDMPIYT	HPRNLP---	-----S	294
35	258	AQMOK-GRVQ	AFLYDG----	-----	YWDTIGTIES	YHANIALAQ	KPHSSVKGFN	CYDARGMIYS	KNHHLPG---	-----A	320
36	233	SMVQARMRVY	SYSRFDENK-	-KEVH----	YWRDGRIDA	FYDANMDLVT	ID---PVFN	LYDPDWPIRT	YQRQCP---	AKTIFGGDPG	308
37	231	-AMLKSHRIM	AYDFTNTTIP	GTEPVE-EHG	YWRDVGTTIDA	YYQAHFDTLG	AT---PRFR	MTRNHWPIYA	-SPDQAE---	SAQIE	306
38	230	-AMCESHRLM	AYDFSSNRVP	GIRDFE-DAS	YWRDVGTTIDA	YYAAHFDTLG	EC---PAFC	MSNPRWPIYA	-APDQTE---	AAQVH	305
39	242	QLID-QARVV	AYPFRDLSTG	EQA-----	YWRDVGTTIDA	FWKTNLELID	VT---PELN	LYDREWPIWT	FQEQLP---	AKFVFDEEDR	316
40	244	AAIG-RDHVQ	AYPFSDFVSG	KQA-----	YWRDVGTVDA	FYRANQELIQ	EE---PELD	LYDDEWPIWT	YQQLP---	AKFMDQRGK	318
41	243	RAVK-NQQVA	AHPFALSCVP	-SSD--VAEP	YWRDVGTTIDA	YWEANIDLTA	TD---PELN	LYDQHWPIWT	YQQLP---	AKFVHNDDDR	321
42	243	SIK-DNNVF	SYAFKDPDSE	NQP-----	YWRDVGTLDA	FWEANMELVT	PQ---PQLD	LYDKAWPIWT	YQEQLP---	SKFIFDDDLR	317
43	232	-KLYPAGNVF	YVRLSDNFIP	G-EPAT-A-	YWRDVGTTIDS	YWEANMDMLK	PE---APFS	LYNKNWPLHT	YHPLP---	ATFR--DPEG	308
44	230	-YLYPQGVF	VYDFTTNTIP	G-EDHN-T-	YWRDVGTTIES	YWEANMDLIQ	ST---PPIS	LYNRKWPMT	YYPALP---	ANFR--NGET	306
45	240	HIVK-HGKAV	AHRFDRSCIR	SHAESA---S	YWRDVGTVDA	YWAANIDLTD	IV---PQLD	LYDHNWPIWT	YGEITPP---	AKFVHDKVGR	318
46	244	SMIK-NNRVM	AFPPRDFVSG	GDA-----	YWRDVGTVDS	LWESNLGLAS	VN---PELD	LYDEAWPIWT	HQEQVPP---	AKFVFDQDNR	318
47	231	RLVQHHK-VQ	AYKFGARGR	VTPDR----	YWRDVGTLDA	YYQANMLLKG	FF---PPLN	LYQRDWPIRT	YRQSP---	ARTINGNSGA	307
48	241	RVVT-SGNAI	AHPFMSCVS	SDPSVE---P	YWRDVGTTIDA	YWAANLDLAS	TI---PSLD	LYDRNWPIWT	HQEQLP---	AKFVRDLNGL	319
49	223	RMVQSGSRVY	VYDFHENRVE	G-EDE--GAG	YWRDIGTIDA	YWAQMDLVS	IQ---PAFN	FYNPRWPIRT	GISHDPP---	AKFVFRDEAN	302
50	244	RAVA-ENQAL	AHPFTLSAIA	TPPFSG---P	YWRDVGTVDA	YWAANLDLAS	TT---PALN	MYDKDWPIWT	YQEQLP---	AKFVHDLDR	322
51	241	EALNRGDTLK	SYLEDD----	-----	YWRDGTIGA	FYESNLALTE	QPK---PPFS	FYDEKFPYIT	RPRFLP---	-----S	301
52	240	AIVR-NGKAM	AHRFSDSCVM	TGLETE---P	YWRDVGTTIDA	FWQANIDLTD	FT---PKLD	LYDREWPIWT	YSQIVPP---	AKFIHDSERR	318
53	243	SMRNSYRNVF	AFPPRDFVGG	DPG-----	YWRDVGTVDA	FYRANLELIG	VE---PELN	LYDEDWPIWT	YQQLP---	AKFIFFNEDR	318
54	233	-RLKSHRFLV	AYDFSSNEIP	GIKPYE-EVG	YWRDVGTTIDA	YFEAKHVDLG	ES---PRFD	AFNPQWPIFS	-SNYQGP---	VARIL	308
55	240	QIVK-YGKAM	AHRFSESCVT	SGLEHE---P	YWRDVGTTIDA	FWQANIDLTE	FT---PKLD	LYDNWPIWT	YAEIVPP---	AKFIHDEDDR	318
56	253	KAVR-AGLAH	AHPFMSCVQ	GGQQSQ---P	YWRDVGTLDA	FWAANLDLAS	VT---PELN	LYDEWPIWT	SQRQLP---	AKFVQDHNGS	331

