MECHANISM OF ACTION AND BINDING MODE REVEALED BY THE CRYSTAL STRUCTURES OF KEY ENZYMES IN PLANTS' SUGAR METABOLIC PATHWAY

CHUA TECK KHIANG BSc (Hons)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY AT THE NATIONAL UNIVERSITY OF SINGAPORE

**DEPARTMENT OF BIOLOGICAL SCIENCES,
FACULTY OF SCIENCE,
NATIONAL UNIVERSITY OF SINGAPORE**

APRIL 2008

For Cherlyn

Acknowledgements

I am grateful to my supervisor, Dr. J. Sivaraman, for an opportunity to persue research in a field of my interest. His patience, perserverance and dedication to science are invaluable lifetime lessons.

I would like to thank Prof. Bharat K Patel, Griffth University, Australia, for the clones of sucrose phosphate synthas (SPS) and fructokinase (FRK). I also thank Prof. Janusz Bujnicki (IIMCB, Warsaw, Poland) for his collaboration on the bioinformatics part of SPS.

I would also like to thank Dr Anand Saxena, Brookhaven National Laboratories (BNL) National Synchrotron Light Source, for assistance in data collection. I wish to thank Mr Shashikant Joshi for extending the Proteins and Proteomics Center facility. I also thank NUS for having given me the opportunity to pursue my Ph.D. with a research scholarship.

My special thanks to all my labmates for their warm and unflinching assistance. In particular, I wish to thank Cherlyn Ng, , Dr. Zhou Xingding, Dr. Li Mo, Sunita Subramanian, Dr Tan Tien Chye, Dr. Kumar

Sundramurthy, Lissa Joseph and for all the scientific/non-scientific discussions and for being such great friends and colleagues.

Table of Contents

Summary

This thesis reports the structure and function of two key enzymes that represents a valid model for the plant enzymes. Plant enzymes are relatively more difficult to isolate and characterize. The plant homologs of the two enzymes taken for this thesis work, namely Sucrose phosphate synthase (SPS) and Fructokinase (FRK), were particularly shown to be highly unstable and could not be characterized. This motivated us to take the *Halothermothrix orenii* as a model system for the plant enzymes to characterize the structure and function. *H. orenii* and plant enzymes share significant sequence homology. A detailed general introduction on the sugar metobolism enzymatic pathway is given in the first chapter.

Sucrose phosphate synthase (SPS; EC 2.4.1.14) catalyzes the transfer of a glycosyl group from an activated donor sugar such as uridine diphosphate glucose (UDP-Glc) to a saccharide acceptor D-fructose 6-phosphate (F6P), resulting in the formation of UDP and D-sucrose-6'-phosphate (S6P), a central and regulatory process in the production of sucrose in plants, cyanobacteria and proteobacteria. The second chapter reports the first crystal structure of SPS from *H. orenii*, and its complexes with the substrate F6P and the product S6P. SPS has two distinct Rossmann-fold domains, A- and B- domains, with a large substrate binding cleft at the interdomain interface. Structures of two complexes show that both the substrate F6P and the product S6P bind to the A-domain of SPS. The donor substrate, nucleotide diphosphate glucose (NDP-Glc), binds to the B-domain of SPS based on comparative analysis of the SPS structure with other related enzymes.

Fructokinase (FRK; EC 2.7.1.4) catalyzes the transfer of phosphate group from an ATP donor to a saccharide acceptor D-fructose resulting in the formation of D-fructose 6-phosphate (F6P). As an irreversible and near rate-limiting step, it is important for regulating the rate and localization of carbon usage by channelling fructose into a metabolically active state for glycolysis in plants and bacteria. The third chapter reports the crystal structure of FRK from *Halothermothrix orenii*, a first representative of any species structurally chracterized, and the possible mechanism of action. FRK possesses a β-sheet "lid" and an α/β (Rossmann-like) fold at its catalytic domain. FRK dimerization is through the lid domain and held in a β-clasp form.

The conclusions and future directions are provided in the fourth chapter. Our findings indicate that the *H. orenii* represent valid models of both plant SPSs and FRKs and thus provide useful insight into the reaction mechanism of the plant enzymes. As SPS has been implicated in stress response and food productivity, structure-based mutagenesis of SPS in plants may result in high yielding crops with greater resistance to osmotic fluctuations in the face of climate changes today.

List of Tables

List of Figures

List of Abbreviation

Å	angstrom (10^{-10} m)
ADP-Glc	adenosine diphosphate glucose
ATP	adenosine triphosphate
BME	β-mercaptoethanol
CAZy	Carbohydrate Active Enzymes database
CNS	crystallography and NMR system
DEAE	diethyl aminoethyl
DLS	dynamic light scattering
DTT	Dithiothreitol
E. coli	*Escherichia Coli*
EDTA	ethylenediamine tetraacetic acid
F6P	Fructose 6-phosphate
FPLC	fast performance liquid chromatography
FRK	Fructokinase
GDP-Glc	guanosine diphosphate glucose
GPGTF	glycogen phosphorylase glycosyltransferase
GT	Glycosyltransferases

H. orenii	*Halothermothrix orenii*
IPTG	isopropylthogalactoside
ITC	isothermal titration calorimetry
kDa	Kilodaltons
LB	Luria-Bertrani
MAD	multiwavelength anomalous dispersion
MALDI-TOF	Matrix Assisted Laser Desorption Ionization –Time of Flight
MS	mass spectrometry
NCBI	National Center for Biotechnology Information
NCS	non crystallographic restraints
nm	Nanometer
OD	optical density
ORF	open reading frame
OtsA	trehalose 6-phosphate synthase
PCR	polymerase chain reaction
PEG	polyethylene glycol
PolydIndx	polydispersity index
rmsd	root mean square deviation
rpm	rotations per minute

S6P	Sucrose 6-Phosphate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SPS	Sucrose Phosphate Synthase
SS	Sucrose Synthase
SeMet	seleno-L-methionine
TP	Triose phosphate

Amino acids

Ala (A)	Alanine
Arg (R)	Arginine
Asn (N)	Asparagine
Asp (D)	aspartic acid
Cys (C)	Cysteine
Gln (Q)	Glutamine
Glu (E)	glutamic acid
Gly (G)	Glycine
His (H)	Histidine
Ile (I)	Isoleucine
Leu (L)	Leucine
Lys (K)	Lysine
Met (M)	Methionine
Phe (F)	Phenylalanine
Pro (P)	Proline

Ser (S)	Serine
Thr (T)	Threonine
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
Val (V)	Valine

Publications

Mechanism of Action and Binding Mode Revealed by the Structure of Sucrose Phosphate Synthase from *Halothermothrix orenii*

Chua Teck Khiang, Janusz M. Bujnicki, Tan Tien Chye, Frederick Huynh, Bharat K Patel and J.Sivaraman (2008). The Plant Cell 20(4)**:**1059-1072.

Chapter I

General Introduction

1.1 Introduction

Plants harness energy from sunlight through a series of chemical reactions to be the earth's primary producers of food. These important reactions are catalysed by enzymes to which functional and structural characterization would greatly aid in increasing the productivity of food to cope with the increasing human population. Plant enzymes, however, are relatively difficult to isolate due to their instability in heterologous systems. Fortunately, these enzymes possess homologs in many bacterial systems that can be well-characterized. This motivated us to use *Halothermothrix orenii* as a model system for understanding plant enzymes through structural chacterization. *H. orenii* and plant enzymes share significant sequence homology. This thesis reports the structures and their derived catalytic mechanisms of two ubiquitous enzymes in all plants, sucrose phosphate synthase (SPS) and fructokinase (FRK), which represent valid models for their plant counterparts.

1.2 Carbon

Carbon is an essential element in all living organisms. About 1900 gigatons of carbon is present and continuously being exchanged between living and non-living components of the biosphere in a biogeochemical process called the carbon cycle. Inorganic carbon in the environment is unusable by organisms and needs to be converted into organic form first. Auxotrophs (e.g. plants) do this through an anabolic pathway called photosynthesis, using atmospheric carbon dioxide, water and sunlight:

$$6CO_{2(gas)} + 12\ H_2O_{(liquid)} + \text{photons} \rightarrow C_6H_{12}O_{6(aqueous)} + 6\ O_{2(gas)} + 6\ H_2O_{(liquid)}$$

There are two stages of photosynthesis. The light-dependant reaction is the first stage, where light energy and cholophyll are used in photophosphorylation and photolysis of water. Products from the light reaction are used in the next stage – known as the light-independant reaction or Calvin cycle, where carbon dioxide is reduced into sugars. The end products of photosynthesis are basic energy sources for all organisms as substrates of respiration, a process through which sugar is oxidized back into carbon dioxide to yield energy for growth and development.

1.3 Key Enzymes of Source and Sink Tissues of Plants

Most plant cells contain chloroplasts for the purpose of photosynthesis. The plant organs involved in carbohydrate production are known as source tissues (Figure 1.1). Most of the carbohydrate produced during photosynthesis converted to sucrose for transport to other areas for storage, growth and respiration. Plant organs that utilize the synthesized sucrose are known as sink tissues. SPS catalyses the production of sucrose-6-phosphate in source tissues, the final substrate in the sucrose synthesis pathway. FRK is a phosphotransferase at sink tissues; it produces fructose-6-phosphate from sucrose breakdown, an important initiation substrate in many catabolic pathways such as glycolysis.

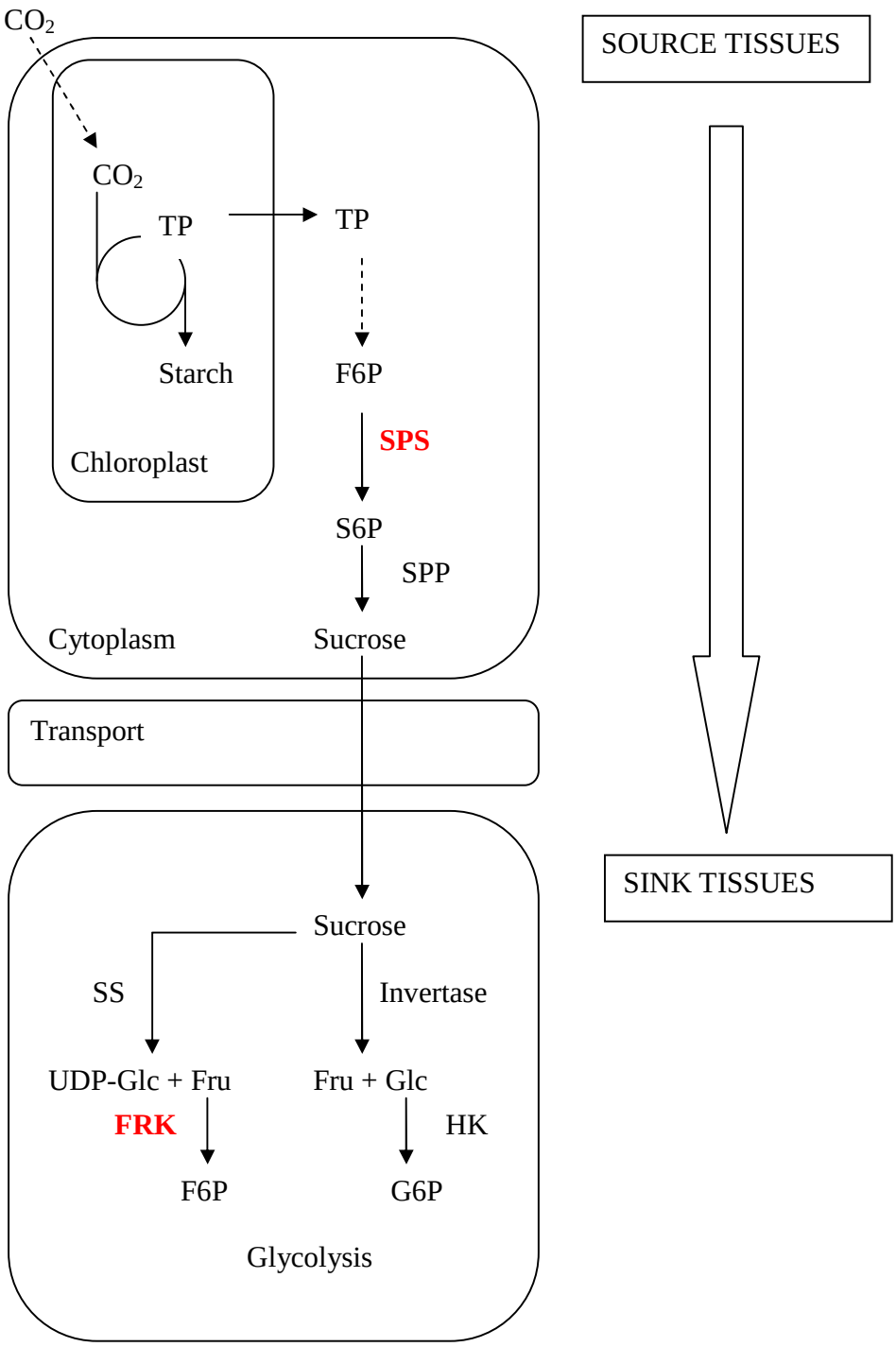


Figure 1.1 SPS and FRK roles in sugar metabolism in plants.

1.4 Sugar

Sugar (from the Sanskrit word *sharkara*) is a type of edible crystalline solid. Scientifically, sugar refers to any type of monosaccharide (simple sugar) or disaccharide. Monosaccharides (Greek: *mono* – 1; *sacchar* – sugar) are the basic building unit of carbohydrates. Examples of monosaccharides include glucose, fructose, galactose, ribose, xylose. Most monosaccharides self-cyclize between an alcohol group and a carbonyl group to form a ring structure (Figure 1.2). Carbon 1 (C1) is the carbon atom of the aldehyde group or the carbon atom immediately adjacent to a ketose group.

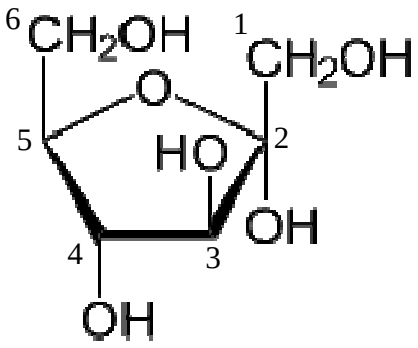


Figure 1.2 Haworth projection of fructose, a monosaccharide.
(http://en.wikipedia.org/wiki/Image:Beta-D-Fructofuranose.svg)

Disaccharides are sugar molecules with two monosaccharide units joined by a glycosidic bond in a condensation reaction between their respective hydroxyl groups. Sucrose (Figure 3) comprises of a fused glucose and fructose unit at a α(1→2) linkage, lactose of galactose and glucose in a β(1→4) linkage, maltose of two α(1→4) linked glucose entities. *alpha*- or *beta*- refers to the stereochemistry of the bond and (1→4) the carbon at which the linkage is formed.

Sugars are central compounds in nature that serve as essential metabolic nutrients and structural components for most organisms. They are also major regulatory molecules that control gene expression, metabolism, physiology, cell cycle, and development in prokaryotes and eukaryotes. In plants, it has been shown that sugars regulate the expression of a broad spectrum of genes involved in many essential processes. Furthermore, sugars affect developmental and metabolic processes throughout the life cycle of the plant. These processes include germination, growth, flowering, senescence, photosynthesis and sugar metabolism.

1.5 Sugar phosphates

Sugar phosphates are abundant in cells and important compounds in nature. They are intermediates common to pathways of synthesis and degradation and therefore the principle site at which pathways converge. Sugar phosphates are derived from breakdown of polysaccharides, photosynthesis and gluconeogenesis. Common examples are triose phosphate (TP), formed during photosynthesis and basic substrates for amino acid and complex carbohydrate synthesis. Glucose-6-phosphate (G6P) and glucose-1-phosphate (G1P) are the basic reactants in starch metabolism, and can be interconverted or converted to fructose-6-phosphate (F6P) by phosphoglucomutase, glucose-6-isomerase for oxidation through the glycolytic pathway. F6P itself is both a substrate and product of sucrose biosynthesis and hydrolysis respectively, while sucrose-6-phosphate (S6P) is an intermediate during synthesis of sucrose. Taken together, sugar phosphates form a pool from which intermediates can be drawn or added to.

1.6 Sucrose

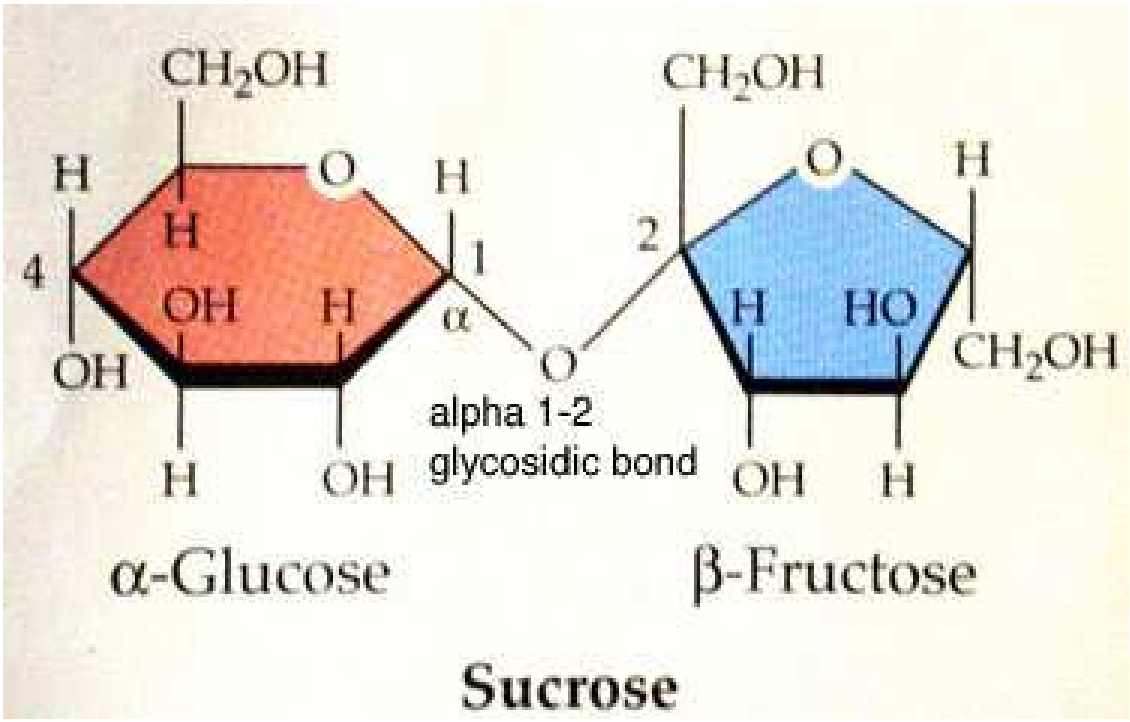


Figure 1.3. Molecular structure of sucrose. α(1→2) disaccharide formed by linking carbon atom 1 of glucose and carbon atom 2 fructose monosaccharides. (http://academic.brooklyn.cuny.edu/biology/bio4fv/page/disaccharide.html)

Sucrose is a α(1→2) disaccharide of glucose and fructose (Figure 1.3). It is solely formed by plants where it has three fundamental and interrelated roles. First, it is the principal product of photosynthesis and accounts for most of the CO_2 absorbed by a plant in this process. Secondly, sucrose is a major transportable metabolite through which carbon is translocated from source to sink tissues through plants' vascular system. Thirdly, sucrose is the main storage sugar in plants, serving as a main source of organic carbons for the synthesis of structural elements and the production of energy in future growth. Lastly, it acts as an osmolyte to prevent water loss in times of stress.

1.7 Sucrose synthesis

Most of the carbon needed for the production of sucrose originate from triose-phosphate molecules produced by the light-independent pathway of photosynthesis, when carbon dioxide is reduced by reacting with ribulose 1,5-bisphosphate to form two molecules of glycerate 3-phosphate. By using ATP and NADPH from the light dependant reactions, glycerate 3-phosphate is further reduced to triose phosphate. Triose phosphate is a three-carbon sugar. One out of six molecules produced will condense to form fructose 6-phosphate, which is then exported to the cytoplasm of a plant cell for sucrose synthesis. Only a small amount of ready-made hexose molecules, produced in the chloroplasts, are transported to the cytoplasm and are utilized for sucrose synthesis. The rest of TP molecules are recycled to form ribulose 1,5-bisphosphate (Figure 1.4).