57	242	RAVA-NRVAV	AHPFSRSCIV	-APDEQFREQH	YWRDAGTIDA	YWDANIDLTA	TV---	PALN	LYDRNWPFVWT	YQQQLPP---	AKFVHNHLDLDR	322
58	241	-MLYPKSRVY	VYDFEQNRVE	G-SDEH-EHG	YWRDVGTTISA	FYEANMDLVA	VT---	PVLN	LYNRRWPIHT	WLRSRPA---	AKFVFSDDD-	320
59	244	YMIKK-YRVY	AHRETESCVG	ASDG---NY	YWRDVGTVDA	YWEANMELTK	VI---	PELN	LYDRHWPIWT	YQEQQLPP---	AKFVFDNADR	321
60	244	YLIDR-YRVY	AHRETESCVG	NQDGV---DS	YWRDVGTTIDA	YWEANMEMVS	VT---	PALD	LYDKSWPIWT	YQEQQLPP---	AKFVFDNADR	322
61	242	BAIKK-YKVF	SYAFQG----	-----	YWEDVGTIKA	YFEANISFSG	KN---	PPFD	FYDENAPIYT	HVRVYLSP---	AKFVFDNADR	300
62	229	-MMFPNGVY	VYDFSTNEIR	G-EPET-SRG	YWRDVGTTIDA	YFEANMDLIS	VT---	PSFD	LYNRYWPLRS	YAPAVPPP---	AKFI--HNOE	307
63	236	QALSN-HRVM	GHIIFDQ----	-----	YWADIGTIRR	FYEVNLELAA	NP----	IFN	LNLFPQVYV	NARFLPPP---	AKFVFDNADR	293
64	254	RAVA-QGRAL	AHPFGMSCVT	RASRGPDAKA	YWRDVGTTIDA	FWAANLDLAS	IT---	PELD	IYDTPWPIWT	YQRQLPPP---	AKFVLDREGK	335
65	220	QMIDDD-VF	AYEFNG----	-----	YWQDVGTLKS	YWEINLELTD	LV---	PEMN	LYDDNWKLLT	RSEEQPPVKF	GP	282
66	256	HMVKN-HAVF	AQNFKHSCMG	KS-----AEP	YWRDIGSIDA	YWAANIALTN	VT---	PDLN	IYDRLWPIWT	RHEQSPPP---	AKFVFDNADR	332
67	258	HMVEK-YRVF	AQSEFQSCVG	MGDD---NTP	YWRDVGTTIDS	YWEASMEMTK	VI---	PDLN	MYDQEWPIWT	YQEQQLPP---	AKFVFDNADR	336
68	244	YIIKK-YKIQ	AHRETESCVG	AQNG---NY	YWRDVGTTIDA	YWEANMELTR	VI---	PELN	LYDREWPIWT	SLEQLPPP---	AKFVFNDEGR	321
69	243	SIILRSYRVY	AHPFSDVQGG	DPG-----	YWRDVGTTIDA	FYANANLELIG	VS---	PELN	LYDEDWPIWT	YQEQQLPP---	AKFIFDNEDR	318
70	249	YMVPR-YRVI	AHRFRHSCTS	SAGSGNPQRC	YWRDVGTTIDA	YWAANIDLTVH	VT---	PDLD	LYDSRWPIWT	YQEQQLPP---	AKFVFDNADR	330
71	243	AIIE-EHNVF	AHPFRDPQOE	GQP-----	YWRDVGTTIDS	FWEANMELVM	PE---	PQLD	LYDPAWPIWT	YQEQQLPP---	AKFIFDDDDR	317
72	225	LLLEEKKLS	AYPFKG----	-----	YWKDVGTVQS	LWEANMDLLK	ED---	SELK	LFERKWKIYS	VNFNQPPQFI	SS	288
73	241	KVVA-QGKAL	AHPFSMSCVS	SNANAP---A	YWRDVGTTIDA	FWAANLDLAS	II---	PELD	IYDENWPIWT	YQRQLPPP---	AKFIPDVNGQ	319
74	246	AGVY-EGVYV	AHPFEKSKG	RNTQG---TI	YWRDVGTTIDS	YWSANIDLVS	EY---	PQLD	MFDESPIRT	VPKQTAP---	TKFFYKHSQA	324
75	259	RQMHR-GKTV	AYLYDG----	-----	YWDIGTIGS	YFEANMALTQ	RPSHNIRGFN	CYDDGGIYS	KNNHLPG---	AKFVFDNADR	321	
76	240	SLIA-NSHVQ	SYAFVD-DHG	EPG-----	YWRDVGTTIAS	YWNANMDLCS	IT---	PELN	LYNEDWPIWT	YQEQMPPP---	AKFAFDDDEGR	313
77	261	HAIK-NDKVY	AYALRDVHEP	DKAG-----	YWRDVGTTIDA	YWKANLELCD	VV---	PELN	LYDEDWPIWT	HQKQTPP---	AKFVFDNADR	336
78	244	KAVG-EGQVV	AHPFQDSCVY	-NSE--KAPA	YWRDVGTTIDA	YWEANIDLTA	TV---	PELN	LYDRSWPIWT	YQEQQLPP---	AKFVHNHADR	322
79	231	RLDTHC-VC	AYRFGGEAGR	VTQDK----	YWRDVGTTIDS	YTYANMELTA	QV---	PELD	LYQPGWPIWT	YHGQNPFP---	ARMAPGSLGQ	307
80	247	AVID-KYIVN	AYPFLDLQSG	EQS-----	YWRDVGTTIDA	YWSANMELIG	VK---	PDLN	LYDKTWPIWT	YQEQTPP---	AKFVFDSDKR	321
81	230	KLIDK-AY	AHQFGGSGR	VTEDD----	YWRDVGTTIDS	LYQANMDLLQ	PV---	SPID	LYQDQWPIWT	YEPQLPPP---	ARTSSDTGN	306
82	243	AIIE-DHQVF	AHPFSDPDS	QQP-----	YWRDVGTTIDS	FWEANMELVT	PE---	PQLN	LYDSNWPIWT	YQEQQLPP---	AKFVFDNADR	317
83	231	-RIVADGMAA	VYDFSDNEVP	GATDR--DRA	YWRDVGTTIDA	FYDAHMDLVS	VH---	PVFN	LYNKRWPIS	ESENLAP---	AKFVN	306
84	246	RCLB-EGTLY	AHPFSRSCMG	RNTEG--DI	YWRDVGTTIDS	FWQSNIDLVS	EH---	PQLD	IYDQSWPIRG	NFVQSYPP---	SKFFYKNANI	324
85	242	GATE-HRKYI	AYPFRDPESG	EQP-----	YWRDVGTVDA	FWEANMELVS	IT---	PELN	LYDQGWPIQT	YQRQLPSS---	AKFVFDNADR	316
86	244	AAIA-YNRVI	AYPFDPKSG	EQP-----	YWRDVGTVDA	FWEANMELIG	KG---	SELD	LYDQDRIWT	YQEQQLPP---	AKFIND-AGH	317
87	238	KIVG-QGNAM	AHPFSMSCVP	SAQFVP--P	YWRDVGTTIDS	FWSANLDLTS	NM---	PQLN	IYDEDWPIWT	YQEQEPPP---	AKFVFDNADR	316
88	250	KALE-EGVLY	AHPFSRSCMG	RNTEG--EI	YWRDVGTTIDS	FWQSNIDLVC	EN---	PQLD	IYDQWPIRG	NFVQTYPP---	SKFFYKKEV	328
89	247	KITK-RGDVL	AHPFELSCVN	SDPSPV--P	YWRDVGTTIEA	YWSANLDLVS	VT---	PELD	MYAKDWPIRT	FITSLPPP---	AKFVQDNHDE	325
90	226	KMLADEGRLY	SYAFSG----	-----	YWRDVGTTIES	LYQANMDLLQ	DE---	PPFE	LSG-KWRIYS	FNPSMPPQFV	GK	288
91	247	RSMK-EGVLY	AHPFSRSCMG	HNTEG--TP	YWRDVGTTIDS	YWNANMDLVT	EY---	PQLN	LFDKDWAIHG	LPTQSMPP---	TKFFCKDNCL	325
92	243	HIMSR-YRMF	AHPFSDSCVA	APG---ESA	YWRDVGTVDA	YWEANMELTK	VT---	PDLN	LYDKTWPIWT	YQEQQLPP---	AKFVFDNADR	320
93	259	YLVR-NRVI	AHPFSDSCVN	MVG---DVP	YWRDVGTVDA	YWEANMELTK	VT---	PELN	LYDDAWPIWT	HQKQLPPP---	AKFIFNDRR	336
94	226	-RLIHSRVI	AYNFADNRIP	GTSSYE-EQP	YWRDVGTTIDA	FYAAHQDVLG	EH---	PRFD	LFNPQWVFN	-SNYQGP---	APRIV	301
95	242	GIID-RYRVQ	AYPFRGKQG	VQA-----	YWRDVGTTIDS	YQANLELIG	VT---	PELN	LYDEWPIWT	YQEQWPPP---	AKFVFDNADR	316
96	229	-RLLRDRHFL	AYDFATNEVP	GIKPYE-KRV	YWRDVGTTIDA	YFDAHQDVLG	DE---	PVFD	MFNPQWPIFS	-SNYQGP---	VARVL	304
97	228	-KMFPRGKVI	YDFETNKIK	G-EKES-T-	YWRDVGTTIES	YWSANMDLVS	KD---	PEFS	LYNRSWPIHT	YYPPLPPP---	ATFV--DVKD	304
98	230	KLIATGS-VF	AYSFCSGKGR	VARDC----	YWRDVGTTIDS	FYDANMDLLQ	PV---	PPMN	LYQKNWAIWT	YEQQYPPP---	ARTVSSATGN	306
99	228	-NMFPRGDFV	VYDFSTNRTT	G-EKEE-V-	YWRDVGTTIDA	YQANMELLE	KD---	APFS	LYNRSWPIHT	YYPPLPPP---	ATFT--DSAN	304
100	241	KLVGNNS-VY	AYKFGDEEGR	VTQDA----	YWRDVGTTIDS	YQSNMDLLK	PT---	SPID	LYQDQWAIWT	YEPQLPPP---	ARTIASVEGN	317
101	226	AMLGDGKSMW	AYQYSG----	-----	YWRDVGTTIQA	YWSANMDLVS	RV---	PQFN	LFDPWPIWT	PNFVKPAHYI	AS	289
102	240	YIIE-HGKAV	AHPFSDSCVR	SDFEHE--P	YWRDVGTTIDA	YQANIDLTD	VV---	PDLD	LYDKSWPIWT	YAEITPPP---	AKFVHDDDR	318
103	235	MMLAKDRVF	AYNFNDNLIP	G-MKPE-ERG	YWKDVGTTIDS	YFEANMELIH	VS---	PQLN	LYNYKWPIHT	NQGNYPPP---	AKTVEDEDDG-	314
104	240	ACLID-DYNVQ	AYLEFDD----	-----	YWEDIGTIEA	FYKANLALTS	QNA---	PPFS	FYHP-APIYT	RPRVLPSS---	S	298
105	245	SQVG-KARIV	AHPFSQSCVY	SVGR---REP	YWRDVGTVDA	YWSANIDLVS	VT---	PALD	LYDADWPIWT	YQMQRPPP---	AKFVFDNADR	323
106	330	GATSLGLRVQ	AYLYDG----	-----	YWEDIGTIEA	FYNANLGITK	KPV---	PDFS	FYDRSAPIYT	QPRVLPSS---	S	390
107	248	AAAR-SHLVQ	TYLFNG----	-----	YWEDIGTIGS	FYEANLALTQ	QPQ---	PPFS	FYDENAPIYT	RPRVLPSS---	S	307
108	241	EALSRGDALK	SYVFDA----	-----	YWEDIGTIGA	FYESNLALTQ	QPT---	PPFS	FYDEKFIYT	RARYLPSS---	S	301
109	240	ACLID-DYNVQ	AYLEFDD----	-----	YWEDIGTIEA	FYKANLALTS	QNA---	PPFS	FYHP-APIYT	RPRVLPSS---	S	298
110	244	MAAQKGMKV	AHLYDG----	-----	YWEDIGTVDA	FFHANLECD	PN---	PKFS	FYDRNAPIYT	QSRFLPPP---	S	303
111	286	KASADGARVQ	AYLFND----	-----	YWEDIGTMKS	FYEANLALAK	DP---	PNFE	FYNAEAPIYT	SPRFLPPP---	A	345
112	325	GAKDAGFKVQ	AFAFDG----	-----	YWEDIGTVEA	FYNANLALTD	PEK---	AQFS	FYDKDAPIYT	MSRFLPPP---	S	385
113	316	SAAKD-HNVV	AYPFYQ----	-----	YWEDIGTIKS	FYEANLALAK	HP---	ATFE	FYDQSPPIYT	SPRFLPPP---	A	374
114	317	ASAKE-IDVK	AYLFND----	-----	YWEDIGTIKS	FYEANLALAE	QP---	PRFS	FYDADKPMYT	SRRNLPPP---	S	375
115	285	GATSIKGRVQ	AYLYDG----	-----	YWEDIGTIAA	FYNANLGITK	KPI---	PDFS	FYDRFPIYT	QPRHLPPP---	S	345
116	320	GATEIGLRVQ	AYLYDG----	-----	YWEDIGTIEA	FYNANLGITK	KPV---	PDFS	FYDRSAPIYT	QPRVLPSS---	S	380
117	326	MAAKD-YDVQ	AYLEFDD----	-----	YWEDIGTIKS	FYEANLALTD	QS---	PNFY	FYDPVKPIYT	SPRFLPPP---	T	384
118	328	RAVLID-HSVQ	ACIFETG----	-----	YWEDVGTIKS	FFDANLALTE	QP---	SKFD	FYDPKTPFIT	SPRFLPPP---	T	386
119	340	SAVKE-HNVQ	AYLFND----	-----	YWEDIGTIKS	FFDANLALTE	QP---	PKFE	FYDPKTPFIT	SPRFLPPP---	T	398
120	340	ASAKE-FYMK	AYLFND----	-----	YWEDIGTIKS	FFDANLALTE	HP---	PRFS	FYDAAKPMYT	SRRNLPPP---	S	398
121	339	AAVME-HNVQ	AYIFKD----	-----	YWEDIGTIKS	FYEANLALAE	EP---	PKFE	FYDPKTPFIT	SPRFLPPP---	T	397
122	332	GATSIKMRVQ	AYLYDG----	-----	YWEDIGTIEA	FYNANLGITK	KPV---	PDFS	FYDRSSPIYT	QPRVLPSS---	S	392
123	339	LAVME-HNVE	AFLFRD----	-----	YWEDIGTIKT	FYEANMGLTE	EF---	PKFE	FYNPKTPIFT	SPRFLPPP---	T	397
124	332	ASAKE-FFIK	AYLFND----	-----	YWEDIGTIKS	FYEANLALTA	HP---	PRFS	FYDATKPMYT	SRRNLPPP---	S	390
125	319	GATSLGLRVQ	AYLYDG----	-----	YWEDIGTIEA	FYNANLGITK	KPV---	PDFS	FYDRSSPIYT	QPRVLPSS---	S	379
126	336	LAVKD-HNVQ	AYLFND----	-----	YWEDIGTIKS	FFDANLALTE	QP---	PKFE	FYDPKTPFIT	SPRFLPPP---	T	394
127	319	AAAKE-INVK	AYLFND----	-----	YWEDIGTIKS	FYEANLALAE	QP---	PRFS	FYDADKPMYT	SRRNLPPP---	S	377