The reaction following triose phosphate production occurs in the cytoplasm. The first step is the priming of glucose by glucose phosphorylase. This involves attaching a UDP moiety:

Glucose-1-phosphate + UTP ↔ UDP-glucose +PPi

The amount of F6P available is held in equilibrium by the interconversion of fructose-1,6-bisphosphate (FBP) and F6P through the action of three enzymes, which are also key regulatory points in the synthesis of sucrose. Cytosolic fructose-1,6-bisphosphatase (cyFBPase) produces F6P from FBP and is inhibited by fructose-2,6-bisphosphate. Conversely, phosphofructokinase catalyzes the backward reaction to FBP from F6P. Pyrophosphate:fructose-6-phosphate-1-phosphotransferase (PFP) is able to

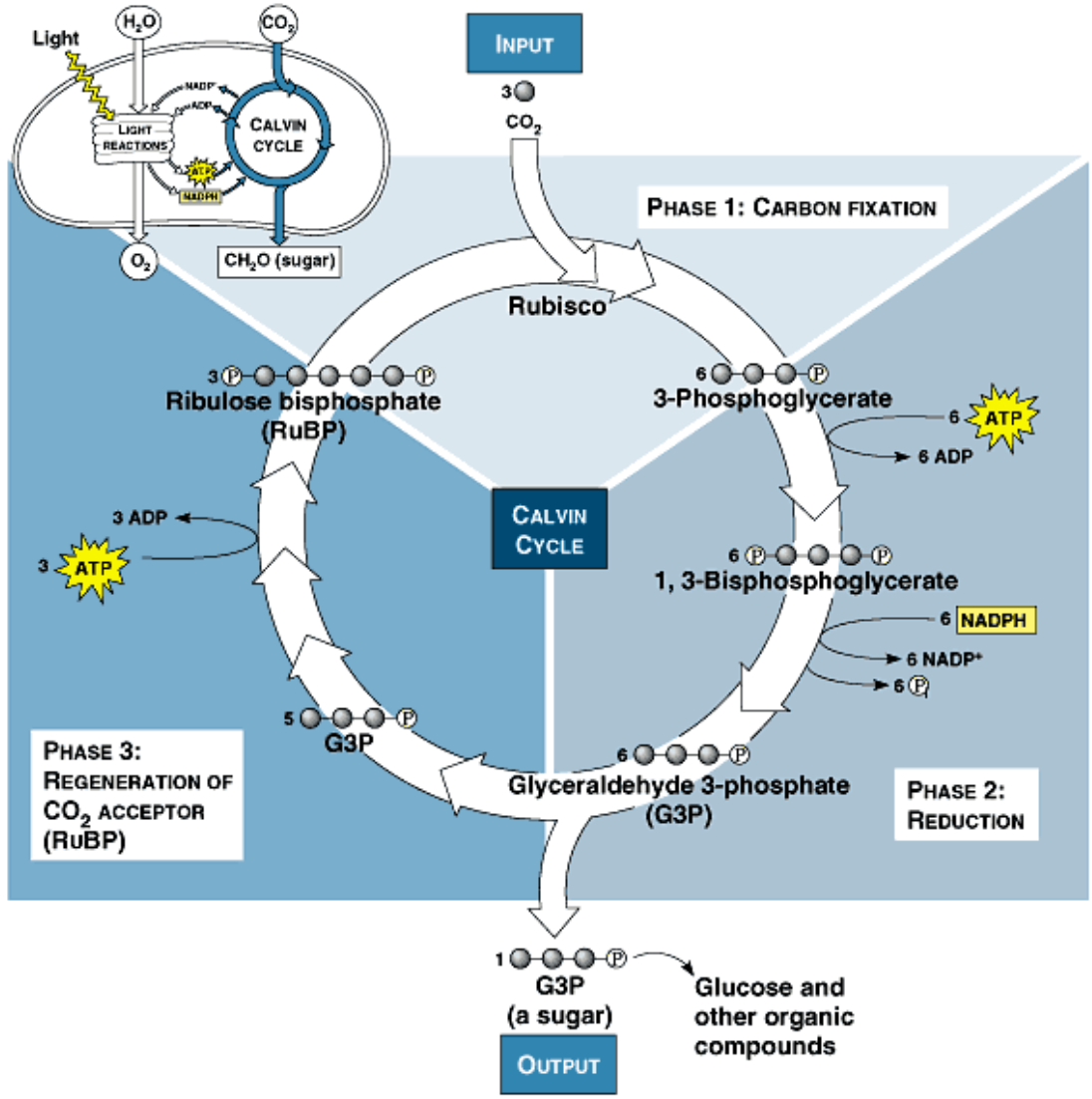


Figure 1.4. The light-independent pathway of photosynthesis

(http://www.msu.edu/~smithe44/calvin_cycle_process.htm).

drive the reaction either way, and the synthesis of F6P is stimulated by fructose-2,6-bisphosphate.

Sucrose phosphate synthase (SPS; E.C. 2.4.1.14) next catalyses the first step in the pathway of sucrose synthesis, by transferring a glycosyl group from activated donor sugar, uridine diphosphate glucose (UDP-Glc) to a sugar acceptor D-fructose 6-phosphate (F6P), resulting in the formation of UDP and S6P:

UDP-glucose + F6P ↔ S6P + UDP (SPS)

Finally, a dephosphorylation of S6P to sucrose by sucrose phosphatase (SPP; E.C. 3.1.3.24) concludes the sucrose biosynthesis pathway. As a large free energy change occurs during this process, the forward reaction is irreversible.

S6P +H2O → sucrose + Pi (sucrose phosphatase)

In an alternative pathway, sucrose synthase is able to bypass the need for S6P and synthesize sucrose directly from NDP-glucose and fructose:

NDP-glucose + D-fructose <=> NDP + sucrose (sucrose synthase)

1.8 Sucrose and environmental stress

In 1979, Munn and co-workers observed that when *Triticum aestivum* was subjected to water stress, its floral apex exhibited an initial elevation of sucrose levels,

followed by increase in amino acid levels (Munn *et al.*, 1979). Subsequently, similar observations were made in other plants (Hubac and Da Silva, 1980); when *Chlorella* cells were plasmolysed by steep increase in sucrose concentration, the rate of sucrose synthesis increased.

This increase was sufficient for a partial restoration of the osmotic volume of the cells. (Greenway and Munns, 1980). It is thus known today that sucrose contributes to osmotic adjustments in a plant and reduces tissue damage to enhance survivability when loss of turgor occurs. In plants surving winter, sucrose contributes to tissue cryo-protection against frost, and sugar content is proportional to freezing tolerance of tissue (Levitt, 1980).

Interestingly, the halophilic bacteria *Dunaliella* has elevated sucrose production in the dark at elevated temperatures when glycerol, its natural osmolyte is used for production of hexose phosphates (Muller and Wegmann, 1978; Wegmann, 1979; Wegman *et al.*, 1980), suggesting the intimate link between the sugar metabolic pathway and osmolytic homeostasis.

1.9 Fate of synthesized sucrose

The rate of sucrose synthesis increases with the rate of photosynthesis. In photosynthetic tissues, sucrose is predominantly exported from cells, most probably by facilitated diffusion and subsequently taken up by the phloem complex by a specific, active sucrose/H^+ co transport mechanism. Once in the phloem complex sucrose is transported to cells in the sink tissues. At least two distinct classes of sink tissues can be distinguished: (1) tissues that are highly metabolically active such as rapidly growing

tissues and (2) tissues that are for storage purposes. Accordingly, the sucrose that arrives will be either used for respiration or starch synthesis.

Sucrose delivered to the sink tissues is cleaved by two mechanisms. In apoplast, cytosol or the vacuole, invertase (EC 3.2.1.26) cleaves sucrose to glucose and fructose; sucrose synthase (SS; EC 2.4.1.13) hydrolyses sucrose to UDP-Glc and fructose in (Keller et al, 1988), tonoplast (Etxeberria E and Gonzalez P, 2003) or inassociation with the plasmalemma (Amor et al., 1995; Carlson and Chourey, 1996). Through either pathway, half of the carbon imported as sucrose into the sink tissues is converted to free fructose, which is phosphorylated and channeled into other pathways.

1.9.1 Starch synthesis

Starch is the dominant storage polysaccharide in plants and an important metabolic substrate in both plants and many herbivores. It is present in all major organs of higher plants, accounting for 65 -75% dry weight of cereal grains and 80% of potato tubers. It is a major immediate product of photosynthesis from sucrose and mobilized in the dark by hydrolysis back to sucrose and transported to respiring tissues.

Starch is a polymer of repeating glucose units; all fructose units derived from the breakdown of incoming sucrose must therefore first be converted to G6P by G6P isomerase. Phosphoglucomutase then transfers the phosphate group from C6 to C1 to produce glucose-1-phosphate (G1P). In the presence of ATP, ADP-glucose phosphorylase catalyses the formation of ADP-glucose and releases an inorganic phosphate in the process. Glucose monomers from invertase action can join the pathway through phosphorylation, while UDP-Glu from SS can be utilized directly. Fructose

requires an additional phosphorylation step by fructokinase (Frk) before initiation into the pathway. G1P and ADP-glucose are the substrate for starch synthase to produce amylase; branching enzyme later synthesizes amylopectin. Together, amylose and amylopectin are known as starch.

1.9.2 Glycolysis

Glycolysis is the initial pathway of carbohydrate oxidation (Figure 1.5). It serves three functions: The generation of high-energy molecules (ATP and NADH) as cellular energy sources as part of aerobic respiration and anaerobic respiration; that is, in the former process, oxygen is present, and, in the latter, oxygen is absent, production of pyruvate for the citric acid cycle as part of aerobic respiration and the production of a variety of six- and three-carbon intermediate compounds, which may be removed at various steps in the process for other cellular purposes.

Glycolysis, through anaerobic respiration, is the main energy source in many prokaryotes, eukaryotic cells devoid of mitochondria (e.g., mature erythrocytes) and eukaryotic cells under low-oxygen conditions (e.g., heavily-exercising muscle or fermenting yeast). It is a catabolic process that takes place in the cytosol and drains the hexose phosphate (specifically F6P) pool. F6P is an important compound in glycolysis because, contary to starch synthesis, all glucose units must be converted to F6P before proceeding. The first committed and rate limiting step converts F6P to F1,6P using ATP as a phosphate donor, through the synchronized action of phosphofructokinase and pyrophosphate:fructose-6-phosphate phosphotransferase. F1,6P is then broken into two molecules of glyceraldehyde 3-phosphate by aldolase. During the phosphorylation of

glyceraldehyde 3-phosphate to 1,3 diphosphoglycerate by pyrophosphate dependant phosphofructokinase using pyrophosphateas a phosphate donor, two molecules of NAD are reduced to NADH. Subsequently in the production of 3-phosphoglycerate, two molecules of ATP are released from the enzymatic transfer of a phosphate group from 1,3-bisphosphoglycerate to ADP by phosphoglycerate kinase. In the final steps of glycolysis, enolase and pyruvate kinase sequentially forms phosphoenolpyruvate and pyruvate respectively.

1.10 Sugar phosphorylation in sucrose catabolism

Phosphorylation of free monosaccharides (glucose and fructose) is not only the initial step of metabolic pathways but also essential for the mobilisation of all hexoses taken up by the cell for downstream processes. Phosphorylation traps a sugar in the cell and furthermore, feedback inhibition by free fructose on sucrose synthase prevents further hydrolysis of sucrose. Therefore, removal of free fructose by phosphorylation helps in establishing sink strength of the tissue and facilitates the formation of a sucrose gradient between the phloem and cells in the sink. A majority of the glucose and fructose phosphorylating activities are thought to be present in the cytosol or associated with the mitochondrial and plastid membranes. Two enzymes are responsible for phosphorylation of sucrose cleavage products fructose and glucose: Fructokinase (FRK; EC 2.7.1.4) catalyzes the transfer of a phosphate group from adenosine triphosphate (ATP) donor to a saccharide acceptor D-fructose resulting in the formation of D-fructose 6-phosphate (F6P). Hexokinases (Hxk) preferentially phosphorylates glucose (Figure 1.1).

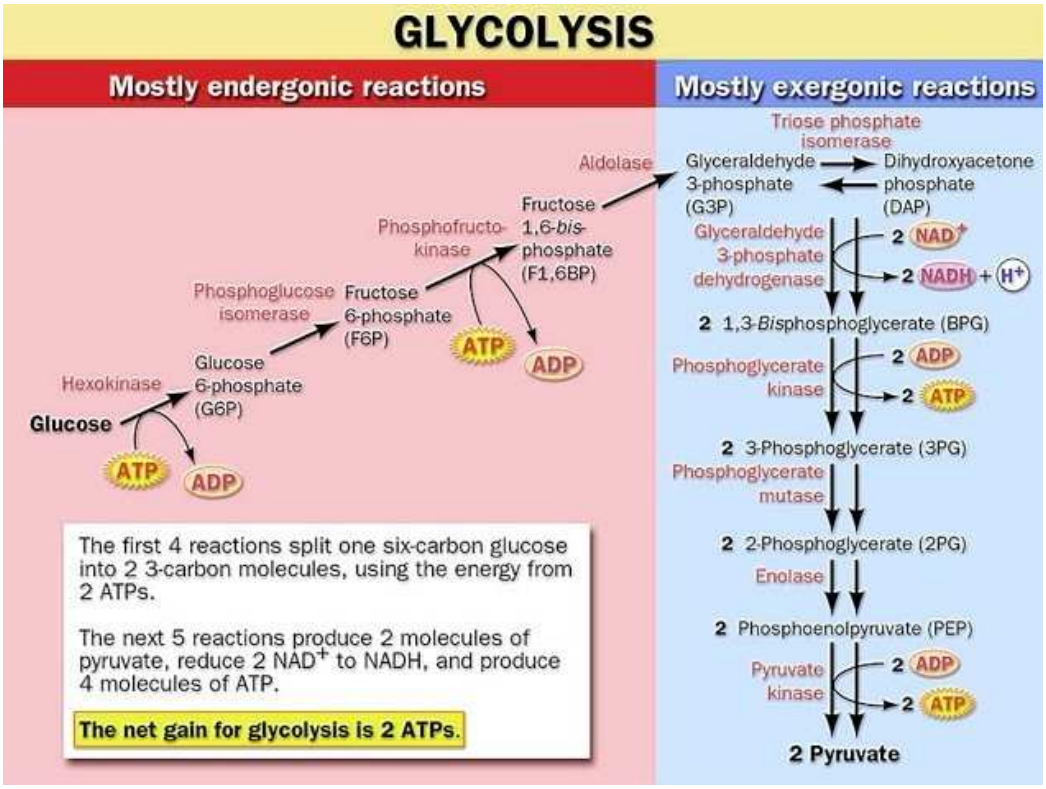


Figure 1.5. The glycolytic pathway.
http://www.biologyclass.net/glycolysis.jpg

1.11 Sugar kinases

Based on sequence and structural classifications, there are three superfamilies of sugar kinases responsible for the phosphorylation of all sugars in a cell.

1.11.1 Hexokinase superfamily

The hexokinase superfamily members represent a class of enzymes that possess an ATPase domain with same basic fold and active site as actin and Hsp70 of the heat shock proteins. There are two distinct domains: the N-terminal domain has a regulatory function and C-terminal catalytic. Members of this family include eukaryotic hexokinases and glucokinases, prokaryotic glucokinase, gluconokinase, xylulokinase, glycerol kinase, fructokinase, rhamnokinase and fucokinases.

1.11.2 Galactokinase superfamily

The galactokinase superfamily is still structurally uncharacterized. However, other sequence studies have shown that all members of this family share common motifs. This family consists of mevalonate kinase and a functionally unrelated homoserine kinase.

1.11.3 Ribokinase superfamily (also known as pfkb family in Prosite sequence collection)

The ribokinase superfamily of proteins consists of fructokinases, *E. coli*'s minor 6-phosphofructokinase, 1-phosphofructokinase, 6-phosphotagatokinases, *E. coli* inosine-guanosine kinase. Following the structure determination of ribokinase (Sigrell *et al.*, 1998), many members of the ribokinase superfamily have been solved to-date. Namely,

THZ kinase (Campobasso er al, 2000), HMPP (Cheng *et al.*, 2002), pyridoxal kinase (Li *et al.*, 2002), AIRs kinase / KDG kinase (Zhang *et al.*, 2004), adenosine kinase (Schumacher *et al.*, 2000) and glucokinase (Ito *et al.*, 2001). In addition, two kinases of unknown function (PDB codes: 1KYH and 1O14) have been identified as part of this superfamily based on their structure, active sites residues, monomer topology, and quaternary structure (Zhang *et al.*, 2004). It was found that the catalytic portion of these enzymes possess a Rossman fold similar to other nucleotide binding proteins.

1.12 *Halothermothrix orenii*

H. orenii is an anaerobic, thermohalophilic bacterium from the class Clostridia. It is found in the sediment of a Tunisian salted lake as a long rod, present only in the 40- to 60-cm layer below the surface. The strain isolated, H168, produced acetate, ethanol, H2, and CO2 from glucose metabolism. Fructose, xylose, ribose, cellobiose, and starch were also oxidized. The optimum temperature for growth was 60º C. No growth was obtained at 42 or 70º C. Strain H168 grew optimally in NaCl concentrations ranging from 50 to 100 g per liter, with the upper and lower limits of growth around 200 and 40 g per liter, respectively. The G+C ratio of the DNA was 39.6 mol%. The phylogeny, physiology, morphology, lipid content, and high G+C content of strain H168 are sufficiently different from those of genera belonging to the family Haloanaerobiaceae to justify the definition of a new genus. The SPS and FRK open reading frames (ORF) were identified in the course of a random sequence analysis of the *H. orenii* genome (Mijts and Patel, 2001). The following chapters of this thesis report the structures and catalytic mechanisms of SPS and FRK, which represent valid models for their plant counterparts.

Chapter II

Mechanism of Action and Binding Mode Revealed by the Structure of Sucrose Phosphate Synthase from *Halothermothrix orenii*

2.1 Introduction

Enzymes sucrose phosphate synthase (SPS; E.C. 2.4.1.14) and sucrose phosphatase (SPP; E.C. 3.1.3.24) are involved in the synthesis of sucrose, a process that is believed to be restricted to plants, cyanobacteria (bacterial ancestors of the plant chloroplasts; Cumino *et al.*, 2002) and some proteobacteria (Lunn, 2002). SPS is a ubiquitously expressed enzyme in plants and green algae. It catalyses the first step in the pathway of sucrose synthesis, by the transfer of a glycosyl group from an activated donor sugar such as uridine diphosphate glucose (UDP-Glc) to a sugar acceptor D-fructose 6-phosphate (F6P), resulting in the formation of UDP and D-sucrose-6'-phosphate (S6P) (Figure 2.1). This upstream, reversible reaction is followed by an irreversible reaction by SPP resulting in the dephosphorylation of S6P to sucrose, which concludes the sucrose biosynthesis pathway.

SPS is proven to be the only enzyme responsible for the formation of S6P (and ultimately, sucrose) from UDP-glucose and F6P, it therefore has major role in the control of sucrose production in leaves. Firstly, there is a close correlation between the rate of sucrose synthesis and the extractable activity of SPS (Stitt *et al.*, 1987). Secondly, three- to seven-fold over-expression of maize SPS in transgenic tomato plants results in a small, but significant increase in leaf sucrose synthesis (Frommer and Sonnewald, 1995). Thirdly, the known regulatory properties of SPS are entirely consistent with this enzyme having an important role in the regulation of sucrose synthesis. High SPS activities found in leaves are subjected to complex regulatory controls involving:

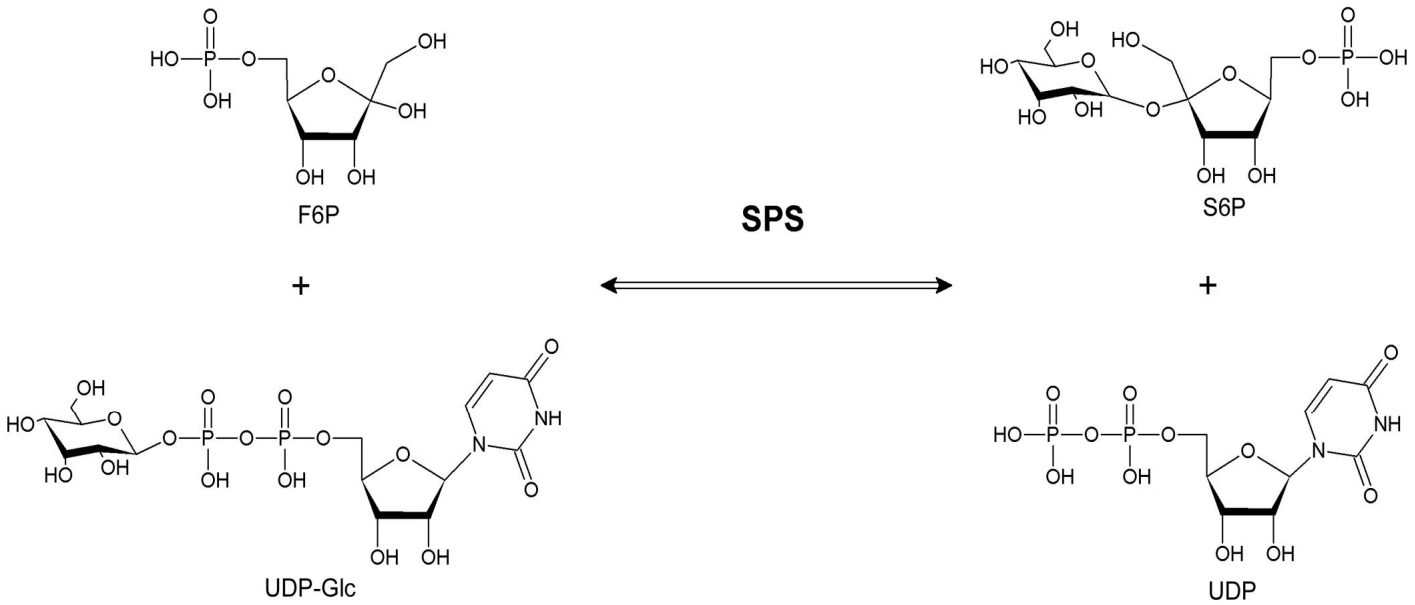


Figure 2.1 A schematic diagram of the reaction involving SPS and F6P. The synthesis of S6P involves the action of SPS (EC 2.4.1.14), which catalyzes the transfer of a glycosyl group from an activated donor sugar such as UDP-Glc to a saccharide acceptor F6P, resulting in the formation of UDP and S6P, a central and regulatory process in the production of sucrose in plants and cyanobacteria.