128	320	GATEIGLRVQ	AYLYDG	----	----	YWEDIGTIEA	FYNANLGITK	KPV	---	PDFS	FYDRSAPIYT	QPRYLPP	----	----	S	380
129	329	RAVLE-HNVQ	TCIFMG	----	----	YWEDVGTIKS	FFDANLALTE	QP	---	SKFD	FYDPKTPFFT	APRYLPP	----	----	T	387
130	331	KALHE-HNVQ	AYVFTD	----	----	YWEDIGTIRS	FFDANMALCE	QP	---	PKFE	FYDPKTPFFT	SPRYLPP	----	----	T	389
131	277	KAADGYHVQ	AYLFND	----	----	YWEDIGTIKS	FFEANLALAK	NP	---	PQFE	FYDARAPIYT	SPRFLPP	----	----	A	336
132	312	NAKDLGMHVQ	AFLYDG	----	----	YWEDIGTIKA	FFDANLACND	PEK	---	AKFS	FYQTGAPIYT	QSREFLPP	----	----	S	372
133	335	YAVKD-HNVQ	AYLFND	----	----	YWEDIGTIKS	FFDANLALTE	QP	---	PKFE	FYDPKTPFFT	SPRFLPP	----	----	T	393
134	343	AAVME-HNVQ	SYNFRD	----	----	YWEDIGTIKS	FYEANLALTE	EP	---	PTFE	FYDPKTPFFT	SPRFLPP	----	----	T	401
135	340	LSAKD-YNVR	AYLFND	----	----	YWEDIGTIKS	FFDSNLALTD	QP	---	PEFQ	FFDPLKPIFT	SPRFLPP	----	----	T	398
136	345	ASAKE-FFIK	AYLFND	----	----	YWEDIGTIQS	FFAANLALTE	HP	---	PRFS	FYDAAKPMYT	SRRNLPP	----	----	S	403
137	331	GATSIGLRVQ	AYLYDG	----	----	YWEDIGTIEA	FYNANLGITK	KPI	---	PDFS	FYDRSSPIYT	QPRYLPP	----	----	S	391
138	334	FSAKE-FYVN	AYLFND	----	----	YWEDIGTIRS	FFEANLALTE	HP	---	GAFS	FYDAAKPIYT	SRRNLPP	----	----	S	392
139	332	AAIRD-HDVQ	GYIFRD	----	----	YWEDIGTIKT	FYEANLALVE	ER	---	PKFE	FYDPDTPFFT	SPRFLPP	----	----	T	390
140	275	ASAKE-YNVQ	AYLFND	----	----	YWEDIGTIKS	FYEANLALTE	QP	---	PKFR	FYDAAKPIYT	SPRYLPP	----	----	T	333
141	272	ASAKE-YNVQ	AYLFND	----	----	YWEDIGTIKS	FYEANLALTE	QP	---	PKFR	FYDAAKPIYT	SPRYLPP	----	----	T	330
142	345	AAANE-YNVQ	AYLFND	----	----	YWEDIGTIKS	FFDANLALTA	QP	---	PKFS	FYDASNPIFT	SPRFLPP	----	----	T	403
143	310	AAASE-YNVQ	AYLFND	----	----	YWEDIGTIKS	FFDANLALTA	QP	---	PKFS	FYDASNPIFT	SPRFLPP	----	----	T	368
144	263	GATQLGMKVQ	AYLFDG	----	----	YWEDIGTIEA	FYNANIGLTK	SP	---	PEFS	FDDKHSPIYT	LPRCLPP	----	----	S	322
145	269	GATQLGMKVQ	AYLYDG	----	----	YWEDIGTIEA	FYHANLGETK	KPV	---	PNFS	FYDRSAPIYT	QARFLPP	----	----	S	329
146	244	GATQLGMKVQ	AYLFDG	----	----	YWEDIGTIEA	FYNANIGLTK	SP	---	PEFS	FDDKHSPIYT	LPRCLPP	----	----	S	303
147	259	ASAKE-YNVQ	AYLFND	----	----	YWEDIGTIKS	FYEANLALTE	QP	---	PKFR	FYDAAKPIYT	SPRYLPP	----	----	T	317
148	324	EAAKN-HNVV	AYPFYG	----	----	YWEDIGTIKS	FPEENLKLKR	HP	---	ATFE	FYDPQSPIYT	SPRVLPP	----	----	A	382
149	323	GAKDAGYKVQ	AYAFKG	----	----	YWEDIGTVEA	FYANLALAD	PSK	---	AQFS	FYDKDPIYT	MSRFLPP	----	----	S	383
150	314	KANELGKHVQ	AFLYKG	----	----	YWEDIGTIEA	FYNANLQOND	PDA	---	PKFS	FYESSGPIYT	QSREFLPP	----	----	S	374
151	313	KAADGYHVQ	AYLFKD	----	----	YWEDIGTIKS	FFEANLALAK	HP	---	PQFE	FYDARAPIYT	SPRFLPP	----	----	A	372
152	244	MAAQMLGVQ	AFLYEG	----	----	YWEDIGTVDA	FFHANLSND	PN	---	PAFN	FHEMNAPIYT	QSREFLPP	----	----	S	303
153	268	RAAADGAKVQ	AYLFND	----	----	YWEDIGTMKS	FFEANLNLAKE	DP	---	PNFE	FYNAEAPIYT	SPRFLPP	----	----	A	327
154	331	AAVRD-HDVQ	SYFFED	----	----	YWEDIGTIKS	FYDANLALTE	ES	---	HKFE	FYDPKPIYT	SPGFLPP	----	----	T	389
155	342	SAVNE-HNVQ	AYLFND	----	----	YWEDIGTIKS	FFDANLALTE	QP	---	PKFE	FYDPKTPFFT	SPRFLPP	----	----	T	400
156	331	AAVRD-HNVQ	SYFFGD	----	----	YWEDIGTIKS	FYDANLALTE	ES	---	HKFE	FYDPKPIYT	SPGFLPP	----	----	T	389
157	343	SAVNE-HNVQ	AYLFND	----	----	YWEDIGTIKS	FFDANLALTE	QP	---	PKFE	FYDPKTPFFT	SPRFLPP	----	----	T	401
158	342	AAVRE-NNVQ	AYFFID	----	----	YWEDIGTIKS	FYDANLALTE	EN	---	PMFK	FYDPKTIYT	SPRFLPP	----	----	T	400
159	340	AAVRE-NNVQ	AYFFND	----	----	YWEDIGTIKS	FYDANLALTE	EN	---	PMFK	FYDPKTIYT	SPRFLPP	----	----	T	398
160	335	SALRD-HKVQ	AYMFRD	----	----	YWKDIGTIKS	FFEANLELTK	QS	---	PNFE	FYDQESPFFT	SPRFLPP	----	----	T	393
161	336	SALRD-HKVQ	AYMFRD	----	----	YWKDIGTIKS	FFEANLELTK	QS	---	PNFE	FYDQETPFFT	SPRFLPP	----	----	T	394
162	335	ASARE-FYMK	AYLFND	----	----	YWEDIGTIRS	FFEANLALTE	HP	---	PRFS	FYDAAKPMYT	SRRNLPP	----	----	S	393
163	316	AAAKE-INVK	AYLFND	----	----	YWEDIGTIKS	FFEANLALAE	QP	---	PRFS	FYDASKPMYT	SRRNLPP	----	----	S	374
164	334	RALHE-HNVQ	AYVFTD	----	----	YWEDIGTIKS	FFDANMALCE	QP	---	PKFE	FYDPKTPFFT	SPRFLPP	----	----	T	392
165	327	GATSTGMRVQ	AYLYDG	----	----	YWEDIGTIEA	FYNANLGITK	KPI	---	PDFS	FYDRSAPIYT	QPRYLPP	----	----	S	387
166	324	GATELGMRVQ	AYLYDG	----	----	YWEDIGTIEA	FYNANLGITK	KPV	---	PDFS	FYDRSSPIYT	QPRYLPP	----	----	S	384
167	340	ASVKE-YNVQ	AYFFGD	----	----	YWEDIGTIKS	FYDANMALTE	ES	---	PMFK	FYDPKTIPT	SPGFLPP	----	----	T	398
168	325	GATSIKRVQ	AYLYDG	----	----	YWEDIGTIEA	FYNANLGITK	KPV	---	PDFS	FYDRSSPIYT	QPRYLPP	----	----	S	385

1	318	---RGSAVSS	VVSGDCIIS	---	---	GAALNRS	LLFTGVRANS	-YSRENAV	LPS	-----	----	----	----	----	----	----	381
2	325	---HGMTLNS	LVSGGCVIS	---	---	GSVVVQS	VLFSSRVRS	-FCNIDSAVL	LPE	-----	----	----	----	----	----	----	388
3	322	KMLDADVTD	VIGEGCVIK	---	---	NCKIHS	VVGLRSCISE	-GAIEDSLL	MGADYYETDA	DRKLLAAKGS	VPIGIGKNCH	IKRAIDKNA	----	----	----	406	
4	341	KIDNCKIKDA	IISHGCELR	---	---	DCSVEHS	IVGERSRLDC	-GVELKDFM	MGADYYQTES	EIASLLAEKG	VPIGIGENTK	IRKCIDKNA	----	----	----	425	
5	318	---HGTAINS	MVSGGDIH	---	---	GAEVRDS	LLFSQVVQP	-GATVHEAVI	LPD	-----	----	----	----	----	----	381	
6	316	---RGMVDS	MVSGGCIVS	---	---	GMVAVDS	LLFSSVVVED	-GSRVDEAVI	LPE	-----	----	----	----	----	----	379	
7	305	---DGHIRTA	SLGAGVVR	---	---	RAVIERS	LLRREVVVEE	-GAEVDSII	MDR	-----	----	----	----	----	----	368	
8	348	KFTTCIQKS	ILCEGSIIE	---	---	ADEITHS	LLGPRTVIGS	-GAIIRD SYL	MGNDYYVSP	--VNDHCKLP	SEPQIGENCI	IKKAIIDKNV	----	----	----	429	
9	322	IVTDSMISNS	LLCEGAVID	---	---	SSNVFHS	VVGIRGVIGK	-NSIIDHSIV	MGNDRYGN	----	----	----	----	----	----	397	
10	321	II TDSMISS	LLCEGCVIN	---	---	TSHVRS	VLGIRSKIGE	-NSVVDQSI	MGNARYGS	----	----	----	----	----	----	396	
11	335	---RGTAVDS	LIAGGCIIS	---	---	GASVKRS	LLFSSVNVHS	-WASVEDSVV	LPD	-----	----	----	----	----	----	398	
12	322	---RQATDS	LISGGCIVS	---	---	GANVRS	VLFSDVRVNS	-YSSIEQSVI	LPK	-----	----	----	----	----	----	385	
13	328	---RQALDS	MVSGGCIVS	---	---	GATVRS	LLFSNVQIRG	-YSTIEDSVI	LPN	-----	----	----	----	----	----	391	
14	320	---RQALDS	MVSGGCIVS	---	---	GATVRS	LLFSNVQVRC	-FSTIEDSVI	LPD	-----	----	----	----	----	----	383	
15	320	---RQALDS	MVSGGCIVS	---	---	GATVRS	LLFSNVQVRS	-YSVLED SVI	LPN	-----	----	----	----	----	----	383	
16	300	KLLDCHVTES	IIGEGCILK	---	---	NCRIQHS	VLGVRRIET	-GCMIEESLL	MGADFYQASV	ERQCSIDKGD	IPVIGIPDTI	IRRAIDKNA	----	----	----	384	
17	319	---QSGNNL	IVCGGCVIS	---	---	GQSIRRS	VLLSSNVKSS	-FCNINEAVL	LPQ	-----	----	----	----	----	----	382	
18	318	---RGMVDS	LVSGGCIVS	---	---	GALVRS	VLFSTGVHLHS	-YSSVEESVL	LPE	-----	----	----	----	----	----	381	
19	310	KMLNSVTES	MIGEGCMIK	---	---	QCRHHS	VLGIRSRIS	-DCTIEDTVL	MGNDFYESSS	ERDTLKARGE	IAAGIGSGTT	IRRAIDKNA	----	----	----	394	
20	328	---RGEAIDS	LVAGGCIVS	---	---	GARVRS	VLFSTVVC	-KSLVKSFI	LPK	-----	----	----	----	----	----	391	
21	294	---RAELDNV	LLGAATIVT	---	---	GAKIRNS	LLRREVVVEP	-GAEIEDSVI	MDY	-----	----	----	----	----	----	357	
22	333	---RGHALDS	SVSGGCIVS	---	---	GATVRS	LLFSNVKVNS	-FSYVEDSVI	LPN	-----	----	----	----	----	----	396	
23	311	GVNLGAALNS	IISGGCIVS	---	---	GATVRS	VLSLNVSVGP	-KSLVEDSVI	LEN	-----	----	----	----	----	----	376	
24	288	---EABVKDS	LVDVGC FVS	---	---	GKVEHS	VLSFNQVKE	-GAQLKDSFI	MSG	-----	----	----	----	----	----	350	
25	306	KYNSSFMQQT	LAADGCIIT	---	---	NANIQNS	VIGVRMLIES	-GAELEGVVC	MGADYYETPA	ERELNRQGTI	PDIGIARGCR	IRHAIDKNA	----	----	----	390	
26	325	---HGMTLNS	LVSGGCIVS	---	---	GSVVVQS	VLFSSRVRS	-FCTIDSSLL	LPD	-----	----	----	----	----	----	388	
27	316	---RGEAIDS	LVAGGCIVS	---	---	GSTLRGT	VLFSSNVVHS	-YGLIEDSVI	LPD	-----	----	----	----	----	----	379	
28	334	---RQALDS	SVSGGCIVS	---	---	GATVRS	LLFSNVKVRS	-YSTVEDSVI	LPN	-----	----	----	----	----	----	397	
29	314	KISNSQINQS	IICEGSIVE	---	---	ASSISNT	ILGPRSVIKK	-GAIIRD SYV	MGNEFYTPP	--VQIKNR-P	STLSIGKDCV	IEHAIDKIVY	----	----	----	394	