1) Metabolite regulation. Spinach SPS is subjected to metabolite dependent post-translational modification (Huber *et al.,* 1989) involving allosteric activation by G6P and inhibition by Pi. Divalent cations such as Mn^{2+} or Mg^{2+} has also been shown activate the enzyme. while UDP competitively inhibits activity with UDP-glucose.
2) Protein phosphorylation. SPS phosphorylation was originally characterized as the mechanism underlying light/dark modulation of SPS activity. There are two kinetically distinct forms of SPS that differ in substrate affinities, sensitivity to inhibition by Pi and activation by G6P: the dephosphorylated (active) and the phosphorylated (inactive) form. Multi-site Seryl phosphorylation: pSer158 reduced F6P and G6P affinity in spinach (McMichael *et al.,* 1993). S158E mutant constitutively deactivated: negative charge responsible for regulating activity – may be involved in activation of SPS in response to stress (Toroser and Huber, 1997). More recently, phosphorylation of SPS has also been implicated in the activation of the enzyme that occurs when the leaf tissue is subjected to osmotic stress.
3) Molecular genetic regulation of gene expression and steady state enzyme protein contents, such as photosynthetic light conditions and osmotic stress that result in changes to endogeneous hormonal factors regulating SPS steady state level. In soybean and spinach, artificial addition of gibberellic acid (GA) upregulated the expression of SPS protein (Cheikh and Brenner, 1992 Cheikh *et al.,* 1992; Walker and Huber, 1989).

The SPS from the photosynthetic cyanobacteria *Anabaena* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 (Lunn *et al.,* 1999, Porchia and Salerno, 1996) has been characterized and its respective putative SPS genes have also been identified in several

other cyanobacterial species, including *Synechococcus* sp. WH 8102 and *Prochlorococcus marinus* (Lunn, 2002). The functional and physiological role of the SPS gene in these photosynthetic prokaryotes, however, is unknown, and it has been speculated that, like in plants, the SPS may play a role in adaptation to osmotic stress. The presence of SPS in prokaryotes suggests that sucrose synthesis is an ancient trait (Cumino *et al.*, 2002, Lunn *et al.*, 1999). The recent identification of a putative SPS in *Halothermothrix orenii*, a non-photosynthetic prokaryote, provided a possibility to answer questions about the molecular and physiological role of SPS enzymes.

H. orenii is an anaerobic, thermohalophilic bacterium from the class Clostridia, with an optimum condition of growth at temperature 60°C in 10% NaCl (Cayol *et al.*, 1994). An open reading frame (ORF) has been identified as SPS in the course of a random sequence analysis of the *H. orenii* genome (Mijts and Patel, 2001). The recombinant *H. orenii* SPS exhibits cross-reactivity with polyclonal antibodies raised against plant SPSs (AgriSera, Sweden) suggesting antigen conservation among the SPSs of bacteria and plants (Huynh *et al.*, 2005).

In this chapter we report the crystal structure of the first SPS from *H. orenii* in the apo form, as well as complexes with the substrate F6P and the product S6P refined at 1.8, 2.8 and 2.4 Å resolutions, respectively. The report on *H. orenii* SPS provides insight into structure and function of SPS from cyanobacteria and plants with which it shares a close similarity. Based on comparative analysis of previously published structures of other GT enzymes, we propose a mechanism for the transfer of the glycosyl group by SPS from NDP-Glc to F6P, leading to the formation of S6P.

2.2 Material and Methods

2.2.1 Cloning, expression and purification.

Primers containing BamH1 and Kpn1 restriction sites at the 5' and 3' ends respectively were used in PCR to amplify the spsA gene. The PCR product was digested by these restriction enzymes, followed by its ligation with the pTrcHisA expression vector (Invitrogen) encoding an N-terminal, non-cleavable His_6 tag (Mijts and Patel, 2001). The plasmid was transformed into BL21 (DE3) and grown in 1 L of LB broth with 0.1mM Ampicillin at 37°C until it reached an optical density (OD_{600nm}) of about 0.6-0.7. The culture was cooled down and induced with 1mM IPTG overnight at 25°C. The *H. orenii* SPS has 499 amino acid residues with a molecular weight of 56.815 kDa. The recombinant *H. orenii* SPS, consisting of a hexahistidine tag and a linker, is expressed as a 61.1 kDa protein. The cells were harvested by centrifugation (9000g; 30min, 4°C) and resuspended in 30 ml of 20mM Tris-HCl pH 7.5, 200mM NaCl and 10mM imidazole and 1 tablet of EDTA-free Complete™ Protease Inhibitor Cocktail (Roche Diagnostics). Selenomethionine-substituted SPS was expressed using methionine auxotroph *E.coli* DL41 in LeMaster medium supplemented with 25mg/L selenomethionine (SeMet). The cells were lysed by sonication, followed by centrifugation at 11000rpm (Eppendorf 5804R) for 30min. Cell lysate was transferred to a chromatography (affinity) column containing Ni-NTA agarose (Qiagen). 1h of incubation was performed at 25°C with gentle agitation. The non-cleavable His_6-tag SPS was eluted with 500mM imidazole following three wash steps to remove non-specific binding. In the 12.5% SDS-PAGE viewed by Coomassie staining, the purified SPS migrated as a single band (Figure 2.5) just between the 66.2kDa and the 45kDa of the protein ladder (SDS-PAGE Molecular Weight Standard, Low-range by BioRad). The recombinant SPS was further purified

using FPLC Hiload 16/60 Superdex200 gel filtration column using AKTA FPLC UPC-900 system (Amersham Biosciences). The recombinant SPS was eluted at the flow-rate of 0.5ml/min, as a single peak (Figure 2.6) between 75ml and 90ml, in a buffer containing 0.2M NaCl and 10mM dithiothreitol (DTT) in 20mM Tris-HCl at pH7.5. This was followed by ultrafiltration to bring to a final concentration of the recombinant SPS to 10 mg/ml (Bradford method, Bradford, 1976).

2.2.2 MALDI-TOF analysis.

The native and SeMet-substituted SPS was further analyzed for the incorporation of selenium on a Voyager STR MALDI-TOF mass spectrometer (Applied Biosystems) by comparing the experimentally measured molecular weight of the native SPS with that of the SeMet protein and confirmed the proper incorporation of selenium (Figure 2.7).

2.2.3 Dynamic Light Scattering (DLS).

Dynamic light scattering measurements were performed at room temperature by a DynaPro (Protein Solutions) DLS instrument (Figure 2.8). The homogeneity of native SPS and SeMet-SPS was monitored during the various stages of concentration steps to avoid aggregation, prior to crystallization. The percentage of polydispersity was 14.1% for all protein samples at about 10 mg/ml.

2.2.4 Isothermal Titration Calorimetry (ITC).

ITC experiments were carried out by a VP-ITC calorimeter (Microcal, LLC) using 0.01-0.02 mM of the SPS in the sample cell and 0.1-0.2mM of F6P in the injector (Figure 2.9). Injection volumes of 4-5 μL each were used and the number of injections

was 60. The heat of dilution for each ligand was measured differentially with the reference cell as a control titration experimental runs for the protein. Consecutive injections were separated by time duration of at least 4 mins to allow the peak to return to the baseline. The ITC data was analyzed using a single site fitting model using Origin 7.0 (OriginLab Corp.) software.

2.2.5 Crystallization.

Initial crystallization conditions were screened at 25°C by hanging-drop vapor-diffusion technique using Hampton Research (Aliso Viejo, CA, USA) crystallization screens and by micro batch under-oil technique using JB crystallization screens (Jena Biosciences, Jena, Germany). Initially, apo and SeMet SPS crystals were plate-like and were obtained after 2 days directly from JB3 screen C2. Further optimization with extensive additive screens (Hampton Research) for best diffraction quality crystals was obtained by hanging-drop vapor-diffusion method using a reservoir solution containing 20% PEG 4000, 0.6M NaCl and 0.1M Na MES pH6.5 with a drop size 1µl of the reservoir solution with 1µl of protein against 1 ml reservoir solution. Crystals had approximate dimensions of 0.45 × 0.3 x 0.2 mm (Figure 2.10). They diffracted up to 1.8 Å and belonged to space group C2 with a = 154.23, b=48.50, c =75.05 Å and β=100.92º.

2.2.6 Data collection, structure solution and refinement.

Crystals were directly taken from the drop, and flash cooled in a N2 cold stream at 100ºK. The apo-SPS crystals were diffracted up to 2.4 Å resolution using an R-axis IV++ image plate detector mounted on a RU-H3RHB rotating anode generator (Rigaku Corp., Tokyo, Japan). Synchrotron data were collected at beam lines X12C and X29,

NSLS, Brookhaven National Laboratory for the SeMet protein (Figure 2.11). Complete MAD datasets were collected at three wavelengths (Table 2.1) using Quantum 4-CCD detector (Area Detector Systems Corp., Poway, CA, USA) to 1.8 Å resolution. Data was processed and scaled using the program HKL2000 (Otwinowski and Minor, 1997).

2.2.7 Structure solution and refinement.

Out of the seven expected selenium sites in the asymmetric unit, five were located by the program SOLVE (Terwillinger and Berendzen, 1999). The N terminal, as well as the C terminal methionine, was disordered. Initial phases were further developed by RESOLVE (Terwillinger, 2000) and improved the overall figure of merit (FOM) to 0.73 which made it possible to build automatically approximately 70% of the molecule. The remaining parts of the model were built manually using the program O (Jones *et al.*, 1991). Further cycles of model building alternating with refinement using the program CNS (Brunger *et al.*, 1998) resulted in the final model, with an R-factor of 0.226 (R_{free}=0.252) to 1.8 Å resolution with reflections I>σI was used in the refinement. The final model comprises of 455 residues (Ile7-Arg462) and 287 water molecules. The N terminal His tag with the linker residues and the C-terminal 32aa were not visible in the electron density map. PROCHECK (Laskowski *et al.*, 1993) analysis shows two residues in the disallowed regions of the Ramachandran plot.

2.2.8 F6P-SPS and S6P-SPS complexes

Two complexes F6P-SPS and S6P-SPS were obtained by soaking crystals of apo-SPS respectively in 20mM F6P and 20mM S6P for 12 hours at 25ºC. Complete datasets of both complexes were collected on an R-axis IV++ area detector with an RU300

rotating anode generator as the X-ray source and diffracted to 2.8 Å and 2.4 Å respectively. Crystals were cryo-protected as described above. The apo-SPS model used to calculate the difference electron density maps revealed the presence of ligands. Two models were refined with CNS (Brunger *et al.*, 1998), combined with manual refitting with the program O and appropriate entries were made in their respective dictionaries. Both F6P-SPS and S6P-SPS complex models consist of residues from Ile7 to Arg462 with 312 and 294 water molecules respectively. The simulated annealing *Fo-Fc* omit map of the ligands are shown in Figures 6a and 6b. Crystallographic statistics are presented in Table 2.1.

2.2.9 Bioinformatics analyses.

Sequence database searches were carried out with PSI-BLAST (Altschul *et al.*, 1997). Sequences of SPS homologs were clustered using CLANS (Frickey and Lupas, 2004) and genuine members of the SPS family were aligned using CLUSTALX. Phylogenetic analyses were done with MEGA 3.1 (Kumar *et al.*, 2004), using the minimum evolution method, JTT matrix, and pair wise deletion of gaps. Protein structure prediction for sequence segments present in SPS homologs from other species but missing from the *H. orenii* SPS was carried out via the GeneSilico metaserver (Kuroski and Bujnicki, 2001). Docking of flexible ADP and UDP structures to SPS was carried out using FlexX (Kramer *et al.*, 1999) with default parameters. Ten top-scoring poses were considered.

2.2.10 Protein Data Bank accession code.

Coordinates and structure factors for the apo, F6P and S6P complexes have been

deposited with RCSB Protein Data Bank (PDB) with code 2R60, 2R66 and 2R68 respectively.

2.3 Results and Discussion

2.3.1 Sequence Analysis

Sequence database searches also revealed a large family of sequences similar to *H. orenii* SPS. Top six homologs of the *H. orenii* SPS exhibit sequence identities varying from 54% for *Petrotoga mobilis* SJ95 to 33% for *Synechocystis* sp. PCC 6803. Similarly, six closest plant SPS homologs show sequence identities of approximately 32% (Figure 3a). Further analysis by CLANS (Frickey and Lupas, 2004) to cluster all sequence homologs into families according to the BLAST sequence similarity P-value is shown in Figure 2.2. SPS sequences form a well-defined group, whose nearest neighbors are sucrose synthases (SS; mostly from plants) and bacterial glycogen synthases. Other homologs, including starch (bacterial glycogen) synthases, are more remotely related. These phylogenetic relationships suggest that the SPS (as well as SS) have originated in Bacteria, and were transferred to plants via the chloroplast endosymbiont. The phylogenetic tree of the SPS family (Figure 2.3) shows several well-resolved branches, among which only green plants are monophyletic, while others comprise genes from cyanobacteria, as well as from diverse other species, suggesting multiple horizontal gene transfers. Thus, SPS from *H. orenii* also appears to have been derived by horizontal gene transfer from cyanobacteria and thus is a member of a sister group of plant SPS enzymes.

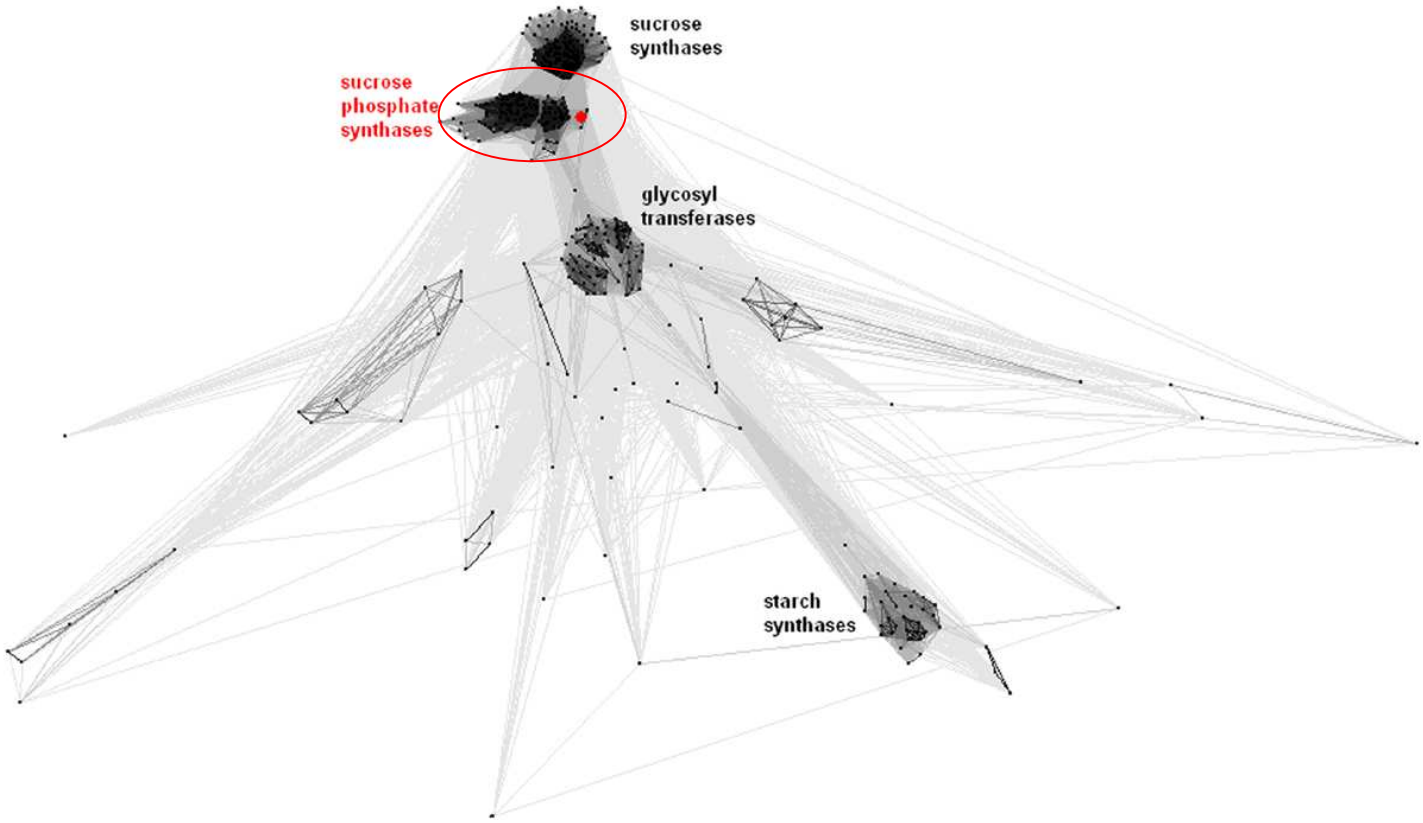


Figure 2.2 Sequence similarity between SPS and its homologs, as visualized with CLANS. Points indicate sequences; lines indicate similarities according to the BLAST P-value (the darker, the more similar). *H. orenii* SPS is shown as a big red dot. Individual families are labeled. All SPS sequences are encircled in red. Sequence similarity is visualized with CLANS (Fricky and Lupas, 2004).

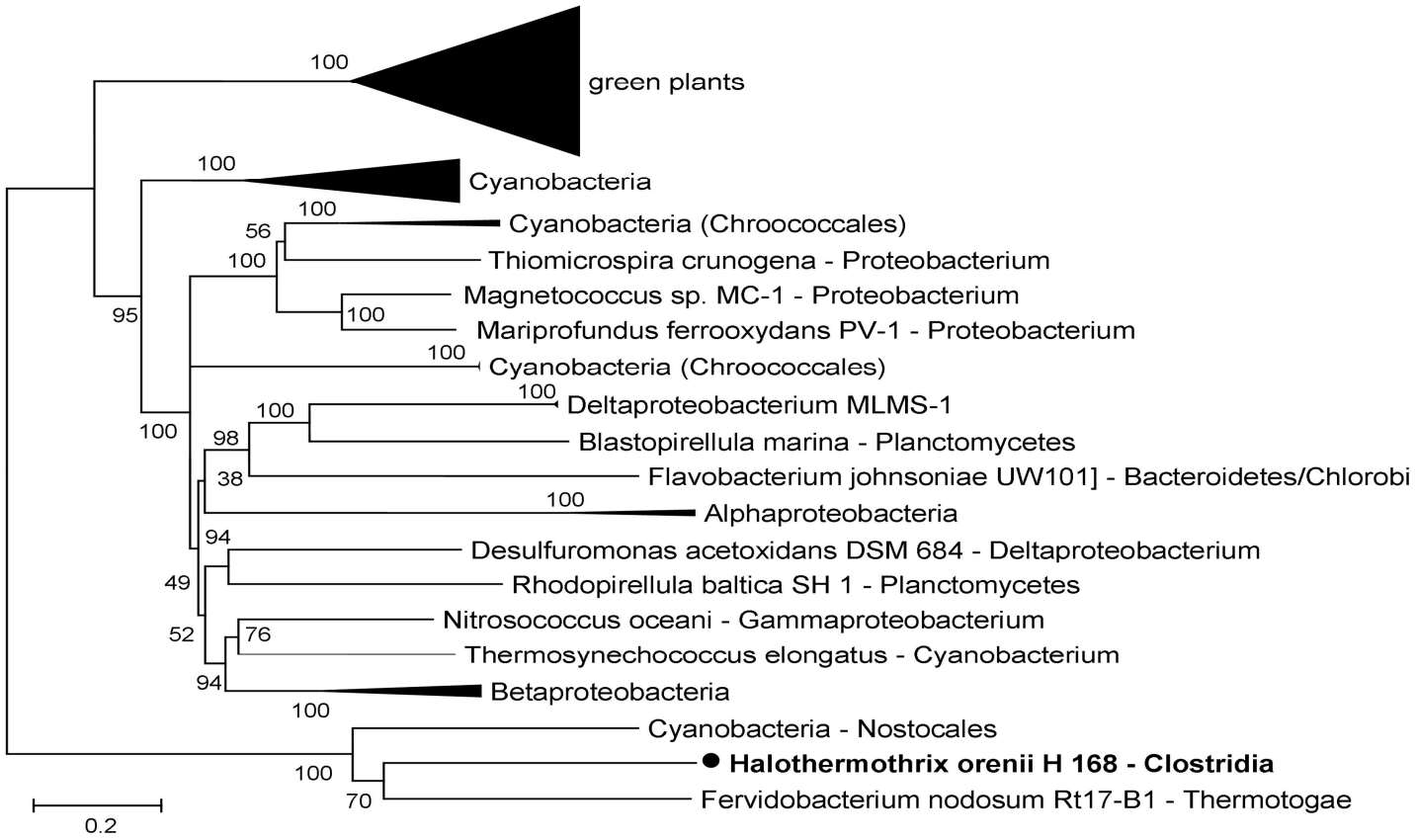


Figure 2.3 Phylogenetic tree of the SPS family. Monophyletic branches are collapsed, shown as triangles and labeled with the taxon's name. Statistical support for individual branches is shown in percent values, according to the interior branch test. *H. orenii* sequence is indicated in bold.

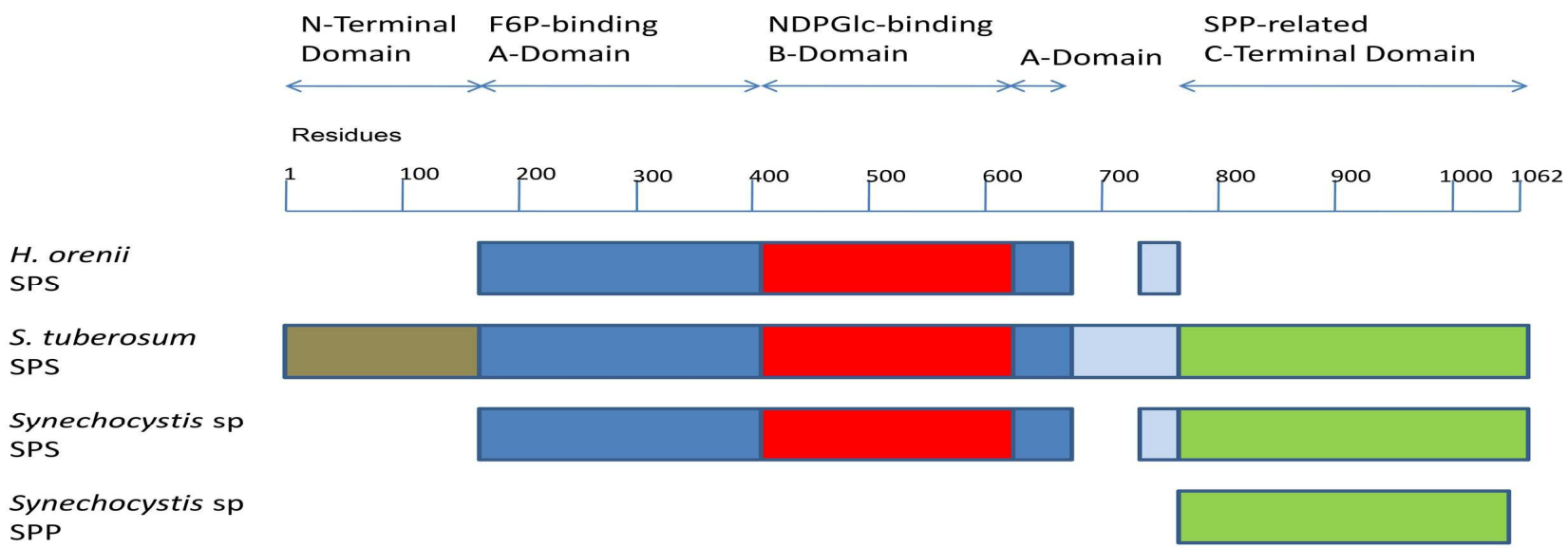


Figure 2.4 Schematic diagram of *H. orenii* SPS with *S. tuberosum* SPS (closest homolog of *H. orenii* SPS belonging to Plant SPS), *Synechocystis* sp. PCC 6803 SPS and *Synechocystis* sp. PCC 6803 SPP. F6P and NDP-Glc binding domains of SPS homologs are deduced by *H. orenii* SPS-F6P complexed structure and docked model analysis, are represented by blue and red bars respectively. The missing N-terminal region in the bacterial SPS which contains the phosphorylation site is represented in brown. The SPP-related C-terminal domain (green) which is joined by a linker (light blue), is present in most bacterial and plant SPS homologs but absent in *H. orenii* SPS.

Approximately 180aa at the N-terminal region of plant SPSs are missing in bacterial SPSs (Figure 2.4). This region containing a phosphorylation site (Ser-162 in maize and Ser-158 in spinach) is involved in light-dark regulation, and is essential for the activation of the enzymatic activity in plant SPSs (Castleden *et al.*, 2004, Curatti *et al.*, 1998, Huber *et al.*, 1989, Lunn, 2002, Lunn and MacRae, 2003). According to our bioinformatics analysis carried out via the GeneSilico metaserver (Kuroski and Bujnicki, 2001), the N-terminal region of plant SPSs is intrinsically disordered. Nonetheless, it exhibits a potential to form several α-helices, which, under some specific conditions, e.g. in the presence of a ligand, could potentially fold to form a stable three-dimensional structure.

Many bacterial and plant homologs also possess an additional C-terminal domain, which is missing from the *H. orenii* SPS (Figure 2.4). This additional domain possesses sequence similarity to the catalytic domain of SPPs, which catalyzes the final step in the pathway of sucrose biosynthesis, by dephosphorylating S6P to sucrose (Lunn, 2002). In plant SPS, the SPP-related C-terminal domain is joined by a linker to the NDP-Glc binding domain. A shorter version of this linker is also present in *H. orenii* and *Synechocystis* sp SPS (Figure 2.4). It has been proposed that in most cyanobacterial SPS this SPP domain is an inactivated pseudo-enzyme because of the absence of conserved Asp residues potentially critical for catalysis (e.g. replaced by Ala4 and Gln6 in the enzyme from Synechocystis sp. 6803), which is further supported by the lack of experimentally detectable SPP activity (Fieulaine *et al.*, 2005, Lunn, 2002, Lunn and MacRae, 2003). However, in some proteobacteria, including *A. ferrooxidans* and *N. europaea*, the SPP-like domain of predicted SPS enzymes contains all of the conserved residues, suggesting that these enzymes are bi-functional with both SPS and SPP

activities (Cumino *et al.*, 2002, Lunn, 2002, Lunn and MacRae, 2003). The absence of the active SPP domain from SPS is usually correlated with the presence of a separate SPP-encoding gene (Lunn, 2002). Thus, we predict that if *H. orenii* is to synthesize sucrose (this capability has yet to be confirmed experimentally), then it must utilize a separate yet to be discovered SPP enzyme.

Under the classification of glycosyltransferases (GTs; EC 2.4.x.y.) based on sequence similarities and stereochemistries of their substrates and products, SPSs are categorized under the CAZy [Carbohydrate Active Enzymes database, (Coutinho and Henrissat, 1999); http://www.cazy.org/] Family 4, known as retaining GTs (MacGregor, 2002, Ullman and Perkins, 1997). GTs are enzymes involved in the biosynthesis of carbohydrates and glycoconjugates (a). In general, GT structures adopt 3 folds, dubbed 'GT-A', 'GT-B' and 'GT-C' (Breton *et al.*, 2006, Gibson *et al.*, 2002, Horcajada *et al.*, 2006, Lunn and MacRae, 2003). The 'GT-A' fold consists of two dissimilar domains with the nucleotide binding domain that resembles a Rossmann fold and another smaller acceptor domain (Breton *et al.*, 2006). The 'GT-C' fold is found in integral membrane GTs (Breton *et al.*, 2006, Liu and Mushegian, 2003). The 'GT-B', also known as the "glycogen phosphorylase glycosyltransferase" (GPGTF) superfamily (Wrabi and Grishin, 2001) comprises of two distinct Rossmann-fold domains: a sugar acceptor and a sugar donor domain. Hence, SPS is generally categorized under the retaining GT-B family. Although some SPSs have been reported to be metal-dependent (Porchia and Salerno, 1996), retaining GT-B is believed to exhibit a metal ion independent mechanism (Breton *et al.*, 2006, 6, Gibson *et al.*, 2002, Liu and Mushegian, 2003) and no metal ion has been identified in the structures solved so far. In addition, it is reported that plant SPSs are

specific for UDP-Glc, whereas bacterial SPSs (*Synechocystis* and *Anabaena*) are not (Curatti *et al.*, 1998, Gibson *et al.*, 2002, Lunn *et al.*, 1999). The *H. orenii* SPS, like the *Synechocystis* SPS, is able to accept other NDP-Glc such as ADP-Glc and GDP-Glc (Huynh *et al.*, 2005).

2.3.2 Characterization of SPS

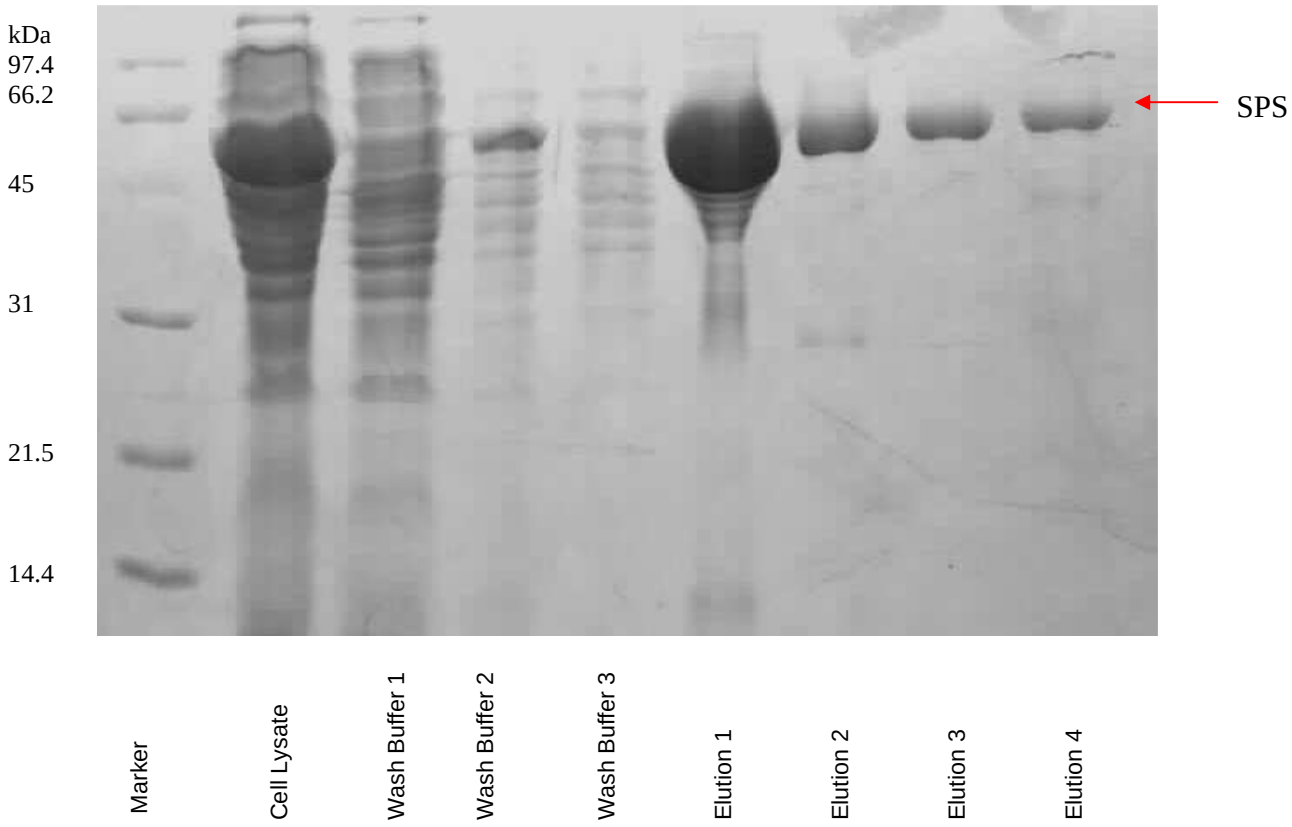


Figure 2.5 SDS-PAGE gel image of *H. orenii* SPS purification. The purified SPS migrated as a single band between 66.2 kDa and 45 kDa (Protein Ladder, SDS-PAGE Molecular Weight, Low-range by BioRad) in 12.5% SDS-PAGE, viewed by Coomassie staining.

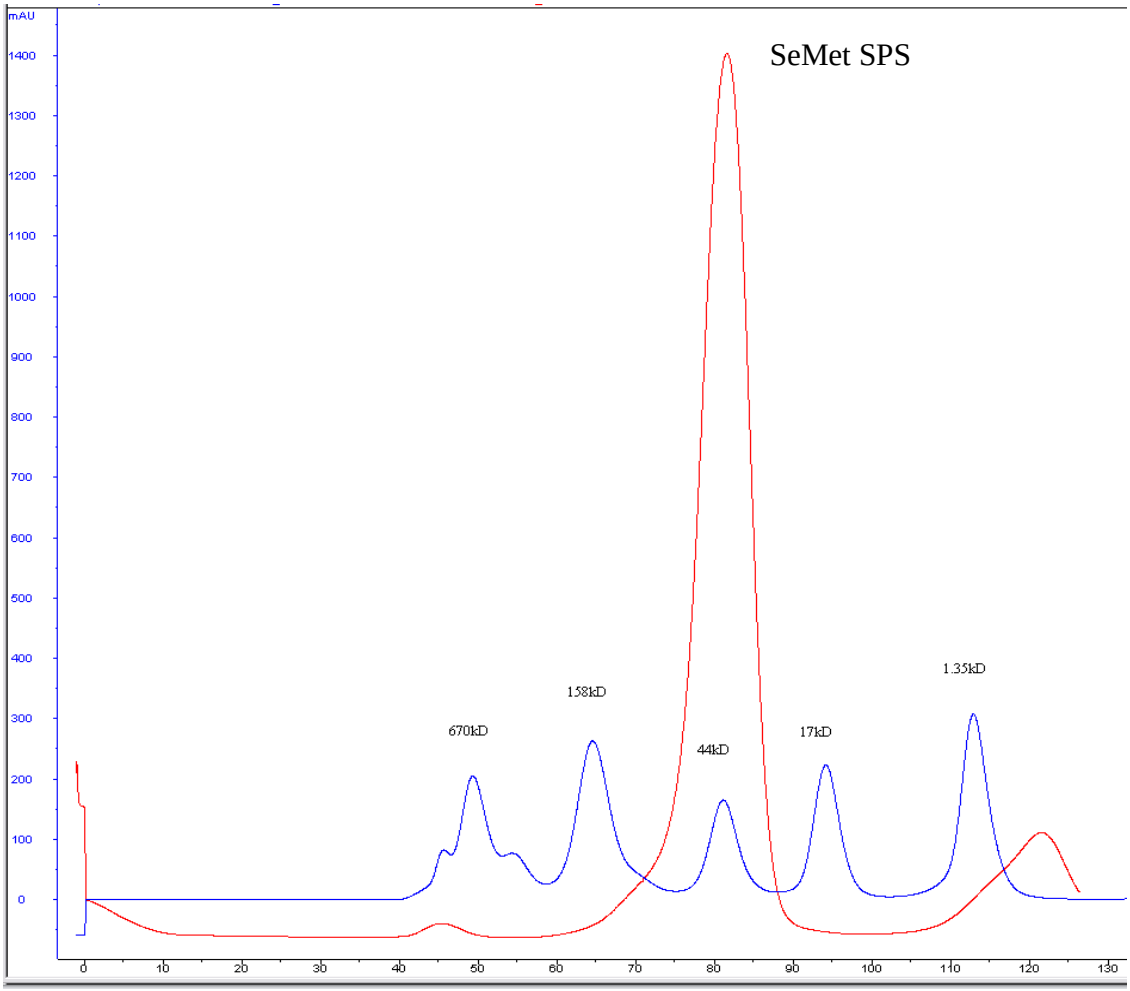


Figure 2.6 Gel filtration profile of SPS. The X-axis indicates the elution volume in mL and the Y-axis indicates the UV absorbance at 280 nm measured in mAU (arbitrary units). The elution profile is for protein injected into FPLC Hiload 16/60 Superdex75 gel filtration column (Amersham Biosciences).

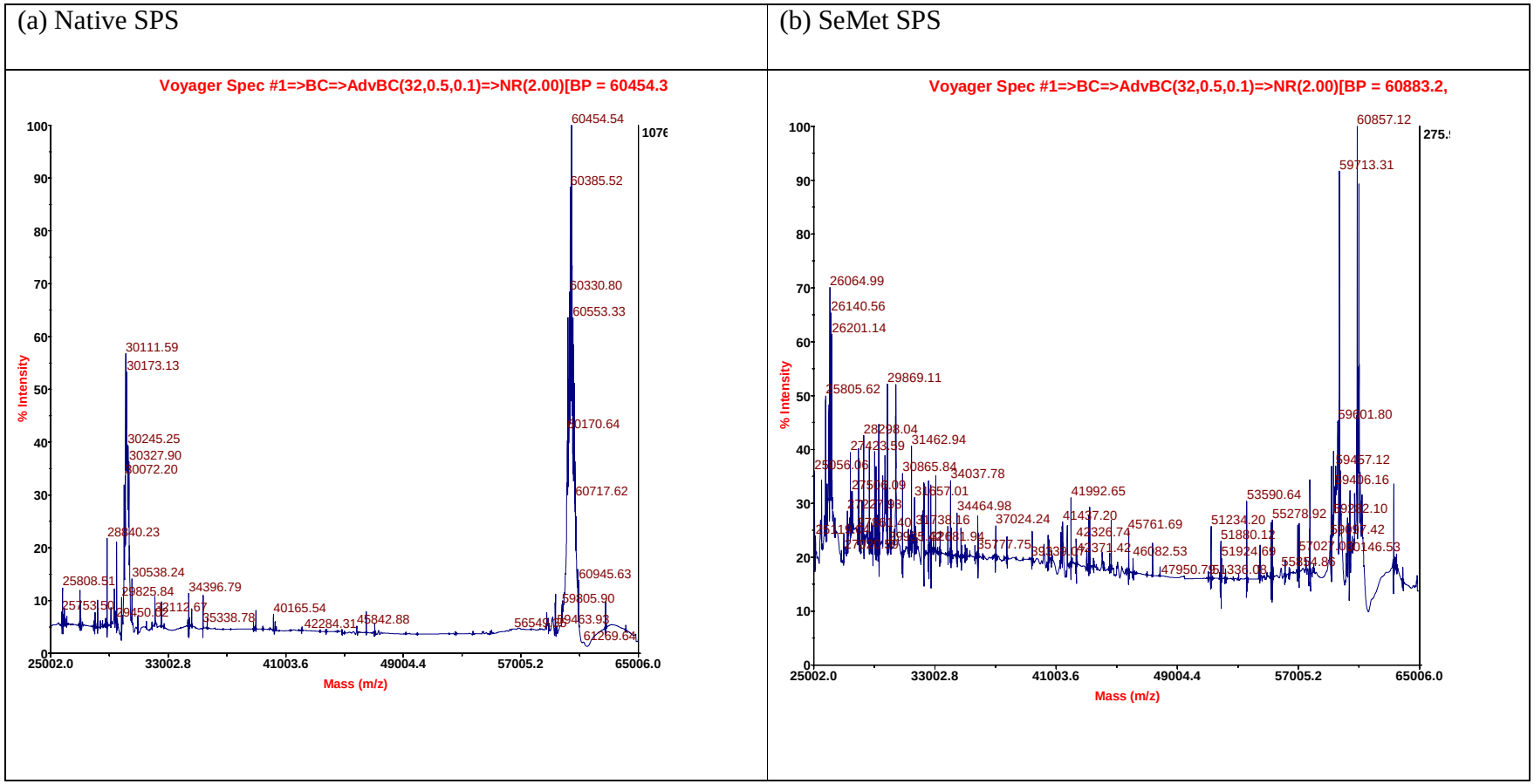
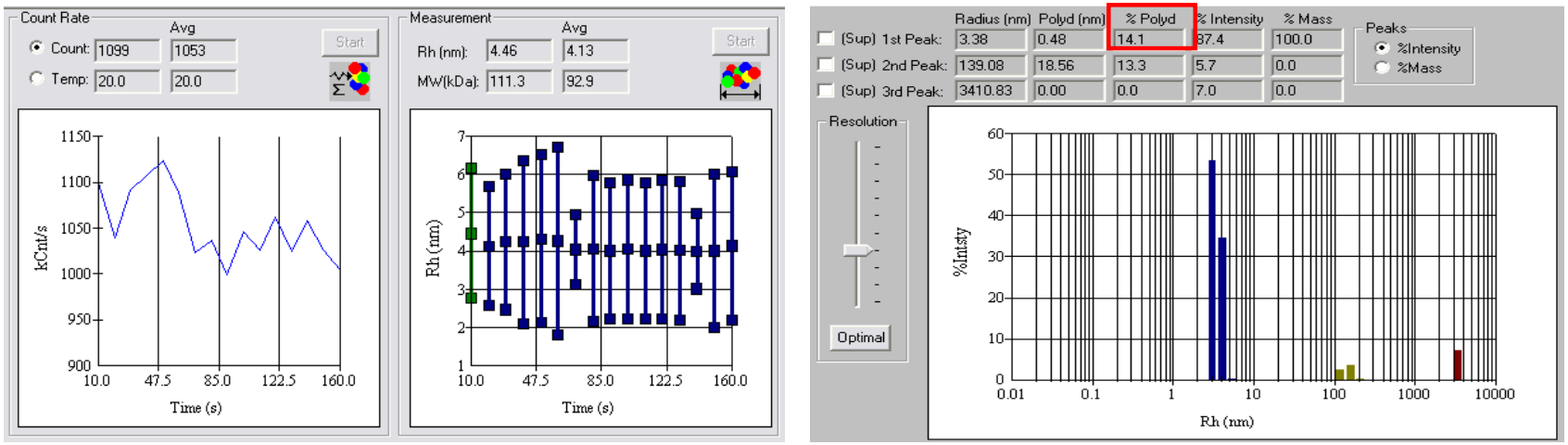


Figure 2.7 MALDI-TOF MS results for native and selenomethionyl SPS. (a) MALDI-TOF MS spectrum for native SPS. **(b)** MALDI-TOF MS spectrum for SeMet SPS.