30	321	---RGLAIES	MVSGGCIVS	---G-AVYRS	VLFSQVRVHS	-YASVNWAVL	LPG	-----	---	---AQIGRHAR	VTRVVVDRC	383
31	315	---QCCLNQT	LIGTGCQLT	---NCRHHT	VLSSNCAIGD	-GASLQGCVL	LPD	-----	---	---VTIEAGAK	LKNVIVDKGV	378
32	289	---AAKQVNS	IVVDGCEIH	---GEVIHS	VLSTNVKVER	-NAKIISVIV	MPD	-----	---	---VYIGENAI	VNKAIVGSNA	351
33	303	KMNFCHISNS	LASEGSIIT	---NAYIVNS	IIGVRTLIES	-GASLDGVYC	MGASYETQE	EKSRNARNGI	PNIGIGKGTI	IRRAIIDQNA		387
34	295	KLNRRAEMNNS	IASEGCVIT	---NAKISDS	VIGVRSATIES	-GSELNGVIC	MGADYYENAE	QRRLNLEAGV	PALGICRNCK	ISHTIIDKNA		379
35	321	VVVSMSISNS	LLCEGSVIE	---SSRVSHS	VVIGRCMIGS	-NSILDHIV	MGNEGYDS	---	---MHG	GALGIGKDCB	IYKTIIDENC	396
36	309	HIQAGLAEDT	LISNGCIIIS	---GATVKRS	LLSPNVRVDY	-YAEVCDISL	FDD	-----	---	---VHIGARAR	VRAAIEEGV	374
37	306	---NGVIHRS	VVSGSIVD	---GASLDNA	MLRRSVVVER	-DARLEHCIV	MER	-----	---	---SRIGRGAQ	VRAAIDQDN	369
38	305	---DGHIRSA	SLGAGVLR	---RATIER	LIRREVVEE	-GAEVADSIV	MDR	-----	---	---TVIGAGAK	IRRAIDQNN	368
39	316	---RGTVVD	MVSGGCIIIS	---GAQLRRS	LLFSSVIVDE	-RTRVEDSVI	LPE	-----	---	---AHIGPGCR	IRNAVIDKYC	379
40	318	---RGMALDS	MVSGGNIIA	---GASVRRS	VLFSRVKVP	-GAEVQEAIV	LPR	-----	---	---VTVEDGCR	IRRAVIDEGC	381
41	321	---RGMALDS	TVSGGCIVS	---G-YVFRS	VLFSRVVHS	-YAKVNWAVL	LPG	-----	---	---VQVGRGAS	LTRVVVDRC	383
42	317	---RGLAVDS	TVSAGCIIIS	---GSTVRKS	VLSSSVHTRS	-YLSLEESV	LHG	-----	---	---SHVGERCK	LKRVVIDSKC	380
43	308	---CET-AVAQS	LIGAGSYIN	---GAKIENS	ILGFRSHVCQ	-NVIIKDSIF	LGN	-----	---	---AKIGAGSR	LTKVILDKDI	372
44	306	---SFC-HIRCK	LISDGCLIT	---GALIKKS	VLGFKCIVGN	-DTEIHESVL	LGE	-----	---	---STIGRNCI	LLKTIIDKDV	370
45	318	---RGLATSS	LVSGGCIIIS	---GSTLTQT	LLFTGVRHHS	-FSTIEQAVI	LPY	-----	---	---VEVGRACE	LKNVVIDRGV	381
46	318	---RGLAIDS	LIAGGCIVS	---GSTVRHS	LLFPRVVRHS	-YCEISDSV	FPN	-----	---	---VEIHRNCK	IRRALIDRYC	381
47	307	---ESVFN	ILAGGVVIS	---GGSVRHS	ILFDIFIDE	-NAIIEKSI	FGG	-----	---	---VHVGTGAR	LQNCIIDQNV	370
48	319	---QGTGTNM	IVCGGCIVS	---GSQISRS	VLSSNVVNS	-FCNIAEAVL	LPQ	-----	---	---VSIQASCR	LKRVVIDRGC	382
49	302	---ARVGLATS	LVSLEGCIIIS	---GGRIHRS	VLSSNVVRNS	-YSSLSHVVA	FED	-----	---	---VKIGRHHV	LKRVVIDKDV	367
50	322	---RGEALNA	LVSGGCIVS	---GSVVRES	VLSSNVLRVS	-YSTIEQAVV	LPD	-----	---	---VQINRHCR	LKVVVIDRHC	385
51	302	KLVDQAITS	IVCEGTILK	---SCSILHC	VLGFRSRIE	-DSVLEDIV	MGADFFESPE	ERIELRKG	TPLGVEG	VKRAILDKNT		372
52	318	---RGMALDS	LVSGGCIVS	---GSEIRSS	LLFTGCRHS	-YSSLSHVVA	LPH	-----	---	---VTVNRKAD	LTNVIDRGC	381
53	318	---RGMALDS	MVSGGCIIA	---GARVSHS	LLFSNVVRVES	-HSEVSDSV	LPD	-----	---	---VTIGKHCY	IRKAILDKGC	381
54	308	---GGIENS	LFSAAVCVHR	---GARVRNC	ILREAVVEA	-GAEEIECII	MDY	-----	---	---SKIKRGR	LRRVIVDRHN	372
55	318	---RGSVAVS	LISGGCIVS	---GSEVRNS	LLFTGCRSHS	-WSTVQHVVA	LPY	-----	---	---VDIGBRAQ	LTRCVIDRGV	381
56	311	---HGTINM	MVSGGCILS	---GSSVSNS	VLFSNVVRHS	-FCTINECVL	LPD	-----	---	---VLINRSCR	LKNVILDRGC	394
57	322	---RGTALDS	TVSSGCIVS	---G-EVNR	LLFSSCRVHS	-YARVNLVSV	LPD	-----	---	---TTVGTTRAR	LTRCVVSDC	384
58	320	---GRRVATS	LVSGGCIVS	---GGQVNS	VLSPDVRINS	-YAQVADSV	MDG	-----	---	---VQIGRHAR	IRRAIDKQV	385
59	321	---CGMATDS	LVSGGCIIIS	---GAKVRS	VLFSDIRVNS	-YASLDVSV	LPK	-----	---	---VDIGRYVT	LKRVVIDKGT	384
60	322	---RGLAIDS	MVSGGCIIIS	---GARVQHS	LLFSDIRVGS	-RSEVSDSV	LSG	-----	---	---ARIGEDVK	LHRVVDLNNC	385
61	301	KVEKASVTSS	LIADGCRIE	---NATIEKC	VIGVRSVQS	-GSTLERVVM	MGSDYEDSD	DIERLNVKHI	EKIGIGKKT	LKNVVIDKNV		385
62	307	---NRVGHANS	AVSSGCLIS	---GALINQS	ILGYRVHVHS	-HSSIEQSV	MGD	-----	---	---TDIGPGCH	IKRAILDKVE	372
63	294	DVQGASLKKT	LLAEGCSIA	---EAKITNS	VIGIRSKIGS	-QVILRDITM	MGADYYETDE	HHAENRRLGR	PDIGVGDSI	TEAAILDKKA		378
64	335	---HGTMVNT	IVSGGCIVS	---GSKVSSS	VLFSGVRIHS	-FCDINEAVL	LPD	-----	---	---VEVGRGAR	LNRVVVDRC	398
65	282	---KQASKS	LISNGALIN	---GKVEN	VISPGVIEE	-NVVIKDSII	CND	-----	---	---SKIKRGTV	INKSIIIDKEV	344
66	332	---RGMALDS	LVSGGCIVS	---GSLVRRS	LLFYDVRVDC	-YSRIEDSV	LPN	-----	---	---VDIGRHVV	LKVVIVEKNC	395
67	336	---RGLAVDS	LVSGGCIVS	---GSTVKRS	VLFSDVRVNS	-YSSIEDSV	LPN	-----	---	---VDVGRHV	LKNVVIDKNC	399
68	321	---TGKATDS	LVSGGCLIS	---GSCVTNS	VLFSDVRVHS	-YCDIEGAVI	LPK	-----	---	---VEIHRNVI	LKNVVIDRGC	384
69	318	---RGMALDS	MVSGGCIIA	---GARIGHS	LLFSNVCVQS	-HTEVVSVI	LPD	-----	---	---VKIGKCH	IRKVIDDKGC	381
70	330	---RGLAVDS	LVSGGCIIIS	---GATVRRS	VLFSNVVRND	GNTLVVSVI	LPN	-----	---	---VRMGEGAR	LKRVVVEKGA	394
71	317	---RGMALDS	TVSGGCIIIS	---GSAVRKS	LLFSNVHVRS	-FCBIEQSVI	LPG	-----	---	---ATINRGCK	IKRAIIDRSC	380
72	288	---DAQVHDS	LIVNEGCVVY	---GNVSHS	VLFGQVTGK	-HATVTSVIV	MPD	-----	---	---VTIGEHVV	TEAIVPEKGL	350
73	319	---HGKAVNT	LVSGGCIVS	---GSHVQNS	VLFSNVVRVES	-FCHVLDAVI	LPG	-----	---	---VTVRGCRC	LTKVVVIDRGC	382
74	324	---RTID-NS	LIGGSGVIT	---DAEISNS	VIFDRVQVE	-GSHVLEAV	LPQ	-----	---	---VRIKNCV	LKRVVIDRNC	386
75	322	IISDSRISSS	LLCEGAMIE	---SGQVNS	VVVRGVIGQ	-GSVFDRSIM	MGSDSYGS	---	---ESF	P-LGIGKNC	IHTIIDENC	396
76	313	---RGAALDS	MVSGGCILS	---GSRVKRS	IVFSGCFLHS	-YSFIKDSVI	LPQ	-----	---	---VDIGRDCR	ITKAIIDKSC	376
77	336	---RGLAVSS	MVSGGAIVS	---GAQVKNS	VLFTNVIVER	-GSVVEAVV	LPK	-----	---	---VKIGPNCR	IRKAVIDEGC	399
78	322	---RGEALDS	SVSAGCILS	---G-SVHNS	LLFSNCRVHS	-YTQIHGAVL	LPE	-----	---	---VQVGRNVR	LTKVVVDRC	384
79	307	---EQQVINS	LLGTGTVVS	---GGTIRHS	LLFTQVQVNE	-NAVVEDSIL	FDG	-----	---	---VHVGAH	LTRCIVDKNV	370
80	321	---RGLAVDS	MVSGGCIVS	---GAKVRHS	LLFSNVVRNS	-YTTIQDTIV	LPE	-----	---	---VNIHRHCR	ITKAIIEKGC	384
81	306	---EGIFINS	LISNGVLI	---GGSVQNS	VLSSNVKIND	-GATVSASIL	FDD	-----	---	---VEVEYSQ	LLNCCIIDKHV	369
82	317	---RGMALDS	TVSGGCIIIS	---GSTIRKS	LLFSNVHVHS	-YSTIEESVI	LPG	-----	---	---ADIGEHQ	LRRIVDSK	380
83	306	---GSAQES	VVAGSIIIS	---AASVRNS	VLSSNVVDD	-GAIVEGSVI	MPG	-----	---	---TRVGRGAV	VRHAILDKNV	369
84	324	---KPVVD-NS	LIGGCVIT	---DASISYS	VLFRIRINE	-GSSIDHSV	LPE	-----	---	---VVIGKNCI	LRHCIIDRHS	386
85	316	---EGKALDS	IVSGGCIVS	---GAEVRS	LLFSQVRVHS	-YSRIEQSV	LPE	-----	---	---VEIGRHR	IKRAVIDRGC	379
86	317	---RGLAIDS	MVSGGDIHQ	---GAEVRHS	LLFSQVLRVP	-RAKIQDAVI	LPD	-----	---	---VVVEGCR	IRRCVIDEGC	379
87	316	---DGVISNT	MVSGGCIVC	---GSKMSNS	ILFSKVRVQA	-FCNLQVVV	LPN	-----	---	---CEIGQSK	LKVVVIDRGC	380
88	328	---RPVD-NS	LISGGCVIT	---DASISNS	VLFRIRINE	-GSEIEHCV	LPQ	-----	---	---VTIGKNCK	LKRCIIDRHS	390
89	325	---HGQMMNS	LIADGCIIIN	---GSTLYSS	VLFPVVRVES	-FCHIEDSVI	LPD	-----	---	---VTVNHHCY	LKRCIERS	388
90	288	---DARVRS	MISEGTMIL	---GTVEN	VIFPGVRVQ	-GAVVRNSV	LPS	-----	---	---AVVGDGAM	VDYAILAQHA	350
91	325	---HGID-NS	LISGGCLIT	---NATITES	VLFRITVAD	-HSHIHQSVI	LPE	-----	---	---VSIKNCI	LQNCIIEKCK	387
92	320	---RGLAVDS	LVSGGCIIIS	---GATVRS	LLFSNVVHS	-FAEVS SVL	LPD	-----	---	---VNIHRGAR	LRRVVVDKGC	383
93	336	---RGLAMDS	LISGGCIIIS	---GATIER	LLFLKVVVD	-YSLIQDSVI	LPN	-----	---	---VEIGRHVT	LKRVVIDKHC	394
94	301	---SGEIIIS	AIGAGSMVK	---GARIHNS	VLRRREVIEE	-DVEIEDCVI	MDY	-----	---	---SIIRGRSR	LKRVVIDRYN	369
95	316	---RGMALDS	MVSGGCIIIS	---GSTVRHS	VLFSDVRVGT	-GSVVDQV	LPS	-----	---	---VHVGEGR	LKRVVIDKGC	379
96	304	---GGEHNS	LLGASVVD	---GVRIRDS	ILRRREAVIED	-DVELDECIV	MDY	-----	---	---TRIGRGA	LRRVIVDRHN	368
97	304	---KRV-KITDS	LISGGSYIQ	---GSTIYKS	VLGFRSNI	-GSEIASESV	LGD	-----	---	---VKIGAGCT	IKRAIIDKDV	368
98	306	---EGIFINS	LIANGVINS	---GGSVQHS	ILSSNVIRND	-SALVIESIL	FDD	-----	---	---VEVEGCK	LHCIIIDKHV	369
99	304	---GRV-QIIDS	LVCNGSYVR	---GSRIEKC	VLGFRSNIAS	-ACDISESIL	LGD	-----	---	---VKVGECCV	LRRVIVDKDA	368
100	317	---QGFINS	MIANGVVIE	---GGSQNS	IFFPKVKVSN	-AAVIDSIL	FED	-----	---	---VEIGKNCH	LQNCIIDKNV	380

101	289	---SACVKK	IIAEGCSVH	---GTVINS	ILFFPGAYIEE	-GAVIQDSII	MSN-----	-----	---SRVCKNAY	INRSIISEQA	351
102	318	---RGSAVSS	VVSGDCIIS	---GAALNRS	LLFTGVRRNS	-YSRIENAVV	LPS-----	-----	---VKIGRHAQ	LSNVVIDHGV	381
103	314	--RRGMNDS	YVCAGCITS	---GSVVRRS	IVGPLTKVNS	-YSLVEDSIL	FEN-----	-----	---VNVGRNVK	IRRAIIDKNI	378
104	299	KLIDQCIAES	IITEGCIK	---QARIFHS	VLGLRSRIS	-GVRIEDSLL	MGADFYETPI	QREESLRRGL	PPVIGIGERC	V LQKAIIDKNA	383
105	323	---RGMAKDS	LVSAGCIVS	---GGAVTGS	LLFNDVRVNS	-YSSVIDTVI	LPM-----	-----	---GDIGRHAR	LTKCILDTCG	386
106	391	KMLDADVTD	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDSLL	MGADYYETAT	EKSLLSAKGS	VPIGIGKNSH	IKRAIIDKNA	475
107	308	KILSSTITES	IISGECILK	---ECQVHRS	VLGVRSRVES	-GCVIDHSLL	MGADYYQDSA	QRSQLRLQKH	IPIGIGANSV	IRRAIVDKNA	392
108	302	KLVDQAQIT	IVGEGSILK	---SCSIHHC	VLGVRSRIS	-DVVLEDSLV	MGSDFYESA	ERIALRKGGG	IPLGVGQGT	VKRAILDKNT	386
109	299	KLIDQCIAES	IITEGCIK	---QARIFHS	VLGLRSRIS	-GVRIEDSLL	MGADFYETPI	QREESLRRGL	PPVIGIGERC	V LQKAIIDKNA	383
110	304	KVQDCEIERS	TIGDGCITKQ	---AKLKNV	MVGLRSTVNE	-GCDLEDTLV	MGADYYESLE	ECDPASLPGC	TPIGIGAGTK	IRKAIIDKNA	388
111	346	KIERCHVKDS	IISHGCAALA	---DCSVEES	IVGLRSRVEA	-GTKIKRIMI	IGADFYESE	KRKAILAAGG	VPVIGIGENTI	IENAIIDKNA	430
112	386	KVMDCDVNMS	IIGDGCVIKA	G---SKIHNS	IIGIRSLIGS	-DCIIDSAMM	MGSDYYETLE	ECEY--VPGC	LPMGVGDGSI	IRRAIVDKNA	469
113	375	TVRNCKVTD	IIAQGSFVS	---DCTINNA	VIGRSIIGQ	-NCTIQDALV	MGADYYESDD	QRATLLKKG	VPVIGIGANSV	IKRAIIDKNA	459
114	376	MVNNSKITDS	IISHGCFLD	---NCRIEHS	VVGVRSRIGS	-NVHLKDTVM	LGADYYETA	ERGELLAEGK	VPIGIGENTT	IQKCIIDKNA	460
115	346	KVLDDADVTD	VIGEGCVIK	---NCKINHS	VVGLRSCISE	-GAIIEDSLL	MGADYYETEA	DKKLLAEKGG	IPIGIGKNSC	IRRAIVDKNA	430
116	381	KVLDDADVTD	VIGEGCVIK	---HCTINHS	VVGLRSCISE	-GAVIEDSLL	MGADYYETEN	DKNVLSETGG	IPIGIGKNSH	IKRAIIDKNA	465
117	385	KVENCKVLNS	IVSHGCFLT	---ECSVEHS	VIGIRSRLEP	-GVQLKDTMM	MGADYYQTEA	ERLSELSVKG	VPVIGIGENTK	IRNCIIDKNA	469
118	387	QLDKCKMKYA	FISDGCFLR	---ECNIEHS	VIGVCSRVSS	-GCELDKDSVM	MGADYYETEE	EASKLLLAGK	VPVIGIGRNTK	IRNCIIDMNA	471
119	399	KVDKCRIVDA	IISHGCFLR	---ECSVQHS	IVGVRSRLES	-GVELDITMM	MGADYYQTES	EIASVLAEGK	VPVIGVGNQTK	IRNCIIDKNA	483
120	399	KIDSSKIVDS	IISHGCSFLN	---NCFIEHS	VVGLRSRINS	-NAHLQDTVM	LGADFYETEA	EVASVVAEKS	VPVIGIGANSV	IKRAIIDKNA	483
121	398	KFDKCRIVNA	IISHGCFLR	---ECTVQHS	VVGERSRLDY	-GVELKDTVM	LGADCYQTEV	EIASLLAEGE	VPIGVGRNTK	IRNCIIDKNA	482
122	393	KMLDADVTD	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETDA	DRRFLAAKGS	VPIGIGKNSH	IKRAIIDKNA	477
123	398	KIEQCQVDDA	IISHGCFLR	---ECSVKHS	IVGERSRLDY	-GVELKDTVM	MGADFYQTES	EIASLLAEGN	VPIGIGRNTK	IRNCIIDKNA	482
124	391	KIDDSKIVDS	IISHGCSFLN	---NCFIEHS	VVGLRSRINS	-NVHLKDTVM	LGADYYETDS	EVASLLAEGR	VPIGIGENTR	IKDCIIDKNA	475
125	380	KMLDADVTD	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETDA	DRRFLMAKGS	VPIGIGKNSH	IKRAIIDKNA	464
126	395	KVEBCRILDA	IISHGCFLR	---ECSVQHS	IVGVRSRLEY	-GVELKDTMM	MGADYYQTES	EIASLLAEGK	VPVIGVGNQTK	IRNCIIDKNA	479
127	378	MVNNSKITDS	IISHGCFLD	---NCRIEHS	VVGVRSRIGS	-NVHLKDTVM	LGADYYETDA	ERRELLAEGN	VPIGIGENTR	IQKCIIDKNA	462
128	381	KVLDDADVTD	VIGEGCVIK	---HCTINHS	VVGLRSCISE	-GAVIEDSLL	MGADYYETED	DKKVLSENGG	IPIGIGKNAH	IRKAIIDKNA	465
129	388	QLDKCKIKDA	SISDGCFLR	---ECSIEHS	VIGVCSRVSY	-GCELDKDSVM	MGADYYETEE	EASKLLLAGK	VPVIGIGRNTK	IRNCIIDINA	472
130	390	KSDKCRIKDA	IISHGCFLR	---ECAIEHS	IVGVRSRINS	-GCELDKDSVM	MGADYYETED	EISRLLEKGG	VPIGVGNQTK	ISNCIIDKNA	474
131	337	KVEKCHVKDA	IISHGCSLA	---DCSVEDA	IIGLRSQIGK	-GCTIKHAMI	IGADYYETDE	QKMALVEAGG	VPVIGIGEGCS	ISNAIIDKNA	421
132	373	KLIDAEVSKC	TIGDGCFIKK	---SKLTNA	MIGLRTNIQE	-DCVIEDVMI	MGADYYEETH	ECED--LPGC	TPIGIGAGTT	IKRAIIDKNA	455
133	394	KVDQCRIVDA	IISHGCFLO	---ECSIKHS	IVGVRSRLES	-AVELMDTMM	MGADYYQTES	EIASLQAEKQ	VPVIGVGNQTK	IRNCIIDKNA	478
134	402	KIDKCRIVDA	IISHGCFLR	---ECTVRHS	VVGERSRLDY	-NVHLKDTVM	LGADYYQTE	EIASLLAEGK	VPIGVGRNTK	IRNCIIDKNA	486
135	399	KIERCQVKDS	IISHGCFLR	---ECSVEHS	IVGVRSRLEY	-GVELKDTMM	MGADYYQTEA	EVAASLAGGK	VPVIGVQETK	IMNCIIDKNA	483
136	404	KIENCKIVDS	IISHGCSFLT	---NSFIEHS	VVGLRSRINS	-NVHLKDTVM	LGADFYETDD	EVAALLAEGR	VPIGIGENTK	IRECIIDKNA	488
137	392	KMLDADVTD	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETDA	DRRFLAAKGS	VPIGIGRNSH	IKRAIIDKNA	476
138	393	KIDNSKIVDS	IISHGCSFLT	---NCLIEHS	IVGIRSRVGS	-NVQLKDTVM	LGADYYETEA	EVASLLAEGK	VPIGIGKNSV	IRKAIIDKNA	477
139	391	KAEKCRMVDS	IISHGCFLR	---ECSIQRS	IIGERSRLDY	-GVELQDTMM	MGADYYQTES	EIASLLAEGK	VPIGIGRDTK	VRKCIIDKNA	475
140	334	KIEKCRVLD	IVSHGCFLO	---ECSVTHS	VIGIRSRVEA	-GAEIQDTMM	LGADFYETEA	EIASMVAEKQ	VPVGVGNQNAK	IRNCILDKNV	415
141	331	KIEKCRIVDA	IVSHGCFLO	---ECSVTHS	VIGIRSRVEA	-GAEIQDTMM	LGADFYETEA	EIASMVAEKQ	VPVGVGNQNAK	IRNCILDKNV	415
142	404	KMEKCRIDS	IVSHGCFLK	---SCSVEHS	LIGVRSRLES	-GVELKDTII	MGADSYETEA	EIAALRAQKQ	VPLGVGEHTT	MRNCLVDKNA	488
143	369	KMEKCRIDS	IVSHGCFLK	---SCSVEHS	LIGVRSRLES	-GVELKDTII	MGADSYETEA	EIAALRAQKQ	VPLGVGEHTT	MRNCLVDKNA	453
144	323	IMHDADIVQS	IIGEGC---	---NCKIYHS	VVGLRSRIAE	-GAVIEDSLL	MGSDFYEQEE	HREHLHSHGG	VPIGIGKYSV	VRKAIIDKNV	404
145	330	KLDADIVDS	IIGEGCLIK	---SCKIHHS	VIGLRSWIAE	-DALVEDALL	MGADFYETDE	ERDALLKGG	VPVIGIGKNSV	VRAIIDKNA	414
146	304	IMHDADIVQS	IIGEGCVIQA	SKKNCKIYHS	VVGLRSRIAE	-GAVIEDSLL	MGSDFYEQEE	HREHLHSHGG	VPIGIGKYSV	VRKAIIDKNV	392
147	318	KIEKCRVLD	IVSHGCFLO	---ECSVTHS	VIGIRSRVEA	-GAEIQDTMM	LGADFYETEA	EIASMVAEKQ	VPVGVGNQNAK	IRNCILDKNV	402
148	383	TVRNCKVDDA	IIAQGSFVA	---DSSISNA	VIGRSIIGS	-GCTVQDALI	MGADYYQSD	QRAALLAAGD	VPVIGIGANSI	ISNAIIDKNA	467
149	384	KVLDDAVSMS	IIGDGCVIKA	G---SKIHNS	IIGIRSLVGS	-DCIIDSAMM	MGADYYETLE	ECEY--VPGC	LPMGVGDGGSV	VRKAIIDKNA	467
150	375	KLIDVQVRS	TIGDGCFIKK	---STISNS	MIGLRTSISE	-GCVIDSMI	MGADYYEETH	ECED--LPDC	TPIGIGAGTV	IRRAIVDKNA	457
151	373	KIEKCHVKDA	IISHGCSLA	---DCCVENA	IVGLRSQVSK	-GCKITERAMI	IGADFYESD	QKAKVIASGG	VPVIGIGBCT	ITNAIIDKNA	457
152	304	KVQDCEIERS	TIGDGCFITK	---AKLKNV	MVGLRSTVNA	-NCDLEDTLV	MGADYYETDY	EAKTSALPGG	VPIGIGAGTK	IRKAIIDKNA	388
153	328	KVERCHVKES	IISHGASLA	---DCQVEES	IIGLRSVVMK	-GCRIKRAMI	IGADFYESDE	KKASLLASGE	VPVIGIGENTI	IENAIIDKNA	412
154	390	KIDKQIVDA	IISHGCFLR	---ECTVQHS	IVGERSRLDY	-GVELDITVM	MGADYYQTES	EIASLLAEGK	VPIGIGRNTK	IRNCIIDKNA	474
155	401	KVEKCKIVDA	IISHGCFLR	---ECSIQHS	IVGVRSRLES	-GVELQDTMM	MGADYYQTEY	EIASLLAEGK	VPIGVGENTK	IRNCIIDKNA	485
156	390	KIDKCRIVDA	IISHGCFLR	---ECTVQHS	IVGERSRLDY	-GVELDITVM	MGADYYQTES	EIASLLAEGK	VPIGIGRNTK	IRNCIIDKNA	474
157	402	KVEKCKIVDA	IISHGCFLR	---ECSVQHS	IVGVRSRLES	-GVELQDTMM	MGADYYQTEY	EIASLVAEKQ	VPIGVGANTK	IRNCIIDKNA	486
158	401	KIDKCRIVDA	IISHGCFLR	---ECTVQHS	IVGERSRLDY	-GVELQDTVM	MGADYYQTES	EIASLLAEGK	VPIGIGRNTK	IRNCIIDKNA	485
159	399	KIDKCRIVDA	IISHGCFLR	---ECTVQHS	IVGERSRLDY	-GVELQDTVM	MGADYYQTES	EIASLLAEGK	VPIGIGRNTK	IRNCIIDKNA	483
160	394	KAICKRIVDA	IISHGCFLS	---ECRVQHS	IVGVRSRLES	-GSELQDTMM	MGADYYQTD	EIATLLKECK	VPIGVGENTK	IRNCIIDKNA	478
161	395	KAICKRIMDA	IISHGCFLS	---ESRVQHS	IVGVRSRLES	-GSELQDTMM	MGADYYQTD	EIATLLKECK	VPIGVGENTK	IRNCIIDKNA	478
162	394	KIDNSKIVDS	IISHGCSFLN	---NSFIEHS	VVGLRSRINS	-NVHLKDTVM	LGADYYETDA	EVVALLAEGR	VPIGIGENTK	IKDCIIDKNA	479
163	375	MISSSKITDS	IISHGCFLD	---NCRVEHS	VVGVRSRVGS	-NVHLKDTVM	LGADFYETDV	ERSDQLAEGK	VPIGIGENTT	IQNCIIDKNA	459
164	393	KSDKCRIKEA	IISHGCFLR	---ECTIEHS	IVGVRSRINS	-GCELDKNAMM	MGADLYETED	EISRLLEKGG	VPIGVGENAK	ISNCIIDMNA	477
165	388	KVLDDADVTD	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETEA	DKQLLAEGGG	IPIGIGKNSH	IKRAIIDKNA	472
166	385	KMLDADITDS	VIGEGCVIK	---NCKIHHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETDA	DRRFLAAKGS	VPIGIGKNSH	IKRAIIDKNA	469
167	399	KIDKCRIVDA	IISHGCFLR	---ECSVQHS	IVGERSRLDY	-GVELQDTVM	MGADYYQTES	EIASVLAEGK	VPIGIGNSNK	VRKCIIDKNA	483
168	386	KMLDADITDS	VIGEGCVIK	---NCKIFHS	VVGLRSCISE	-GAIIEDTLL	MGADYYETEA	DKSFLAAKGS	VPIGIGRNSH	IKRAIVDKNA	470
1	382	VIEPGLIVGE	DPELD----	---AKR-FR	RTESEG----	-----	-----	-----	-----I	CLITQSMIDK	417
2	389	VIEPEGMVIGE	NAEED----	---ARR-FY	RSEEG----	-----	-----	-----	-----I	VLVTREMLRK	424