Msr#	Time(s)	Temp(C)	Count Rate	Ampl	Diff Coeff	Radius(nm)	Polyd(nm)	PolydIndx	MW(KDa)	%Mass	Baseline	Sos Error
1	10.0	20.0	1099796	0.641	471.	4.46	1.70	0.15	111.	100.0	1.004	59.3
2	20.0	20.0	1039549	0.618	509.	4.13	1.56	0.14	92.9	100.0	1.001	28.8
3	30.0	20.0	1090917	0.602	495.	4.24	1.77	0.17	98.9	100.0	1.002	39.8
4	40.0	20.0	1107292	0.603	495.	4.24	2.13	0.25	98.9	100.0	1.003	34.7
5	50.0	20.0	1123592	0.607	486.	4.32	2.19	0.26	103.	100.0	1.003	38.3
6	60.0	20.0	1088783	0.610	493.	4.26	2.46	0.33	100.	100.0	1.001	36.2
7	70.0	20.0	1023287	0.612	520.	4.04	0.923	0.05	88.3	100.0	1.000	21.1
8	80.0	20.0	1037226	0.616	517.	4.06	1.91	0.22	89.4	100.0	1.000	21.1
9	90.0	20.0	999829	0.623	524.	4.01	1.78	0.20	86.7	100.0	1.000	14.3
10	100.0	20.0	1045831	0.609	519.	4.04	1.82	0.20	88.6	100.0	1.001	21.3
11	110.0	20.0	1026111	0.610	524.	4.01	1.78	0.20	86.7	100.0	0.999	18.5
12	120.0	20.0	1062040	0.601	520.	4.04	1.80	0.20	88.2	100.0	1.000	17.4
13	130.0	20.0	1024498	0.588	524.	4.01	1.80	0.20	86.8	100.0	1.001	20.0
14	140.0	20.0	1058264	0.591	525.	4.00	0.983	0.06	86.2	100.0	1.001	17.4
15	150.0	20.0	1026429	0.611	524.	4.01	2.00	0.25	86.6	100.0	1.000	19.8
16	160.0	20.0	1004535	0.637	508.	4.13	1.95	0.22	93.2	100.0	1.001	28.0
Aves:												
Mono		20.0	1053623	0.611	510.	4.13	1.79	0.19	92.9	100.0	1.001	27.3
Bi-1		0.0	0	0.000	0.000	0.000	----	----	0.000	0.0	0.000	0.000
Bi-2				0.000	0.000	0.000			0.000	0.0		

Figure 2.8 Dynamic Light Scattering results for SeMet SPS. The % Polydispersity, molecular weight and SOS error are indicated by red, green and blue boxes respectively.

(a) **(b)**

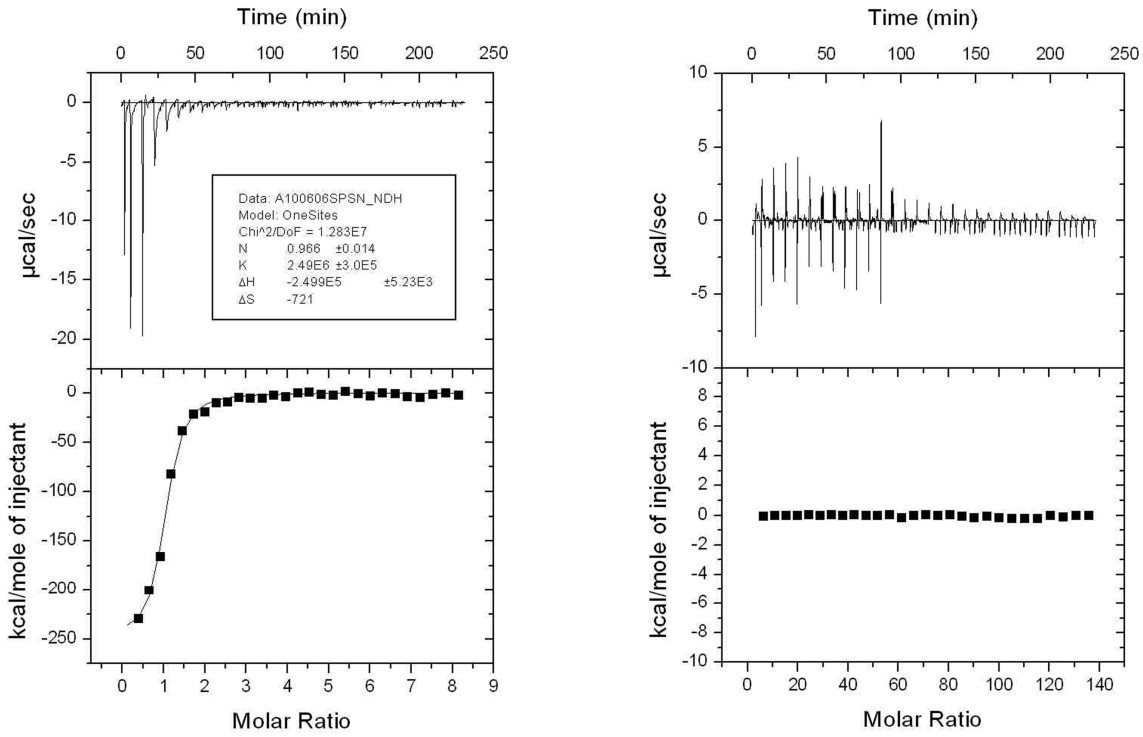


Figure 2.9 ITC profile of *H. orenii* SPS and substrate F6P. (a) Baseline subtracted raw ITC data for injections of F6P is indicated in the upper panel of the ITC profiles shown. The peaks normalized to 1:1 ligand and protein molar ratio were integrated as is shown in the bottom panel. (b) Control experiment: ITC data same as (a) except no F6P. The solid dots indicate the experimental data and the best fit to the experimental data were obtained from a non-linear least squares method of fitting using a one-site binding model depicted by a solid line. The Gibbs free energy change ($\Delta G = -2.26 \times 10^6$ kcal/mol) implies a favorable enzymatic reaction.

2.3.3 Crystallization and data collection

Diffraction quality crystals were obtained according to the procedure as described in the "Method and Materials" section.

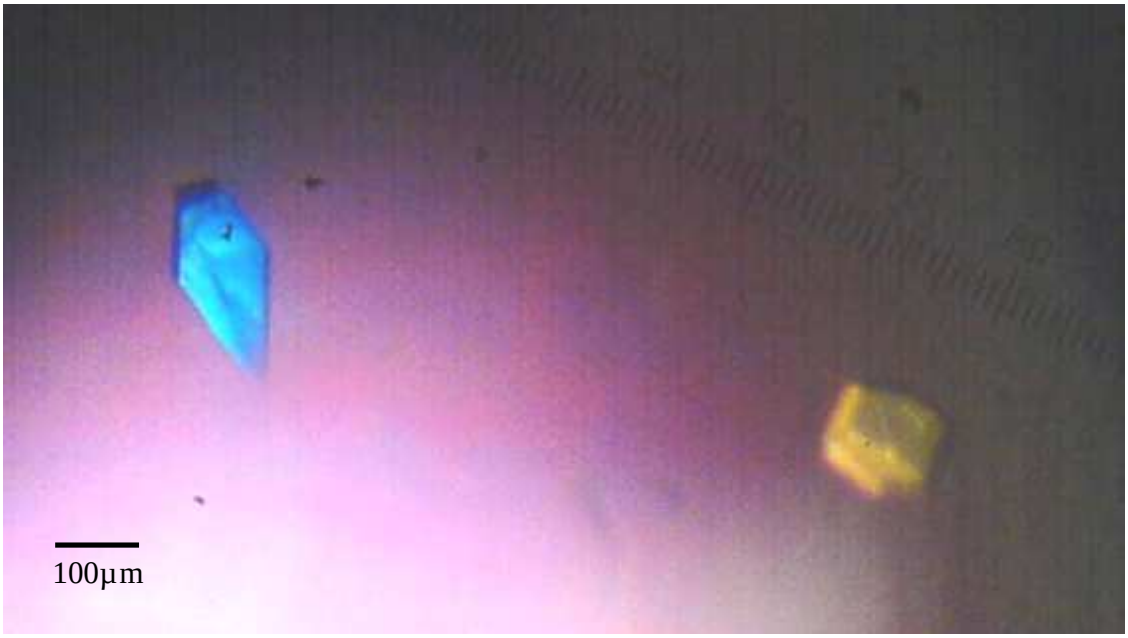


Figure 2.10 Crystals of SeMet SPS. SeMet SPS crystals were obtained by the hanging drop vapor diffusion method.

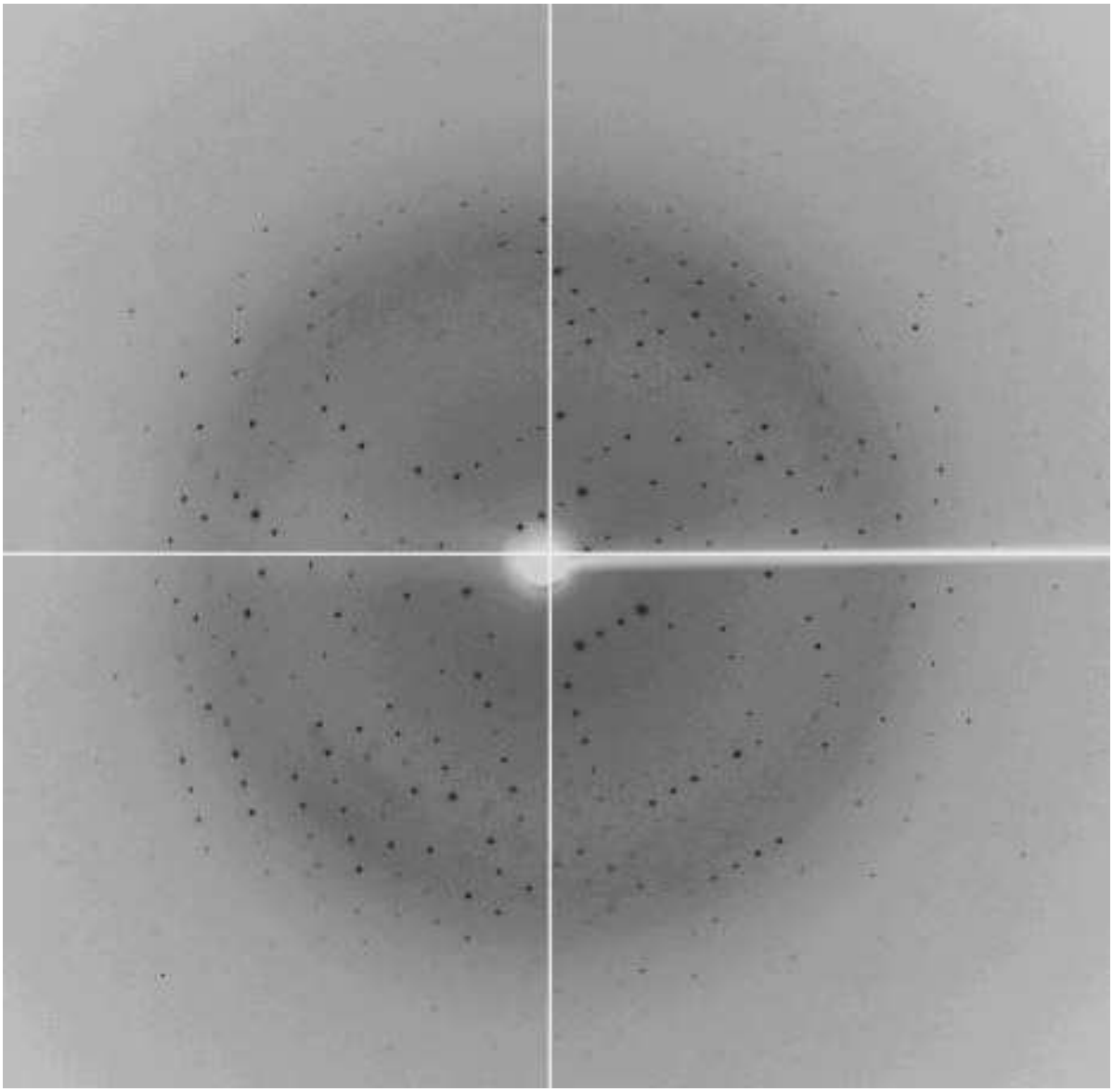


Figure 2.11 Sample diffraction pattern of SeMet SPS crystal. Sample diffraction pattern collected from ADSC Q210 diffractometer system at X12C beamline (NSLS, BNL) for SeMet SPS crystal.

Table 2.1 Data collection and refinement statistics

Data set	Peak	Inflection	Remote	F6P Complex	S6P Complex	High Resolution
Data collection						
Resolution range (Å)	50.0-2.0 (2.1-2.0)	50.0-2.1 (2.1-2.0)	50.0-2.0 (2.1-2.0)	50.0-2.7 (2.8-2.7)	50.0-2.3 (2.4-2.3)	50.0-1.8 (1.9-1.8)
Wavelength (Å)	0.9788	0.9794	0.9600	1.5418	1.5418	0.9788
Observed reflections > 1	211355	220102	243363	55492	92403	266501
Unique reflections	37156	34739	37435	13734	25868	48898
Completeness (%)	99.3	99.8	99.9	99.0	99.8	95.7
Overall (I/σ)	17.9 (1.8)	15.0 (3.4)	14.7 (2.8)	13.4 (3.1)	15.5 (2.0)	12.1 (1.8)
R_{sym} (%)[a]	5.6 (18.8)	6.2 (27.0)	5.9 (32.2)	7.9 (32.2)	5.0 (33.5)	6.0 (29.0)
Refinement and quality[b]						
Resolution range (Å)				20.0-2.8	20.0-2.4	20.0-1.8
R_{work} (no. of reflections)[c]				0.210 (11318)	0.210 (17949)	0.226 (36349)
R_{free} (no. of reflections)[d]				0.267 (1282)	0.268 (1998)	0.252 (4056)
rmsd bond lengths (Å)				0.005	0.009	0.010
Rmsd bond angles				1.0	1.3	1.4
Average B-factors (Å^2)[e]						
Main-chain				42.101	40.386	32.639
Side-chain				44.314	46.533	35.458
Ramachandran plot						
Most favored regions (%)				86.7	87.2	89.7
Additional allowed regions (%)				12.3	11.8	9.2
Generously allowed regions (%)				0.5	0.5	0.5
Disallowed regions (%)[f]				0.5	0.5	0.5

[a] $R_{sym}=|I_i-<I>|/|I_i|$ where I_i is the intensity of the ith measurement, and $<I>$ is the mean intensity for that reflection.
[b] For all models, reflections with $I>\sigma I$ was used in the refinement.
[c] $R_{work}=|F_{obs}-F_{calc}|/|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[d] R_{free}=as for R_{work}, but for 10% of the total reflections chosen at random and omitted from refinement.
[e] Individual *B*-factor refinement was carried out.
[f] Residues in the disallowed regions are well defined in the electron density map

2.3.4 Overall Structure.

The structure of recombinant SPS from *H. orenii* was solved by Multi-wavelength Anomalous Dispersion (MAD) method from synchrotron data and refined to a final R-factor of 0.226 (R_{free}=0.252) at 1.8 Å resolution. The structure of F6P-SPS complex was refined at 2.8 Å resolution, to an R-factor of 0.210 (R_{free}=0.267). Data for the S6P-SPS complex were collected to 2.4 Å resolution, and refined to an R-factor of 0.210 (R_{free}=0.268). All three models have been refined with good stereochemical parameters (Table 2.1). Statistics for the Ramachandran plot from an analysis using PROCHECK (Kuroski and Bujnicki, 2001) for these three models gave approximately 88% of non-glycine residues in the most favored region, with Tyr128 and His151 in the disallowed regions. Interestingly, these two residues are well-defined in the electron density map and are key amino acids involved in the substrate binding and reaction. It is worth mentioning that a similar observation of key substrate-recognizing residues in the forbidden region has been previously reported for the SPS structural homologs such as trehalose 6-phosphate synthase, OtsA (PDB code 1GZ5, Gibson *et al.*, 2002), and glycogen synthase (PDB code 1RZU, Buschiazzo *et al.*, 2004). The hexahistidine tag present in the protein used for crystallization is not visible in the electron density and, additionally, 37 C-terminal residues (Lys463-Glu499) are disordered. The asymmetric unit consists of a SPS/complex molecule. This monomer observation is consistent with the gel filtration result.

The SPS molecule consists of two domains (A-domain: Ile7-Gly229 and Tyr443-Arg462, and B-domain: Val230-Arg442) that form a deep substrate binding cleft at the interface with a dimension of approximately 20 Å wide and 30 Å deep. Each domain

topology is similar to a Rossmann fold (Figure 2.12). The A-domain (mostly the N-terminal residues) has a central core β-sheet consisting of eight mostly parallel β strands (β1↑β2↑β3↓β4↑β5↑β6↑β7↑β8↑) flanked on three sides by seven helices; of which three are 1-2 turn small helices. The B-domain (mostly comprising of C-terminal residues) has a central parallel β-sheet of six strands (β9↓β10↓β11↓β12↓β13↓β14↓) flanked by nine α-helices (three of which are 1-2 turn helices). The A- and the B-domains are connected through the loops Pro228-Val230 and Arg442-Gln446. The latter loop is considered as a part of the kink crossing over the domains and connecting two α-helices, a general feature for enzymes belonging to the GT-B fold superfamily (Breton *et al.*, 2006, Gibson *et al.*, 2002, Horcajada *et al.*, 2006). The A- and B-domains superimpose with an rmsd of 3.2 Å for 104 Cα atoms and exhibit 11.5% sequence identity.

2.3.5 Structural Comparisons to Other Proteins.

Structural comparison of *H. orenii* SPS with other protein structures was performed using the program DALI (Holm and Sander, 1993). Significant structural similarities were found with glycogen synthase, trehalose 6-phosphate synthase (OtsA) and glycogen phosphorylase (Figure 2.13 and Figure 2.14), all of which belong to the GT-B family and possess catalytic mechanisms of retaining GTs. The closest structural similarity is observed between SPS and *Agrobacterium tumefaciens* glycogen synthase complexed with ADP (PDB code 1RZU) from the CAZy Family 5, yielding an rmsd of 4.4 Å for 365 Cα atoms, with approximately 11% sequence identity. This is followed by *Escherichia coli* trehalose 6-phosphate synthase complexed with G6P-UDP (OtsA; PDB

code 1GZ5) from the CAZy Family 20 (rmsd = 4.4 Å for 269 Cα atoms; approximately 13% sequence

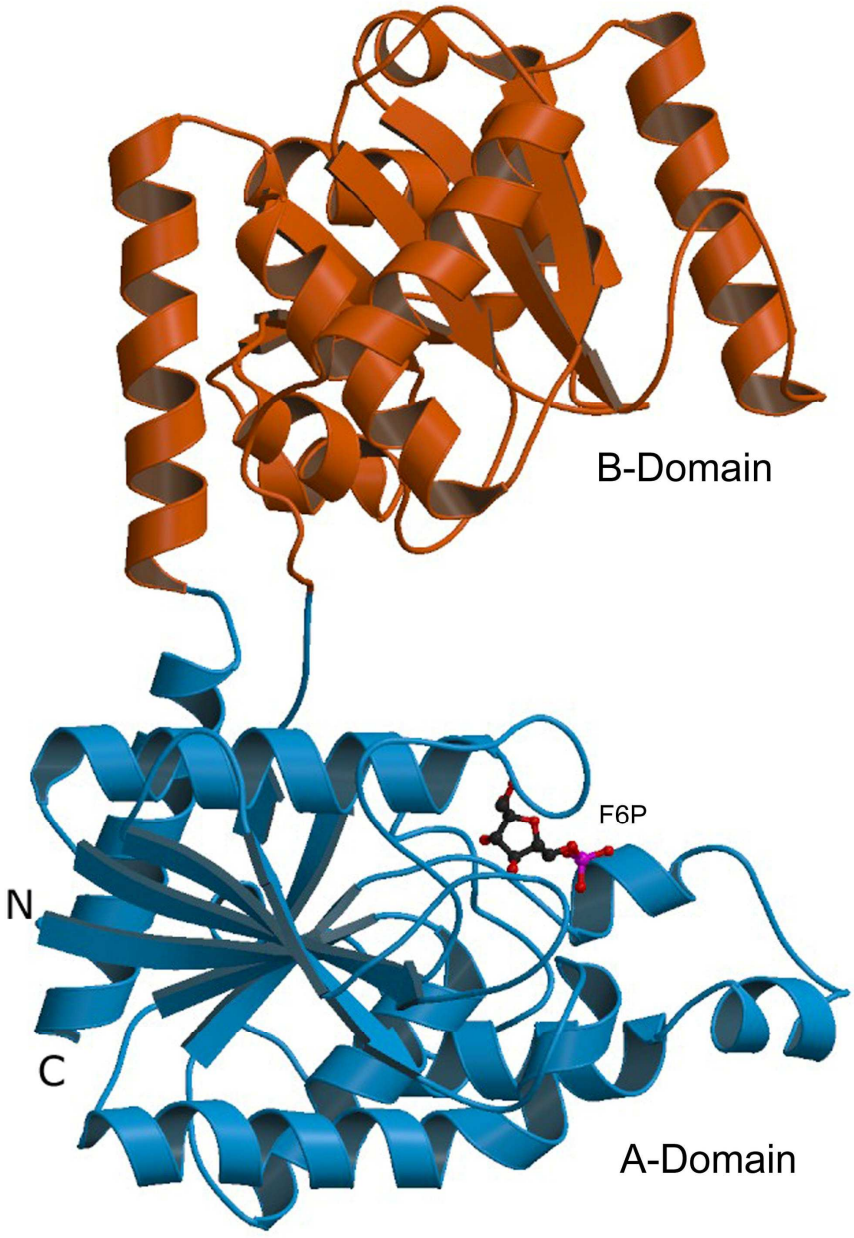


Figure 2.12 Ribbon diagram showing the structure of SPS. A-domain (residues 7-229; 443-462) is depicted in blue and the B-domain (residues 230-442) in red. The bound substrate molecule D-Fructose-6-Phosphate (F6P) is depicted as a ball-and–stick

representation. The N- and C-terminals are labeled. This figure was prepared using the programs MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt *et al.*, 1997).
identity) and *Oryctolagus cuniculus* glycogen phosphorylase complexed with glucopyranose spirohydantoin (PDB code 1A8I), from the CAZy Family 35 (rmsd = 4.2 Å for 311 Cα atoms; approximately 7% sequence identity).