3	407	RIGDNVKIIN	KDNVQEAARE	TD	----	GY	FIKSG	----	----	----	----	----	I	VTVIKDALIP	446
4	426	KIKKNVSIIN	KDGVQEAADRP	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	VVVVKGQQLKP	404
8	430	RIGKGVQLIN	KQQLTRYES	-	E	----	LV	FIRDG	----	----	----	----	I	IVVPRGQVLP	467
9	398	RIGNGVKLTN	IQQYKYDSP	DG	----	KL	VVRDG	----	----	----	----	----	I	IIIPRGTKIP	437
10	397	CIGNGVKLQN	LKGYIKYDSP	DK	----	KL	FVRDN	----	----	----	----	----	I	IIVPQGTHIP	436
11	399	RIPEGMIVGV	DPEED	----	----	RKR	FH	VSPKG	----	----	----	----	I	TLVTAEMLGQ	434
12	386	RIPDGMIEIV	NLELD	----	----	RKR	FH	ITEQG	----	----	----	----	V	VLVTPDMLGQ	421
13	392	QIPEGLKVG	NPDED	----	----	RKH	FY	VTDG	----	----	----	----	I	TLITPEMLGQ	427
14	384	QIPEGLEIGF	DPVAD	----	----	KKH	FY	VTDG	----	----	----	----	I	TLITPEMLRQ	419
15	384	IIPPGLEVGF	DPVED	----	----	RKH	FY	VTEG	----	----	----	----	V	TLVTPESLQ	419
16	385	RIGHDVKIIN	KDNVQEAADRE	SQ	----	GF	YIRSG	----	----	----	----	----	I	VVVLKNAVIT	424
17	383	AIPDGTIVIGE	DPVSD	----	----	AER	FY	RTDDG	----	----	----	----	V	VLVTPPALRQ	418
18	382	RIPAGMTIGF	DAEDD	----	----	ARR	FH	VSADG	----	----	----	----	V	VLVTVAMLEA	417
19	395	RIGKNVMIVN	KENVQEANRE	EL	----	GF	YIRNG	----	----	----	----	----	I	VVVIKNVITIA	434
20	392	VIPDGMIVIGE	DPVED	----	----	AKR	FH	VTEGG	----	----	----	----	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	RIGENCISIGY	EREGYEDGDY	G	----	YY	HVKDG	----	----	----	----	----	I	IIVIAKNVLP	429
26	389	HIPEGMIVGE	NADED	----	----	SAR	FY	RSEGGGGVSD	SGYAGKVRGK	IEPLGLFVFR	LDLLIRLSLL	IRLNLFIRMN	LLIILTTFFFK	468	
27	380	VIPEGLVIGY	NHDDD	----	----	RARGFR	VSEKG	----	----	----	----	----	V	VLVTRMLGQ	416
28	398	VIPEGLVAGY	DVKAD	----	----	RKR	FY	VTEKG	----	----	----	----	I	TLITPEMLGQ	433
29	395	NIGDGVQLIN	KDRLLTYDG	-	E	----	HV	FIRDG	----	----	----	----	V	IIVPRGADLP	432
30	384	VIPDGMIVIGE	DPAAD	----	----	AAR	FY	RTEG	----	----	----	----	I	VLVTRMLRA	419
31	379	TVPANLAIDD	INIAK	----	----	HFG	FS	VSEKG	----	----	----	----	V	VLITQQVIDS	414
32	352	IIRKNSVVDG	G	----	----	RNV	AI	GLY	----	----	----	----	E	EIKTGMVAN	381
33	388	RIGNGCRIGI	DNIPRAEGDY	P	----	MY	SIHDG	----	----	----	----	----	I	IVINKNAVIA	426
34	380	RIGDNQIGV	SGKTYEDGEH	GPH	----	G	EF	YSSAG	----	----	----	----	I	IVIRKNAIIP	421
35	397	SIGNGVKLSN	LKGYSHYDSP	DG	----	KL	FVRDG	----	----	----	----	----	I	TIIPRGTKLP	436
36	375	TVPPGFSIGY	DREAD	----	----	VRR	FP	VSEGG	----	----	----	----	V	VIVFNNIDL	410
37	370	DIPAHERIGF	DLEAD	----	----	RKR	FH	VTASG	----	----	----	----	I	VVVPRQFFKP	405
38	369	FIPPGMRIGY	DREAD	----	----	RAR	FH	VTDG	----	----	----	----	I	VVVAKGQQLKP	404
39	380	HIEAGTVIGE	DPEAD	----	----	AQR	FT	VSPNG	----	----	----	----	V	VLVTPMLGQ	415
40	382	RIPPGMIVIGE	DLETD	----	----	RER	FH	VTPGG	----	----	----	----	V	VLVTAEMLGQ	417
41	384	VIPDGMIVIGE	DAELD	----	----	SQR	FH	RSPNG	----	----	----	----	I	TLVTKTMLAR	419
42	381	HIPAGLTIGY	DREQD	----	----	IENGFR	VTEKG	----	----	----	----	----	I	TLVTSQMLKA	417
43	373	IIAPNTIIGE	NLEED	----	----	RKN	FT	VSDG	----	----	----	----	V	IAIAKGSRIG	408
44	371	HIADNVQIGV	NLEED	----	----	KRR	FT	VSEEG	----	----	----	----	I	VVVPKGARIG	406
45	382	RIPDGLVVG	DPELD	----	----	EQR	FR	RTENG	----	----	----	----	I	CLITQPMIDR	417
46	382	KIPEGTIVIGY	NLEED	----	----	KKR	FH	VSPKG	----	----	----	----	V	VLVTPDMLGQ	417
47	371	SIPPGEQIGY	DLVKD	----	----	RNR	FT	VSEKG	----	----	----	----	I	VVVPKDFVVF	406
48	383	HIPDGTIVIGE	DPVRD	----	----	GER	FY	RVDG	----	----	----	----	V	VLVTAEAALKR	418
49	368	EIPAGAEIGF	NLEED	----	----	RKK	WF	VSEGG	----	----	----	----	I	VVVPKRAKID	403
50	386	VIPERTVIGE	DAEAD	----	----	ARR	FH	RTEGG	----	----	----	----	V	VLVTRMLDR	421
51	387	RIGDNVVIIN	KDRVEEADKP	EL	----	GF	YIRNG	----	----	----	----	----	I	VVVVKNATIA	426
52	382	VIPEGLVIGQ	EPEED	----	----	ARW	FR	RSEGG	----	----	----	----	I	VLVTPDMLDA	417
53	382	KVPDGMIVIGE	DLEED	----	----	KKR	FY	VTEEE	----	----	----	----	V	VLVTPMLGQ	417
54	373	IIAPGTQIGY	DYQED	----	----	TRR	YH	VSPSG	----	----	----	----	I	VVVPRGKLSF	408
55	382	KVPPGLVVG	DPEED	----	----	AKW	FR	RTEGG	----	----	----	----	I	VLITQSMMLDA	417
56	395	VLPEGMIVIGE	DPELD	----	----	AKR	FE	RTSNG	----	----	----	----	V	VLVTKMLRA	430
57	385	HIPPGMIVIGE	DPDDD	----	----	ARR	FR	RTEG	----	----	----	----	V	TLITRKMLDQ	420
58	386	QVPPKVEIGY	DHEQD	----	----	RARGFT	VTEEG	----	----	----	----	----	L	VVVPKSYIFK	422
59	385	RIPDGMIEIV	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	VIIIPKNTIVK	423
62	373	VVAPGTIIGE	DPELD	----	----	RQR	FQ	VSEGG	----	----	----	----	I	VVVPKARVVG	408
63	379	RIGRNVHIRF	LP--DRPDSE	TD	----	QW	AIRDG	----	----	----	----	----	L	VVVPKSAIIP	416
64	399	VIPDEMIVIGE	DADAD	----	----	AAR	FE	RVDNG	----	----	----	----	V	VLVTRMLKR	434
65	345	IIGSECQIGC	SGEKKANFEQ	PE	ILHSGLN	VIGKG	----	----	----	----	----	----	A	EVPAETCIEK	389
66	396	CIPDGSAGV	DPEED	----	----	RRY	FY	VSPGG	----	----	----	----	V	TLVTAALAEK	431
67	400	KIPDGMIEVGV	NIEED	----	----	RKR	FH	VSENG	----	----	----	----	I	TLITPEMLGQ	435
68	385	SIPEGMQIGV	DLALD	----	----	AKR	FY	VSEKG	----	----	----	----	I	TLVTPDMLGQ	420
69	382	NVPDGTIVIGE	DLEED	----	----	KRR	FY	VTEEG	----	----	----	----	V	VLVTPMLGQ	417
70	395	IIPPGLVVGE	DPVED	----	----	ARR	FH	RTPGG	----	----	----	----	V	TLITPESLQ	430
71	381	EIPAGLEIGF	DRKTD	----	----	EENGFR	VSKKG	----	----	----	----	----	I	VLVTRDMLSE	417
72	351	VLPDGAVIRS	E	----	----	KDIQE	VLLVS	----	----	----	----	----	E	EFVKEKELI	380
73	383	TLPEGMIVIGE	DPVAD	----	----	AQR	FD	RSEGG	----	----	----	----	V	VLVTRMLAN	418

74	387	TIPDGMQIGV	DAELD	---	---	KQR-FR	VSQGG	---	---	---	---	---	---	---	---	V	VLVTKTMLKA	422
75	397	CIGNGVRLQN	LQGHKDYDSE	DG	---	---	KL	VVRDG	---	---	---	---	---	---	---	I	IIVPRGTQIP	436
76	377	VIAPGTIIGE	DRAED	---	---	EKR-FY	VDENG	---	---	---	---	---	---	---	I	VLVTPDMLGQ	412	
77	400	VIPEGTVIGY	DAEAD	---	---	RKA-YT	MSAGG	---	---	---	---	---	---	---	V	VLVTPPEMLGQ	435	
78	385	RIPDGLVVGE	DPDDD	---	---	ARR-FY	RSEGG	---	---	---	---	---	---	---	V	TLITPRMLEK	420	
79	371	HIPPERIGF	NHAAD	---	---	AAR-FV	ISESG	---	---	---	---	---	---	---	I	TVVPRKNIHFL	406	
80	385	EIPEGTVIGE	NRAED	---	---	EKR-FH	VSPGG	---	---	---	---	---	---	---	V	VLVTPDMLGQ	420	
81	370	KIPPRTKIGV	NRAED	---	---	AAR-FT	ISDRG	---	---	---	---	---	---	---	I	VVVPESEYKFE	405	
82	381	VIPAGLIVGH	DKAQD	---	---	LANGFR	VSPKG	---	---	---	---	---	---	---	I	TLVTSMDLKR	417	
83	370	VVGPGEMVGV	DLEKD	---	---	RER-FA	ISAGG	---	---	---	---	---	---	---	V	VAVGKGVWI	404	
84	387	VIPDGMRIIG	DKESD	---	---	RKR-FR	VSSSGK	---	---	---	---	---	---	---	V	VLVTPAMLKR	423	
85	380	RLPEGTVLGE	DHEAD	---	---	AKR-FR	VIDNG	---	---	---	---	---	---	---	I	VLVTPKMLGQ	415	
86	381	RIPRETIVIGE	DDVAD	---	---	RER-FF	VSPKG	---	---	---	---	---	---	---	V	VLVTAEMLGQ	416	
87	380	RIPEGTVIGE	DEELD	---	---	AQR-FY	RSPNG	---	---	---	---	---	---	---	V	VLVTQEMFAK	415	
88	391	VIPDGMIEIGV	NLELD	---	---	RQR-FR	VSSGG	---	---	---	---	---	---	---	V	VLVTPSMMLKQ	426	
89	389	TIPEGTVIGM	NAEED	---	---	SAR-FH	RTEEG	---	---	---	---	---	---	---	I	VLVTRMLEQ	424	
90	351	VMEEGARVIG	E	---	---	ETQIT	VIPEG	---	---	---	---	---	---	---	E	TVSAASTAQR	382	
91	388	VIPDNVFGV	DKEHD	---	---	KARGFR	LSSSGK	---	---	---	---	---	---	---	V	TLVTPMLTK	425	
92	384	KIPDGLVVGE	NPEED	---	---	AKR-FH	VIKNG	---	---	---	---	---	---	---	I	TLITPEMLGQ	419	
93	400	KIPDGSFAGI	VPERD	---	---	KQC-FH	VTERG	---	---	---	---	---	---	---	I	TLITPEMLGQ	435	
94	365	DIAPNSRIGF	DAAAD	---	---	GAR-YT	VSEGG	---	---	---	---	---	---	---	V	VVVPKG	397	
95	380	RIPDHTIIGV	SDEED	---	---	ARR-FY	ISPGG	---	---	---	---	---	---	---	V	RVVTPPEMLGQ	415	
96	369	HIEPGERIGF	DPDVD	---	---	RQR-FH	VSESG	---	---	---	---	---	---	---	I	TVVPRGRASY	404	
97	369	EIAAGTIIGE	DLELD	---	---	RKR-FH	VSDEG	---	---	---	---	---	---	---	I	VVIKAGSKVG	404	
98	370	KIPPYTEIGL	NPIED	---	---	RKR-FH	ISERG	---	---	---	---	---	---	---	V	VVVPESEYQFS	405	
99	369	DIAPGTQIGV	NLKED	---	---	KKH-YH	VSDDG	---	---	---	---	---	---	---	V	VVIPKGARVG	404	
100	381	KVPDGTQIGL	DSLAD	---	---	AKR-FH	ISKQG	---	---	---	---	---	---	---	V	IIVPSSYQFE	416	
101	352	IIGEKARLGE	GPDVFN-EYK	PG	---	IYDSGIT	VVGEK	---	---	---	---	---	---	---	S	SIPADAVIGK	395	
102	382	VIPEGLIVGE	DPELD	---	---	AKR-FR	RTEESG	---	---	---	---	---	---	---	I	CLITQSMIDK	417	
103	379	TIPDGTITIGY	DHGED	---	---	RRRGT	VTEESG	---	---	---	---	---	---	---	I	VVVSPE	412	
104	384	RIGNDVRIILN	KERFDSADHP	ER	---	GF	YIRHG	---	---	---	---	---	---	---	I	VIVPKDVIIP	423	
105	387	RIPEGLVIGE	DPILD	---	---	AKR-FH	VTEQG	---	---	---	---	---	---	---	I	TLVTPDRLAL	422	
106	476	RIGDNVKIIN	SDNVQEAARE	TD	---	GY	FIKSG	---	---	---	---	---	---	---	I	VTVIKDALIP	515	
107	393	CIGRDVKIIN	KDNVVEENRE	DQ	---	GF	YIRSG	---	---	---	---	---	---	---	V	VVIKNAVIP	432	
108	387	RIGENVTIIN	KDRIEADRA	DQ	---	GF	YIRNG	---	---	---	---	---	---	---	I	VVVVKNASIL	426	
109	384	RIGNDVRIILN	KERFDSADHP	ER	---	GF	YIRHG	---	---	---	---	---	---	---	I	VIVPKDVIIP	423	
110	389	RIGENCOIIN	EAGVMDKDCS	SE	---	GY	IIRDG	---	---	---	---	---	---	---	I	IVVIKDAVIK	428	
111	431	RVGKNCVITN	KDNIEDLADE	ER	---	GV	FIRNG	---	---	---	---	---	---	---	I	VTIIRNCTIP	470	
112	470	RIGPKQIIN	KDGVKEANRE	DQ	---	GF	VIKDG	---	---	---	---	---	---	---	I	VVVIKDSHIP	509	
113	460	RVGKNVKIVN	KEGVTEGTR	AE	---	GI	YIRSG	---	---	---	---	---	---	---	I	VVIDKDALVP	499	
114	461	RIGKNVVISN	SEGVDEADRT	SE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVVLKNAIIA	500	
115	431	RIGDNVKIILN	ADNVQEAAME	TD	---	GY	FIKGG	---	---	---	---	---	---	---	I	VTVIKDALLP	470	
116	466	RIGENVKIIN	FDNVQEAARE	TE	---	GY	FIKSG	---	---	---	---	---	---	---	I	VTVIKDALIP	505	
117	470	RIGKNVVIMN	SENVQEAARE	TE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVVLKNAVIL	509	
118	472	RIGKNVVITN	SKGIQEAADHP	EE	---	GY	YIRSG	---	---	---	---	---	---	---	I	VVILKNATIN	511	
119	484	KIGKDVIIIN	ADGVQEAADRP	SE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TAVLKNATIK	523	
120	484	RIGKNVVIAN	SEGIQEAADRS	ME	---	GF	YIRSG	---	---	---	---	---	---	---	V	TVILKNSVIQ	523	
121	483	KIGKDVIMN	KDGVQEAADRE	EE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TIIEKATIE	522	
122	478	RIGDNVKIIN	GDNVQEAARE	TD	---	GY	FIKSG	---	---	---	---	---	---	---	I	VTVIKDALIP	517	
123	483	KIGKDAVIVN	KDGVQEAADRP	DD	---	GF	YIRSG	---	---	---	---	---	---	---	I	TIIEKATIK	522	
124	476	RIGKNVVISN	SEGIQEAADRS	LE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TIILKNFTIK	515	
125	465	RIGDNVKIIN	SDNVQEAARE	TD	---	GY	FIKSG	---	---	---	---	---	---	---	I	VTVIKDALLP	504	
126	480	KIGRDVVIAN	ADGVQEAADRP	SE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVILKNATIN	519	
127	463	RIGKNVVISN	SEGVVEADRT	SE	---	GF	YIRSG	---	---	---	---	---	---	---	V	TVVLKNSIIA	502	
128	466	RIGENVKIIN	FDNVQEAARE	TE	---	GF	YIRSG	---	---	---	---	---	---	---	I	VTVIKDALIP	505	
129	473	RIGKNVVITN	SKGIQEAADHP	EE	---	GY	YIKSG	---	---	---	---	---	---	---	I	VVILKNATIK	512	
130	475	RVGRNVSIITN	TEGVQEAADRP	EL	---	GY	YIRSG	---	---	---	---	---	---	---	I	VVILKNATIK	514	
131	422	RIGKNCIITN	AAGVEDLEDE	EN	---	GI	YIRSG	---	---	---	---	---	---	---	I	VTIIRNATIP	461	
132	456	RIGMDCIIN	KDNVQEAARE	DK	---	GY	IIKDG	---	---	---	---	---	---	---	I	VVICKDAIIP	495	
133	479	KIGRGVVITN	ADGVQEAARE	EE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVIMENATIN	518	
134	487	KIGKDVIVIN	KDGVQEAADRP	EE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TIIEKATIE	526	
135	484	RIGKNVVIAN	KDHVEEADRP	SE	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVVLKNSSEIK	523	
136	489	RIGKNVVIAN	SEGIQEAADRS	SE	---	GF	YIRSG	---	---	---	---	---	---	---	V	TIILKNSVIQ	528	
137	477	RIGDNVKIIN	SDNVQEAARE	TD	---	GY	FIKSG	---	---	---	---	---	---	---	I	VTVIKDALIP	516	
138	478	RVGKNVVIAN	SEGIQEAADRS	SD	---	GF	YIRSG	---	---	---	---	---	---	---	I	TVILKNSIIK	517	
139	476	KIGKNVIMN	KDGVQEAADRP	EE	---	GF	YIRLG	---	---	---	---	---	---	---	I	TVIVEKATIQ	515	
140	419	RIGKNVVIAN	KDNVQEAARE	SE	---	GY	YIRSG	---	---	---	---	---	---	---	I	TVILKNATIA	458	
141	416	RIGKNVVIAN	KDNVQEAARE	SE	---	GY	YIRSG	---	---	---	---	---	---	---	I	TVILKNATIA	455	
142	489	RIGSHVITN	TDGVQEAARE	SE	---	GI	YIRSG	---	---	---	---	---	---	---	I	TVVVKNSIVK	528	
143	454	RIGSHVITN	TDGVQEAARE	SE	---	GI	YIRSG	---	---	---	---	---	---	---	I	TVVVKNSIVK	493	
144	405	RIGRNVRIIN	KDNVLEAARE	TE	---	GY	FIKNG	---	---	---	---	---	---	---	I	VTIIKDAVIV	444	