However, the superimposition of individual domains of SPS and its homologs exhibit a good fit. The A-domain of the SPS superimposes on the corresponding domains of glycogen synthase, trehalose 6-phosphate synthase (OtsA) and glycogen phosphorylase with a rmsd of 2.6 Å for 197 Cα atoms, 3.2Å for 191 Cα atoms and 2.9 Å for 184 Cα atoms respectively. Similarly, the B-domain of SPS superimposes on the corresponding domain of these same homologs with a rmsd of 2.8 Å for 178 Cα atoms, 3.2 Å for 188 Cα atoms and 3.1 Å for 187 Cα atoms respectively. Thus, the comparison of the full length SPS with its structural homologs shows variations in the relative disposition of A- and B-domains of these enzymes (Figure 2.14). This type of flexibility in two-domain enzymes is not unusual and has been reported for several two domain enzymes (Breton *et al.*, 2006, Buschiazzo *et al.*, 2004, Horcajada *et al.*, 2006, MacGregor, 2002). Furthermore, these structural comparisons suggest a possibility of different conformations of GT-B structures. Structures of SPS (or its two complexes) and the glycogen synthase-ADP complex may represent an open conformation (Buschiazzo *et al.*, 2004), whereas the trehalose 6-phosphate synthase-G6P-UDP complex may represent a closed conformation (Buschiazzo *et al.*, 2004, Gibson *et al.*, 2002).

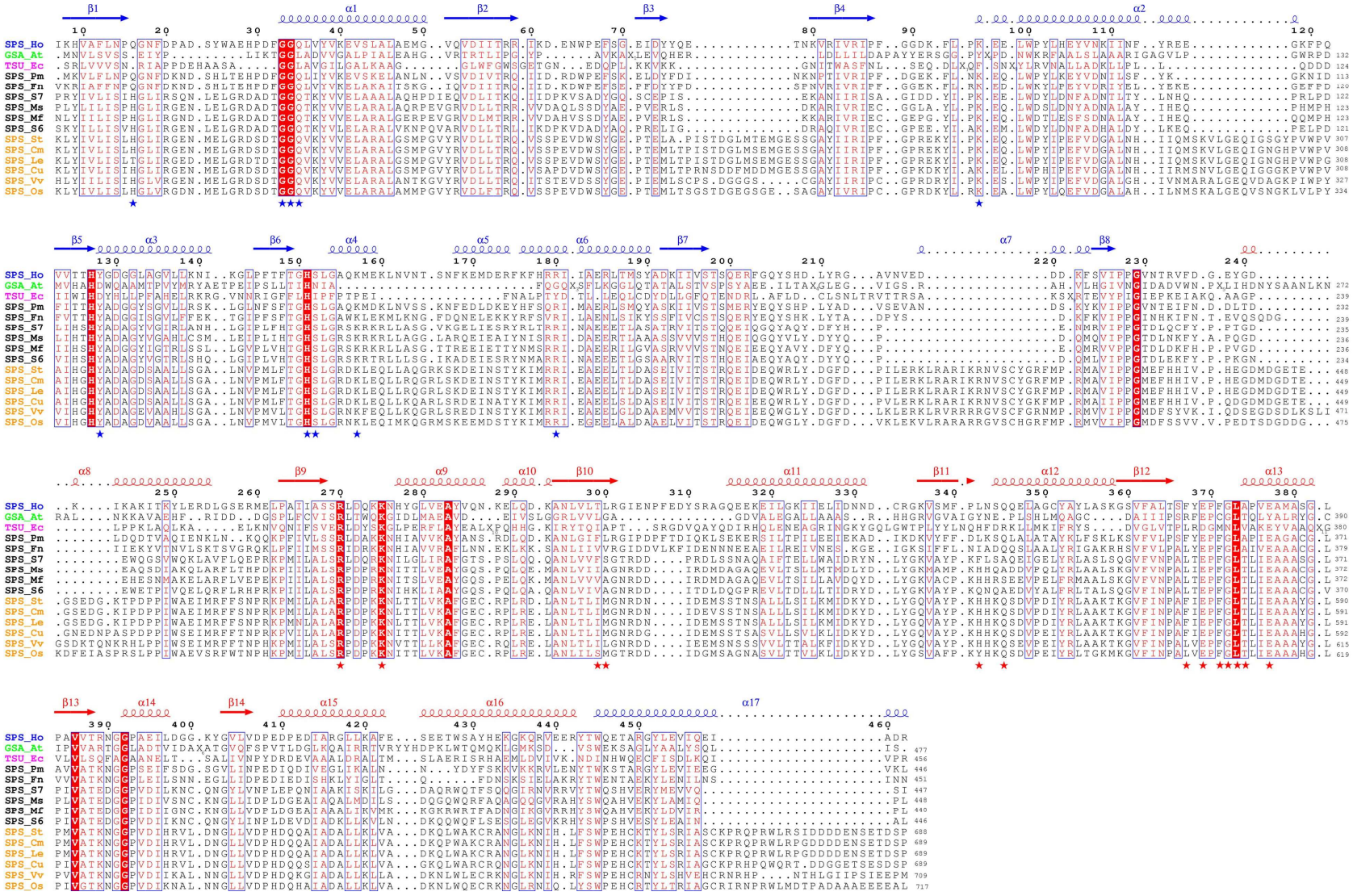


Figure 2.13 Structure based sequence alignment of *H. orenii* SPS. (a) Top 3 rows: Structure based sequence alignment of SPS (blue), Glycogen synthase (PDB code 1RZU, green) and Trehalose 6-phosphate synthase (PDB code 1GZ5, magenta). The independent domains are superimposed. The amino acids are in one-letter codes; the conserved residues are highlighted. Strictly conserved residues are shaded red with semi-conserved residues lettered in red. Secondary structural elements of *H. orenii* SPS belonging to A- and B-domains are shown in blue and red respectively. This figure was created using the program ESPript (Gouet *et al.,* 1999). Middle 4-9th rows: Sequence alignment of *H. orenii* SPS (top, blue) with the closest six SPS homologs (black) was carried out using ClustalW (Chenna *et al.,* 2003) and ESPript (Gouet *et al.,* 1999). Bottom 10-15th rows: Sequence alignment of *H. orenii* SPS (top, blue) with the closest six SPS plant homologs (orange). Key substrate binding residues in A- and B-domains are indicated by blue and red asterisks respectively. Suffix: SPS_Ho: SPS, *H. orenii;* GSA_At: Glycogen Synthase, *Agrobacterium tumefaciens* (1RZU, Buschiazzo *et al.,* 2004); TSU_Ec: Trehalose 6-phosphate synthase OtsA, *E. coli* (1GZ5, Gibson *et al.,* 2002); SPS_Pm: SPS, *Petrotoga mobilis* SJ95; SPS_Fn: SPS, *Fervidobacterium nodosum* Rt17-B1; SPS_S7: SPS, *Synechococcus* sp. PCC 7002; SPS_Ms: *Magnetococcus* sp. MC-1; SPS_Mf: SPS, *Mariprofundus ferrooxydans* PV-1; SPS_S6: SPS, *Synechocystis* sp. PCC 6803; SPS_St: *Solanum tuberosum*; SPS_Cm: *Cucumis melo*; SPS_Le: *Lycopersicon esculentum*; SPS_Cu: *Citrus unshiu*; SPS_Vv: *Vitis vinifera*; SPS_Os: *Oryza sativa*.

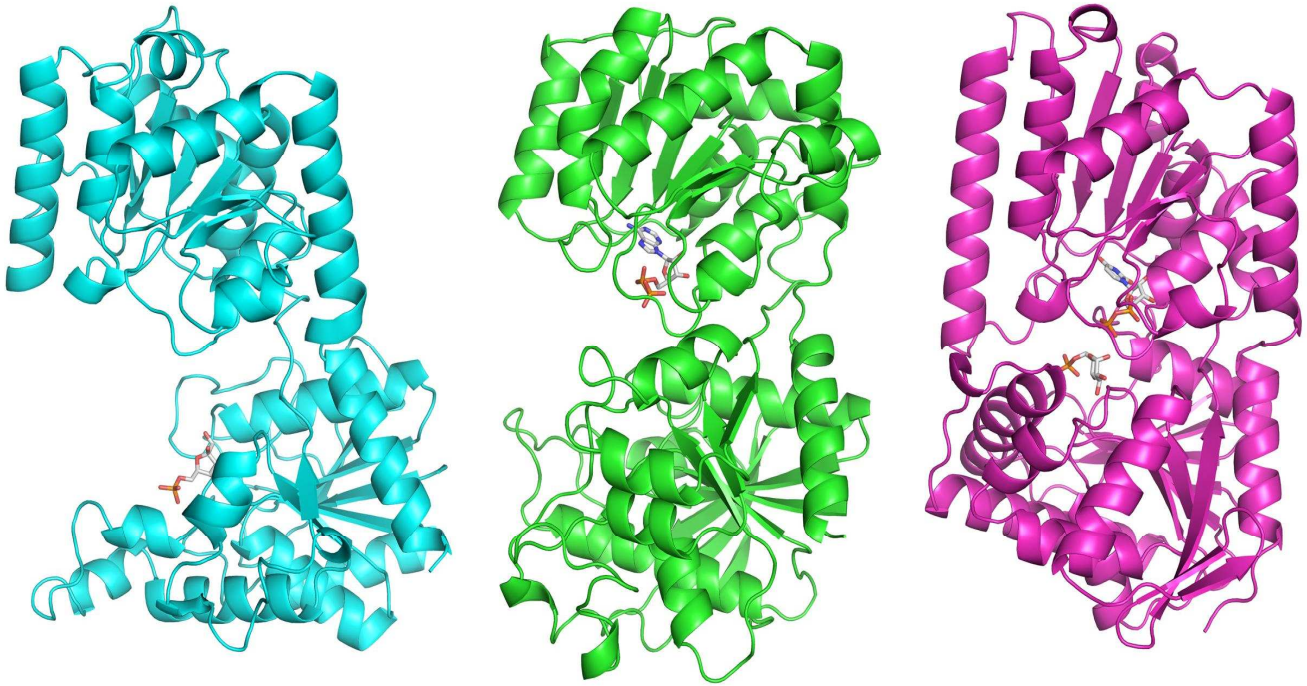


Figure 2.14 Ribbon diagrams showing three complex structures side-by-side: SPS-F6P complex (left, cyan), Glycogen synthase-ADP complex (centre, green; PDB code 1RZU, Buschiazzo *et al.*, 2004) and Trehalose 6-phosphate synthase-G6P-UDP complex (right, magenta; PDB code 1GZ5, Gibson *et al.*, 2002). The ball-and-stick representation shows the bound F6P, ADP and G6P-UDP respectively. The open and closed conformation of the substrate binding cleft observed in SPS-F6P (open), Glycogen synthase-ADP (open) and Trehalose 6-phosphate synthase-UDP (closed). These figures were prepared using the program PyMOL (DeLano, 2002).

The transformation of GT-B from an open to the closed conformation involves a small twist between the two domains (Buschiazzo *et al.*, 2004), bringing to close the substrate binding cleft. For instance, in the open GT-B conformation (e.g. *H. orenii* SPS), the entrance of the substrate binding cleft is over 20 Å, whereas in the closed GT-Bs (OtsA) it is approximately 6 Å. To illustrate this, a closed model of SPS was generated by independently superimposing the A- and B-domains of SPS on the respective domains of the closed OtsA-UDP-G6P complex structure. A figure was prepared by superimposing the B-Domain of the closed SPS-UDP model on the open SPS-F6P complex (Figure 2.15). By comparing the open SPS-F6P complex and the closed SPS-UDP model, the conformational change upon domain movement is shown.

Comparison of three-dimensional structures of SPS, OtsA and glycogen synthases indicates that catalytic domains of these enzymes are evolutionarily related. This is further supported by the presence of several invariant residues at the substrate binding sites. This structural similarity exists despite their low sequence identities, suggesting that the structure is often more conserved than the primary sequence. However, sequence identities among all SPS (including plant SPSs) are much higher than sequence identities of these three structural homologs (mentioned above) (Figure 2.13). These observed similarities of sequences, and overall structures suggest a common structural and mechanistic framework for all SPS enzymes. Therefore, the structure of *H. orenii* SPS is a valid model for the catalytic domain of plant SPSs providing valuable insight into the reaction mechanism of the plant enzyme that had not been available previously.

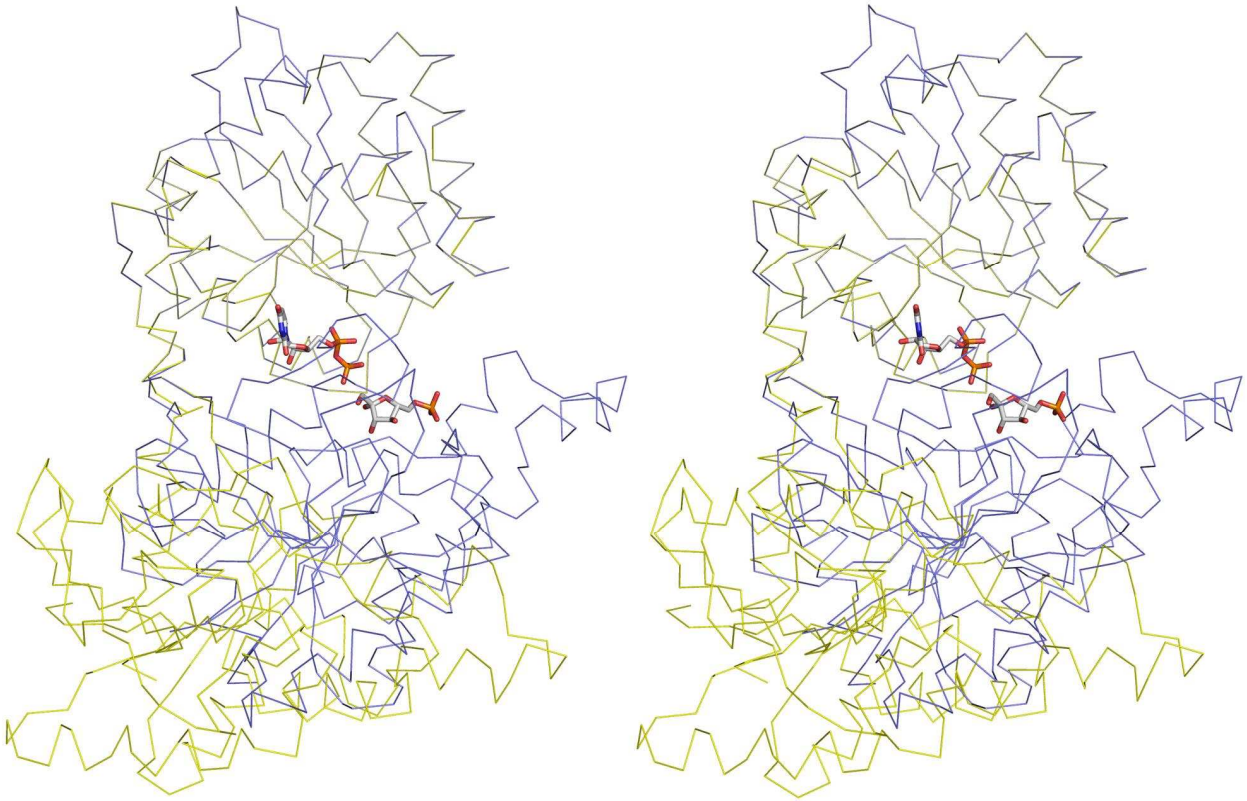


Figure 2.15 Superimposed, stereo diagram of the open SPS-F6P complex (yellow) and the closed SPS-UDP model (blue). The ball-and-stick representation shows the bound F6P and UDP at the substrate binding cleft observed in the open and closed SPS. The superimposition was performed with DALI (Holm and Sander, 1993) and O program (Jones *et al.*, 1991). These figures were prepared using the program PyMOL (DeLano, 2002).

2.3.6 SPS-F6P complex.

Prior to the crystallization of the SPS-F6P (enzyme-substrate) complex, the formation of the complex was verified by Isothermal Titration Calorimetry (ITC) experiments. The molar ratio between SPS and F6P was determined to be 0.966 (~1), suggesting a 1:1 complex (Figure 2.9a and 2.9b). In order to obtain this complex, we soaked the apo-SPS crystals in a solution containing the substrate F6P and collected a complete X-ray diffraction data-set. The difference electron density map clearly showed a substrate molecule bound to one of the two domains of SPS (Figure 2.16a). The F6P binds in a deep depression in the A-domain, in the interdomain interface cleft (Figure 2.17a and 2.17b). The substrate is located between two helices of A-domain such that α4 is close to the phosphate group and α1 is close to the sugar side of F6P. Side chains lining the binding pockets are from Gln16, Gly33, Gln35, Lys96, Tyr128, Ser152, Lys157 and Arg180 (Figure 2.16a and 2.18). These residues are conserved among the bacterial and plant SPSs (Figure 2.13). In addition, the structure and sequence analyses reveal that the binding residues of SPS to the fructose moiety of F6P (Gly33, Gln35, Lys96 and Tyr128) and to the diphosphate group of UDP-Glc (Arg270, Lys275, Glu369 and Phe367) are conserved in plant and bacteria sucrose synthases (SS). A total of nine hydrogen bonds and several hydrophobic interactions are formed between F6P and the SPS molecule. Of these, four strong hydrogen bonding contacts (<3 Å) are found between the phosphate group of F6P and highly conserved residues of SPS such as Tyr128, Ser152, Lys157 and Arg180. In the substrate binding cleft region adjacent to the F6P binding pocket there are several well-ordered water molecules, which could be replaced by the incoming second substrate NDP-Glc.