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	RYYRIR-----	423
54	409	YARNSRG--- KGLGYAE -----	422
55	418	RAAALG-----	423
56	431	LGTEELNEED 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 GSTVTRPQT -----	417
66	432	MAEWRA-----	437
67	436	IVHSIR-----	441
68	421	DLRVVI-----	426
69	418	KYRYRIR-----	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	MAEKDAAKAL EKAAKQTAE L 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	DIPDVTIR--- --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	<i>Agrobacterium tumefaciens</i>
2	62288123	P0A6V1.2	<i>Escherichia coli</i>
3	232164	P23509.2	<i>Solanum tuberosum (potato)small subunit</i>
4	232166	Q00081.1	<i>Solanum tuberosum (potato)large subunit</i>
5	499947973	WP_011628707	<i>Alkalilimnicola ehrlichii</i>
6	499558932	WP_011239715	<i>Aromatoleum aromaticum</i>
7	46399384	CAF22833	<i>Candidatus Protochlamydia amoebophila UWE25</i>
8	499777686	WP_011458420	<i>Chlamydia felis</i>
9	499185705	WP_010883245	<i>Chlamydia pneumoniae</i>
10	71845846	AAZ45342	<i>Dechloromonas aromatica RCB</i>
11	499798856	WP_011479590	<i>Methylobacillus flagellatus</i>
12	499425082	WP_011112546	<i>Nitrosomonas europaea</i>
13	499953628	WP_011634362	<i>Nitrosomonas eutropha</i>
14	499699341	WP_011380075	<i>Nitrosospira multiformis</i>
15	499308001	WP_010998776	<i>Nostocaceae</i>
16	499807020	WP_011487754	<i>Paraburkholderia xenovorans</i>
17	499782104	WP_011462838	<i>Rhodoferrax ferrireducens</i>
18	451782402	AGF53371	<i>Synechocystis sp. PCC 6803</i>
19	499631839	WP_011312573	<i>Thiobacillus denitrificans</i>
20	74056688	AAZ97128	<i>Thiobacillus denitrificans ATCC 25259</i>
21	503413850	WP_013648511.1	<i>Nitrosomonas sp. AL212</i>
22	504622405	WP_014809507	<i>Desulfomonile tiedjei</i>
23	488191742	WP_002262950	<i>Streptococcus mutans</i>
24	504267776	WP_014454878	<i>Spirochaeta africana</i>
25	488143381	WP_002214589	<i>Yersinia pestis</i>
26	503550450	WP_013784526	<i>Alteromonas naphthalenivorans</i>

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

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

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

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

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

165	357145851	XP_003573789	<i>Brachypodium distachyon</i>
166	357462397	XP_003601480	<i>Medicago truncatula</i>
167	357473317	XP_003606943	<i>Medicago truncatula</i>
168	357495273	XP_003617925	<i>Medicago truncatula</i>

## REFERENCE 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 adenyltransferase (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 Research*, 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 research and 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., *Cloning, 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 microsporoc 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 Díez, 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 Díez, 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 Rhodospseudomonas 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 adenylyl 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. Cifuentes, 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.