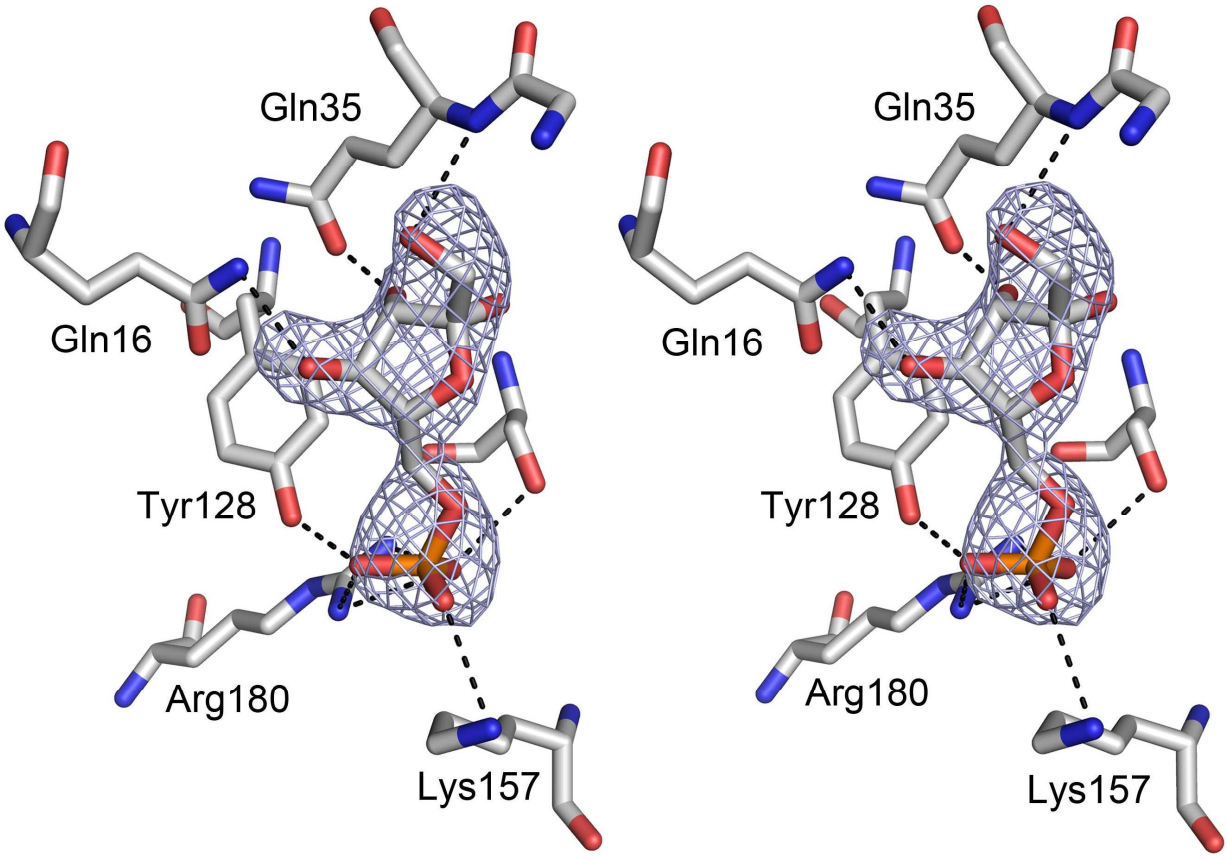
The binding of F6P does not cause any major conformational changes in the SPS structure. Furthermore, the superimposition of SPS apo- and F6P complex- structures (rmsd of 0.266 Å for 455 Cα atoms) reveals no domain movement. Only the key side-chains, Gln35, Lys157 and Arg180, show a small movement towards the F6P. It must be emphasized that the substrate was soaked into the crystal; therefore no major structural rearrangement of SPS was anticipated.

2.3.7 SPS-S6P complex.

Similar to the SPS-F6P complex, apo-SPS crystals were soaked in a solution containing S6P (product) and a complete X-ray diffraction data-set was collected. The electron density map clearly showed the presence of one S6P molecule bound at the A-domain in the domain interface cleft (Figure 2.16b). The location of the product molecule, S6P, is in the same region as F6P of the F6P-SPS complex, between the two helices of A-domain such that α4 is close to the phosphate group and α1 is close to the sugar side of S6P. Similarly, the S6P binds in a deep depression in the A-domain, at the domain interface cleft (Figure 2.17c). The overall hydrogen bonding contacts of F6P and S6P complexes are the same except for His151 (Figure 2.18). In SPS-F6P complex, His151 has no interaction with F6P molecule due to the absence of the glycosyl group. Here in the SPS-S6P complex, His151 forms a strong hydrogen bond (<3 Å) with the O atom of the transferred glycosyl group (Figure 2.16b). A total of thirteen hydrogen bonding contacts and several hydrophobic interactions are formed between S6P and SPS molecules. Noteworthy, similar to SPS-F6P complex, five strong hydrogen bonding contacts (<3 Å) are found between the phosphate group of S6P and the highly conserved

residues (Tyr128, Ser152, Lys157 and Arg180). The binding of S6P does not cause any major conformational changes in the SPS structure. The superimposition of apo-SPS and S6P complex structures (rmsd of 0.266 Å for 455 Cα atoms) reveals no significant differences. Key side-chains, such as Gln35, Lys157 and Arg180, interacting with S6P show a small movement (<1 Å) towards S6P.

(a)



(b)

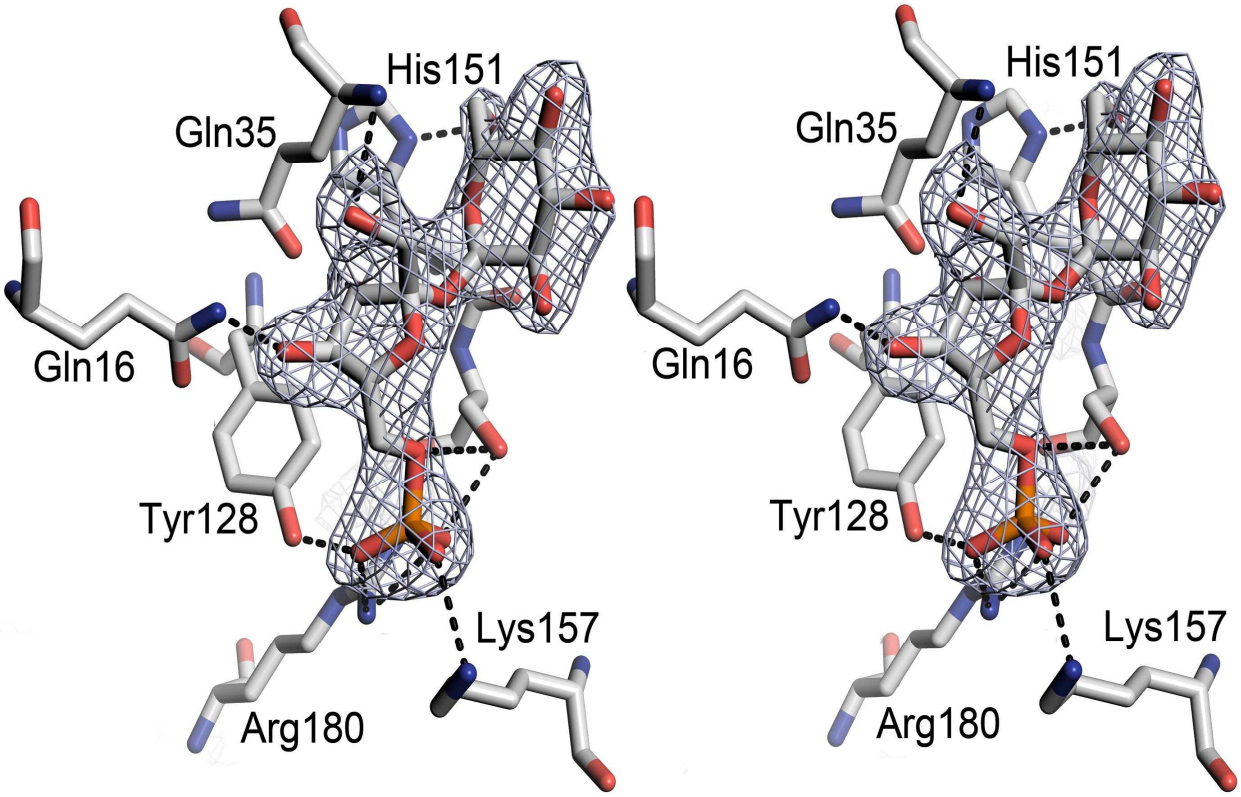


Figure 2.16 Simulated-annealing *Fo-Fc* omit map of (a) F6P and (b) S6P in the substrate binding site of SPS contoured at a level of 3.0σ. All atoms within 3.5 Å of F6P and S6P were omitted prior to refinement and map calculation. For figure clarity, maps are shown only for the substrates F6P and S6P and not all binding residues are shown. The hydrogen bonding contacts are shown in black dashes. Atoms are shown in gray (C), blue (N), red (O) and orange (P). This figure was prepared using the program PyMOL (DeLano, 2002).

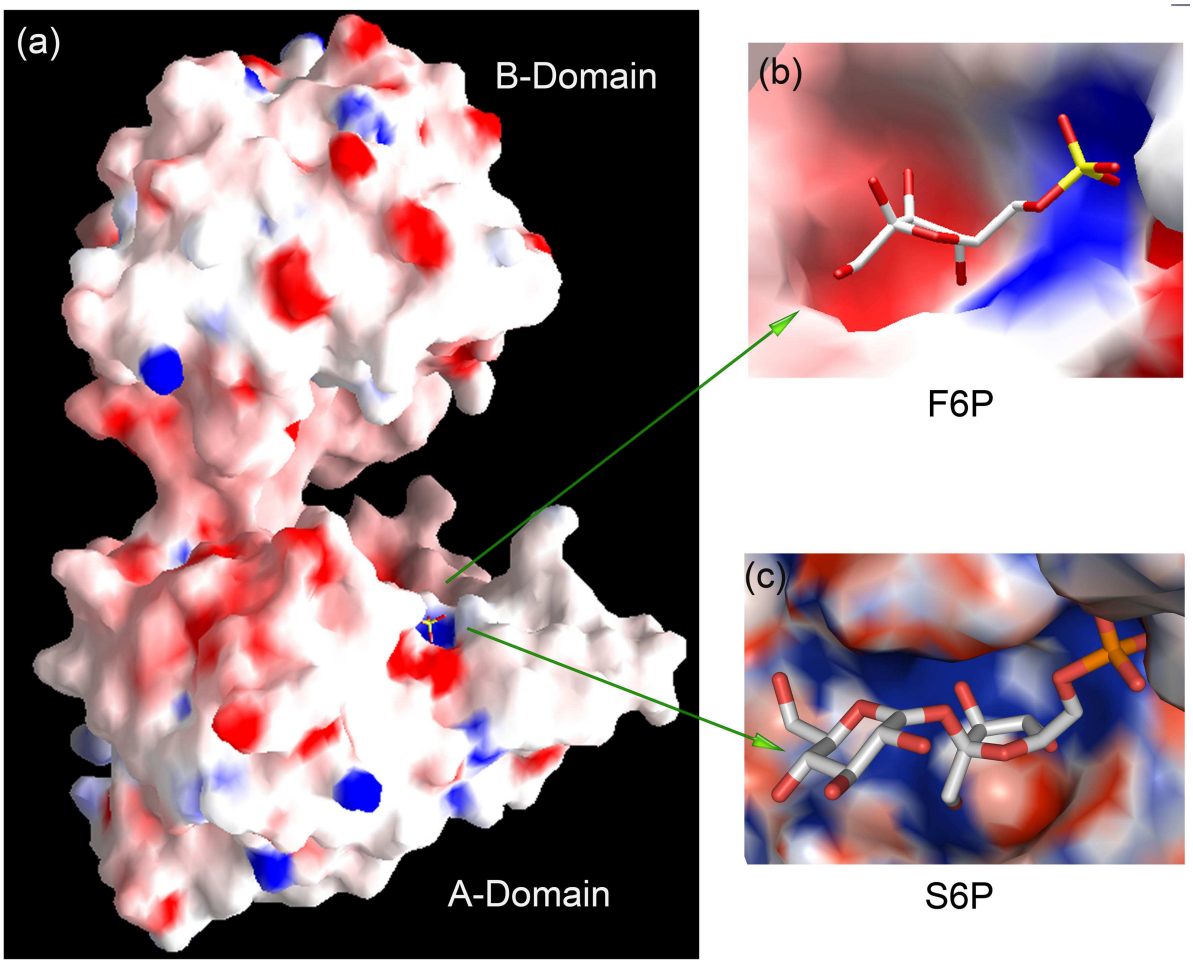


Figure 2.17 (a) Molecular surface of SPS showing the distinct two domains separated by a large substrate binding cleft. The bound F6P/S6P molecule is shown in the cleft region. The orientation is same as of Figure 2.12. **(b) Close-up view of the F6P binding site.** Atoms are shown gray (C), red (O) and yellow (P). These figures were produced using GRASP (Nicholls *et al.*, 1991). **(c) Close-up view of the S6P binding site.** Atoms are shown gray (C), red (O) and orange (P). This figure was produced using PyMOL (DeLano, 2002) using electrostatic potential generated by APBS (Baker *et al.*, 2001)

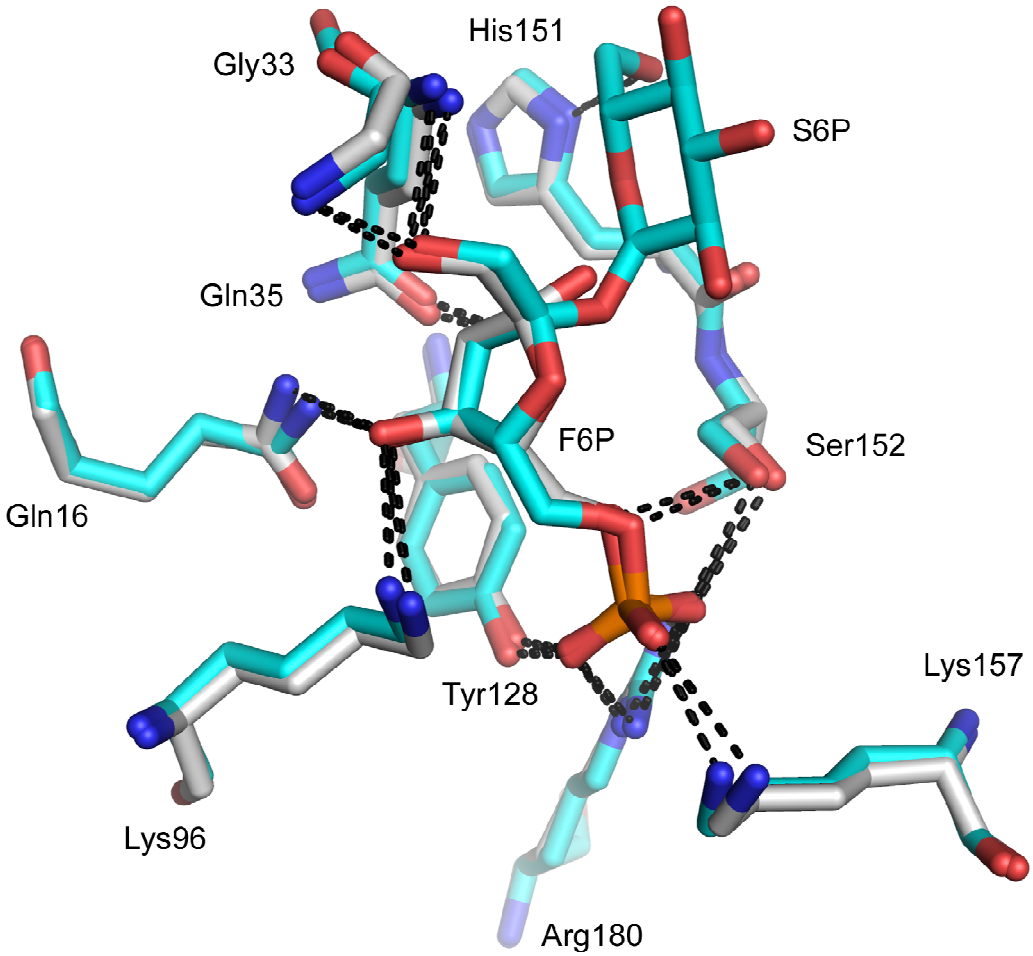


Figure 2.18 Superimposition of F6P-SPS and S6P-SPS complexes. The important difference between the 2 complexes is the hydrogen bond contact of ND1 atom of His151 and the glycosyl group-O6 atom of S6P in the S6P-SPS complex. The hydrogen bonding contacts are shown in black dashes. In the F6P-SPS complex and S6P-SPS complex, the C atoms are shown in gray and cyan respectively. The rest of the atoms are shown in blue (N), red (O) and orange (P). This figure was prepared using the program PyMOL (DeLano, 2002).

2.3.8 Putative ADP / UDP binding pocket.

It is reported that plant SPSs are specific for UDP-Glc, whereas bacterial SPSs (*Synechocystis* and *Anabaena*) are not (Curatti *et al.*, 1998, Gibson *et al.*, 2002, Lunn *et al.*, 1999). The recombinant *H. orenii* SPS, like the *Synechocystis* SPS, is able to accept other NDP-Glc such as ADP-Glc and GDP-Glc (Huynh *et al.*, 2005). Although we did not obtain the position of a second NDP-Glc substrate in the SPS structure through crystallization, the binding site of ADP-Glc and UDP-Glc can be predicted by a comparison with the structure of glycogen synthase-ADP complex (Buschiazzo *et al.*, 2004) and trehalose 6-phosphate synthase (OtsA)-UDP complex (Gibson *et al.*, 2002) respectively (Figure 2.14). The overall architecture of the nucleotide binding site is very similar in SPS, glycogen synthase and trehalose 6-phosphate synthase. In the open form of SPS, ADP/UDP is predicted to bind to a pocket on the B-domain of the interdomain cleft adjacent to the A-domain F6P binding pocket. This binding pocket is lined up by Ser268-Arg270; Thr299-Ile303; Pro370-Ser381 and Pro341-Tyr352.

To provide independent support for the predicted NDP-Glc binding site in *H. orenii* SPS, we carried out computational docking of NDP molecules with the FlexX algorithm (Kramer *et al.*, 1999). The result gave ten docked models of UDP to *H. orenii* SPS (Figure 2.17a). A similar result was obtained in the course of ADP docking to SPS with a similar orientation of the ligand (Figure 8b). Furthermore, to validate the proposed NDP-Glc binding site of SPS, we superimposed B-domains of SPS-UDP/ADP docked models on the trehalose 6-phosphate synthase OtsA-UDP (Figure 2.21) and glycogen synthase-ADP (Figure 2.22) respectively. It clearly shows the agreement between the NDP-Glc predicted binding pocket and key conserved residues of SPS, trehalose 6-

phosphate synthase (OtsA) and glycogen synthase. Arg270, Lys275, Glu369 and Glu377 are key residues and highly conserved among the plant SPS homologs (Figure 2.13). The NDP-Glc binding pocket remains the same for both the open form (SPS, Glycogen synthase) and closed form (OtsA) of these GTs. By analogy, we propose that the NDP-Glc will occupy the same binding pocket in the closed form of SPS. It should be noted that for the binding of NDP-Glc the side-chains of Arg270 and Arg301 may have a different orientation. Docked ADP and UDP at the SPS NDP-binding site form several hydrogen bonds and van der Waals interactions with the SPS molecule (Figure 2.19 and 2.20). It is known that bacterial SPSs exhibit little specificity against NDPs whereas plants SPSs are more specific to UDP-Glc than other NDP-Glc (Curatti *et al.*, 1998, Gibson *et al.*, 2002, Jones *et al.*, 1991). Most of residues interacting with the diphosphate group and the ribose moiety of NDP-Glc are well conserved in both plants and bacteria, whereas, residues interacting with the base moiety of the NDP-Glc are less conserved among bacteria than plant SPSs. Above considerations may indicate why plant SPS are specific for UDP, while bacterial SPS do not discriminate among NDP (Figure 2.19 and Figure 2.21). Unlike plant SPS, both plant and bacterial SS show similarities to bacterial SPS utilizing NDP-Glc as glycosyl donor (Porchia *et al.*, 1999). Based on the docked models (Figure 2.19 and 2.20) and sequence analysis, three nucleotide binding residues of *H. orenii* SPS (Thr299, Leu300 and Leu342; Figure 2.13) are identified. In contrast, the corresponding positions in plant SPS are substituted by conserved large side chain residues, Ile, Met and His (Figure 2.13). These variations also suggest a possible basis for the more diverse binding modes of bacterial SS, plant SS and bacterial SPS, and the stringent binding mode of plant SPS to UDP-Glc.

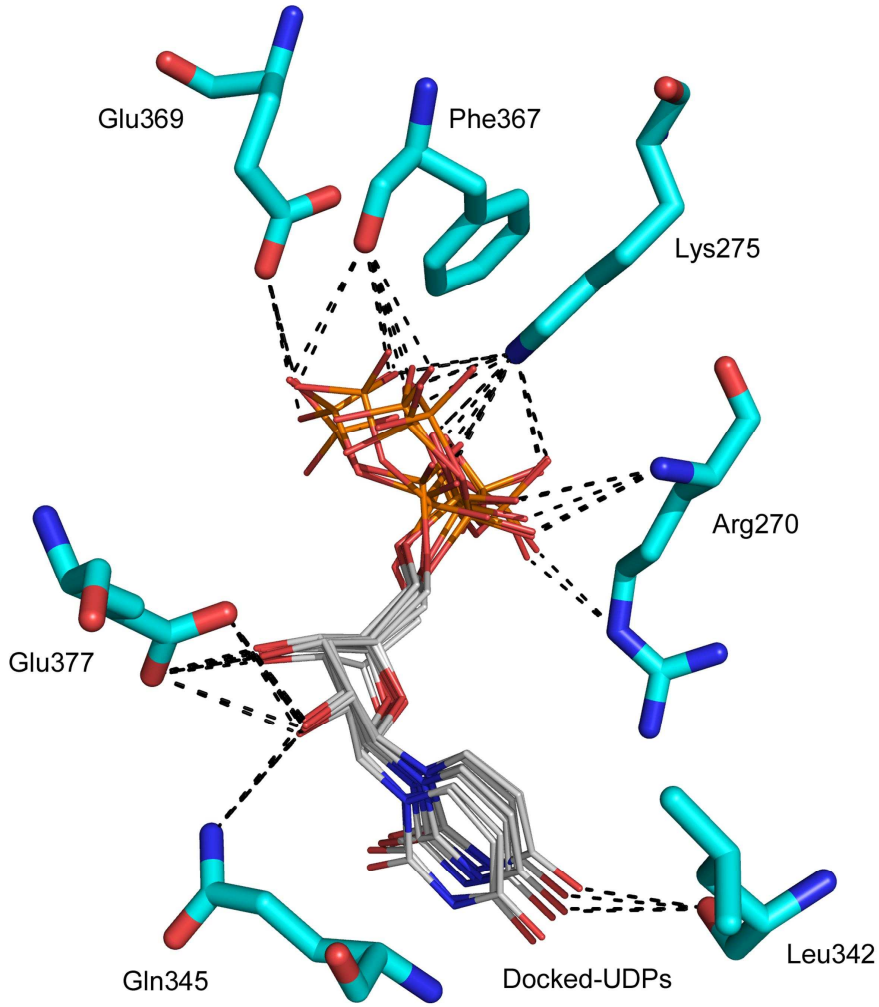


Figure 2.19 Ten docked models of UDP interacting with the binding residues of *H. orenii* SPS. Docked models of UDP (gray) obtained by the 'ab initio' method reveal the NDP-Glc binding mode of *H. orenii* SPS (cyan). From these models, most of the conserved residues are found to interact with the diphosphate group and the ribose moiety of the NDPs. Atoms are shown red (O) and orange (P). Hydrogen bonding contacts are shown in black dashed lines. This figure was prepared using the program PyMOL (DeLano, 2002).

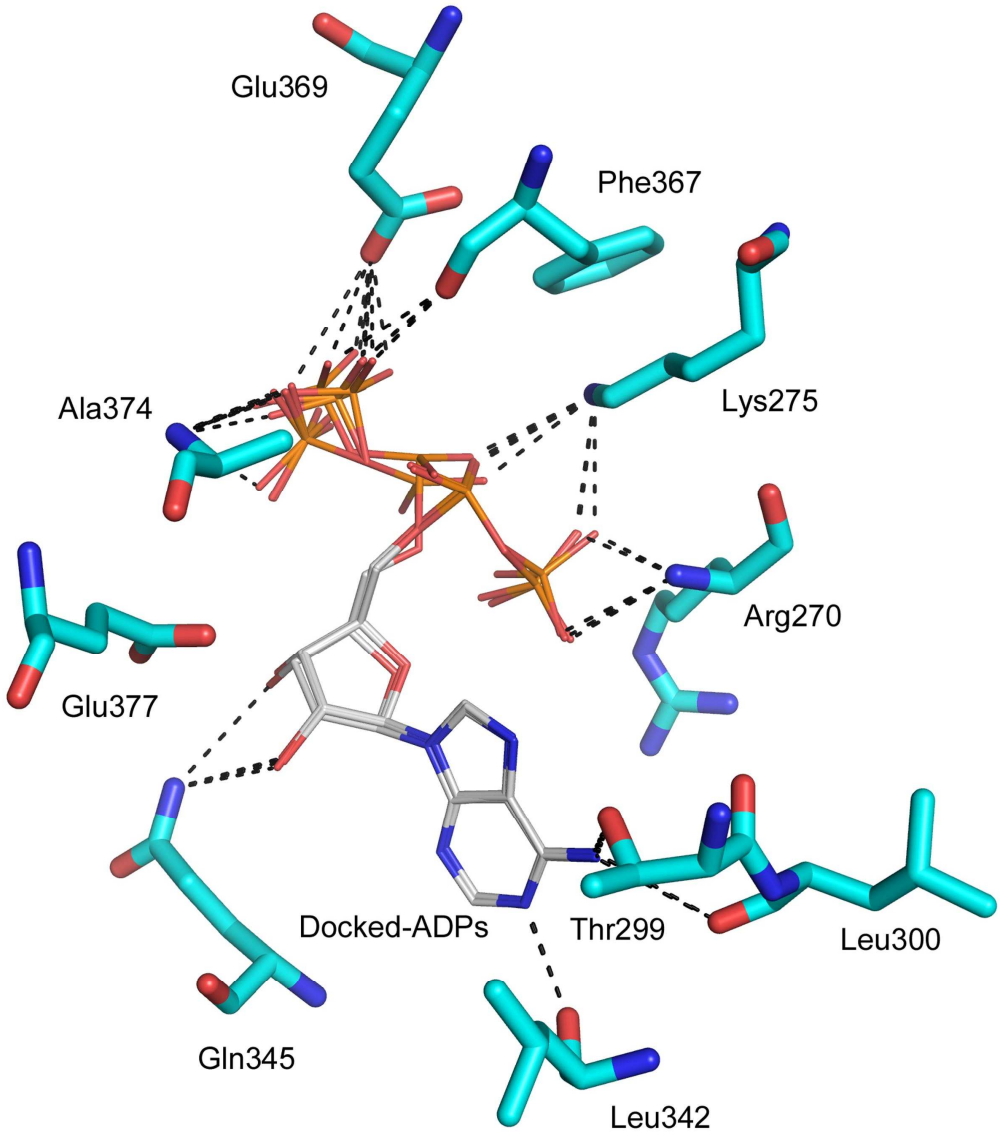


Figure 2.20 Ten docked models of ADP interacting with the binding residues of *H. orenii* SPS. Docked models of ADP (gray) obtained by the 'ab initio' method reveal the NDP-Glc binding mode of *H. orenii* SPS (cyan). From these models, most of the conserved residues are found to interact with the diphosphate group and the ribose moiety of the NDPs. Atoms are shown red (O) and orange (P). Hydrogen bonding contacts are shown in black dashed lines. This figure was prepared using the program PyMOL (DeLano, 2002).

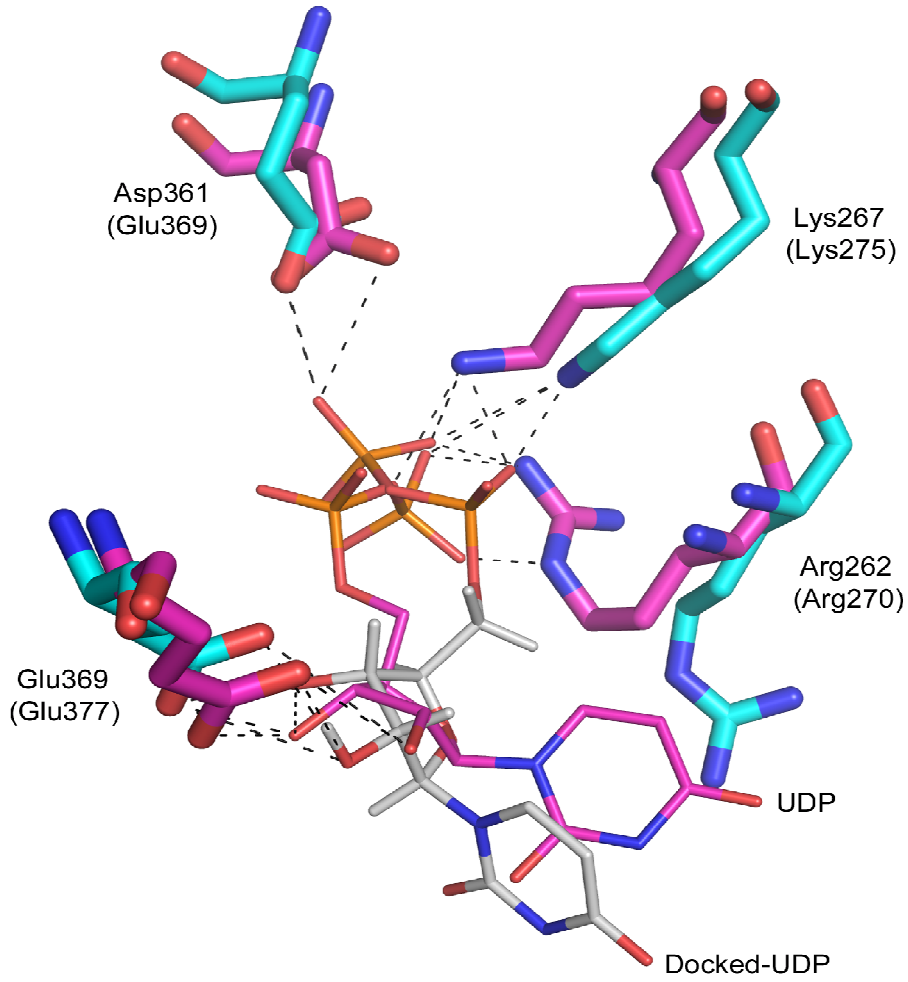


Figure 2.21 Superimposition of one docked-UDP ligand and the actual UDP ligand. The superimposition of the conserved, binding residues of SPS (cyan) and trehalose 6-phosphate synthase (magenta; PDB code 1GZ5, Gibson *et al.*, 2002) interacting with one UDP docked model (gray) and UDP ligand (magenta) respectively. The NDP-Glc binding mode of *H. orenii* SPS was deduced from the UDP-bound docked model by an 'ab initio' method. For figure clarity, only four of the conserved, key residues of *H .orenii* SPS: Arg270, Lys275, Glu369 and Glu377, and the corresponding residues of trehalose 6-phosphate synthase. The hydrogen bonding contacts are shown in black dashes. This figure was prepared using the program PyMOL (DeLano, 2002).

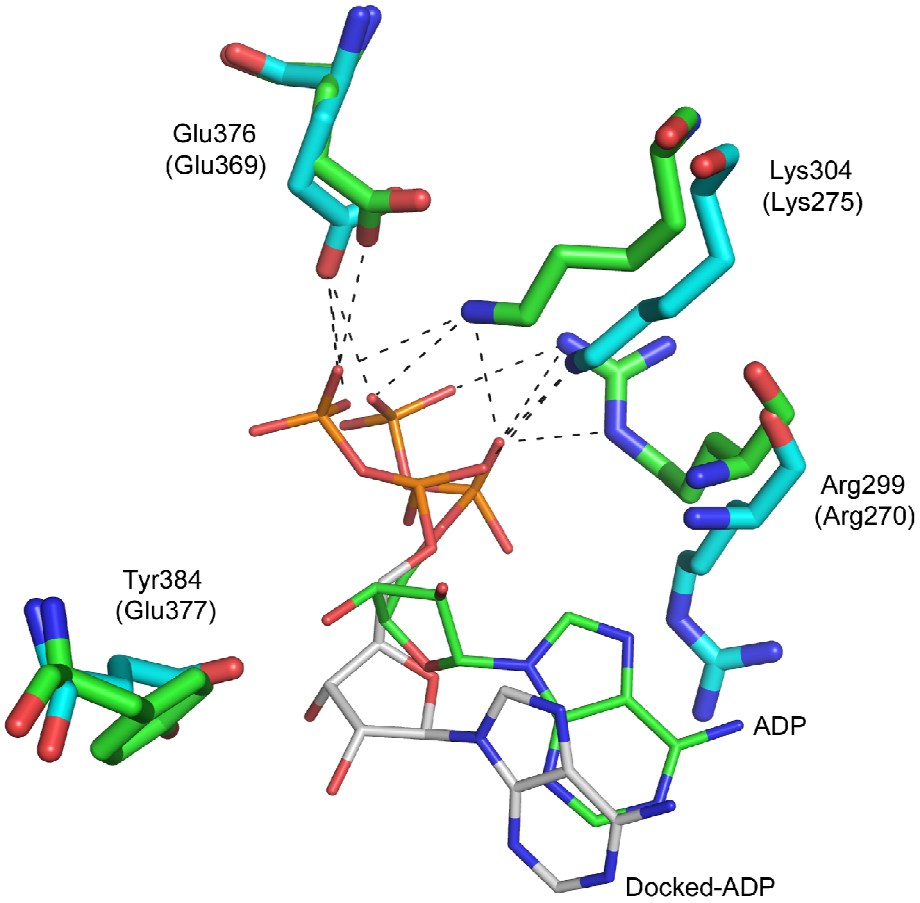


Figure 2.22 Superimposition of one docked-ADP ligand and the actual ADP ligand. The superimposition of the conserved, binding residues of SPS (blue) and corresponding residues of glycogen synthase (green; PDB code 1RZU, Buschiazzo *et al.*, 2004) interacting with one of the ADP docked models (gray) and its ADP ligand (green) respectively. The docked models of ADP deduced the NDP-Glc binding mode of *H. orenii* SPS by the 'ab initio' method. For figure clarity, only four of the conserved, key residues of *H .orenii* SPS: Arg270, Lys275, Glu369 and Glu377, and the corresponding residues of glycogen synthase are shown. The hydrogen bonding contacts are shown in black dashes. This figure was prepared using the program PyMOL (DeLano, 2002).

Cid *et al.* (2000) proposed the E-X_7-E motif for the C-terminal of GTs, in which the Glu residues have a catalytic role in the reaction (Cid *et al.*, 2000). This motif is also found in *H.orenii* SPS and residues Glu377 and Glu369 (Figure 2.13) are known as the motif positions 1 and 2 respectively (Cid *et al.*, 2000, Gibson *et al.*, 2002, Liu and Mushegian, 2003, Wrabi and Grishin, 2001). In SPS-UDP docked models, we observed that the carboxylate group of Glu369 interacts with the distal phosphate group of a few UDP docked models while that of Glu377 interacts with the ribose moiety in all ten docked models of UDP. Since all NDP-Glc share the same ribose and disphosphate backbone, we speculate that these conserved Glu residues will most certainly play an equivalent role in binding to other NDP-Glc donors.

2.3.9 Mechanism of action

The successful crystallization of the *H. orenii* enzyme provides the first opportunity to understand the structure of SPS from any organism. Based on our structural and bioinformatics analysis of the NDP-Glc binding pocket, in particular the detected similarity to retaining GTs of known structure (see above), we propose a possible mechanism of SPS action.

The inverting GT-A, in the presence of a DXD motif, adopts a divalent metal ion dependent catalytic mechanism, whereas the retaining GT-B, in the absence of such motif, exhibits the metal ion independent mechanism (Breton *et al.*, 2006, Buschiazzo *et al.*, 2004, Gibson *et al.*, 2002, Liu and Mushegian, 2003). Although the mechanism of retaining GTs is not well understood, Gibson *et al.* (2002) proposed a putative transition state for the transfer of glycosyl group by OtsA, which is metal ion independent (Gibson

et al., 2002). To verify the property of *H. orenii* SPS to bind to divalent metal ion with and without substrates, we made several attempts using Isothermal Titration Calorimetry (ITC) and co-crystallization / soaking experiments to trap the Mg^{2+} ion. None of the results supported the binding of Mg^{2+} under the conditions tested, and in the crystal structure no electron density corresponding to a divalent metal-ion was observed near the diphosphate groups of the docked NDPs.

Figure 2.23 shows superimposed model of catalytic regions of 2 complexes: (i) open SPS-F6P and (ii) the closed SPS-S6P-UDP model. The closed model of SPS was generated by independently superimposing the A- and B-Domain of SPS on the respective domains of the closed OtsA-UDP-G6P complex structure. Gly33, Gly34 and Gln35 of domain-A of SPS are highly conserved among SPS homologs (Figure 2.13). These three residues are also corresponding to the conserved Gly-Gly-Leu motif of OtsA (Gibson *et al.*, 2002). In the open SPS-F6P and SPS-S6P complex crystal structures, F6P (or S6P) forms hydrogen bonds with the main chain amide of Gly33 and Gln35 (Figure 2.18). In the closed SPS model (Figure 2.23), the main chain amide of Gly34 is found interacting with the diphosphate group of UDP at B-Domain, while Gly33 and Gln35 maintain the interactions with F6P (or S6P). In the case of closed OtsA structure, UDP at the B-domain interacts with the main-chain amides of the corresponding two glycines of the Gly-Gly-Leu motif at A-domain (Gibson *et al.*, 2002). Although the role of the SPS Gly33 is different from its corresponding Gly in OtsA, both the second glycine of the Gly-Gly-Leu motif in SPS and OtsA binds to the diphosphate group of UDP at their respective B-domain. Hence, based on the closed model of SPS and the OtsA complex

structure, Gly34 of SPS may play a crucial role in providing a linkage between NDP-Glc and A-domain, and may also involved in domain closure upon substrates binding.

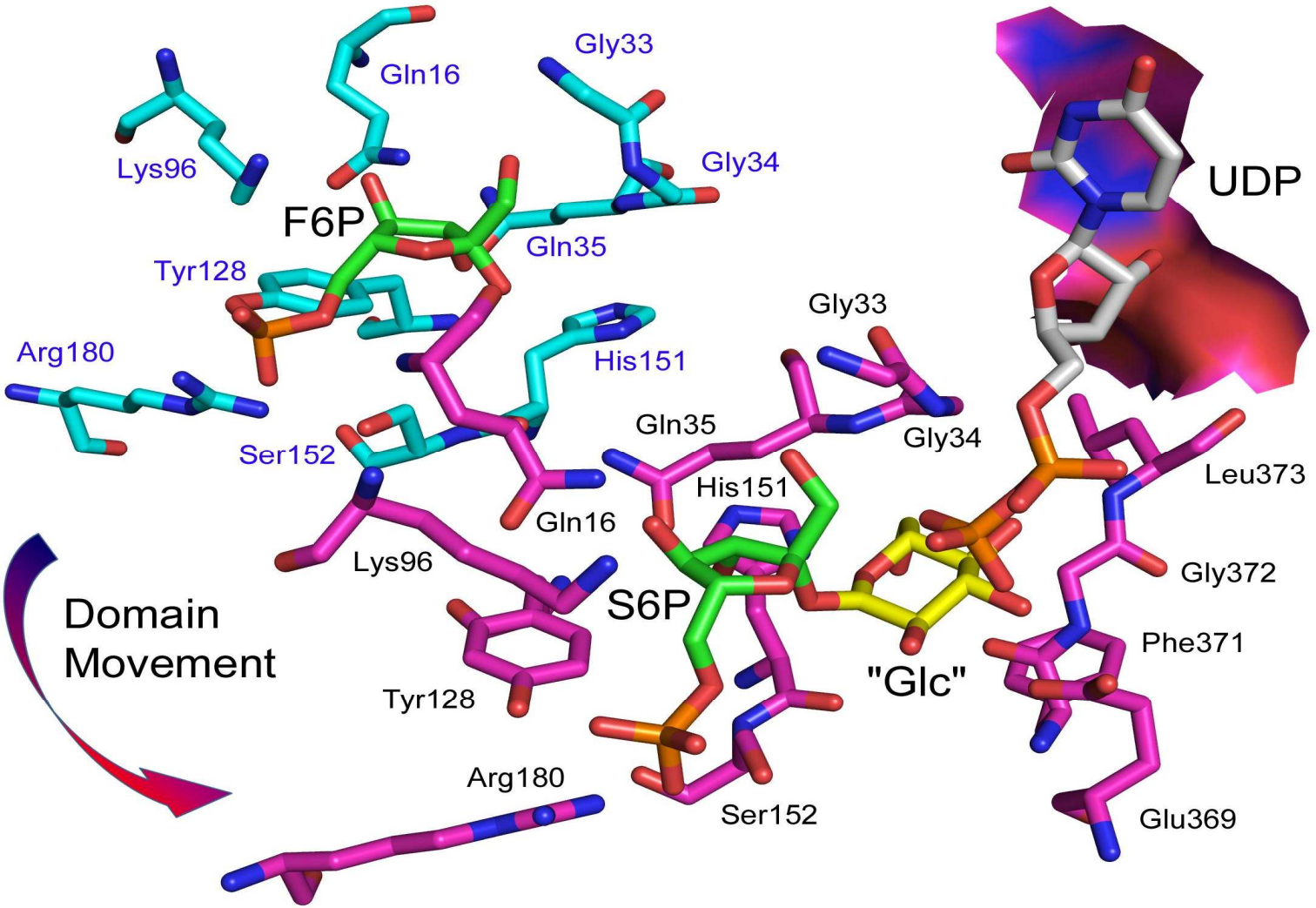

Gln16
Gly33
Lys96
Gly34
F6P
Gln35
Tyr128
Arg180
His151
Ser152
Gln16
Lys96
Domain
Movement
Tyr128
Arg180
Gly33
Gln35
Gly34
His151
S6P
"Glc"
Ser152
UDP
Leu373
Gly372
Phe371
Glu369

Figure 2.23 Superimposition of the catalytic regions of the open SPS-F6P complex (cyan) and the closed SPS-S6P-UDP model (magenta). SPS residues proposed to bind to the glycosyl group of S6P (or UDP-Glc) and its binding residues from A-domain are shown. An arrow illustrates the movement of the binding residues from A-domain upon domain closure. Carbon atoms of F6P and UDP are shown in green and gray, respectively. The carbon atoms of S6P are also shown in green with its glycosyl group shown in yellow. The rest of the atoms are blue (N), red (O) and orange (P). The superimposition was performed with DALI (Holm and Sander, 1993) and O program (Jones *et al.*, 1991). This figure was prepared using the program PyMOL (DeLano, 2002).

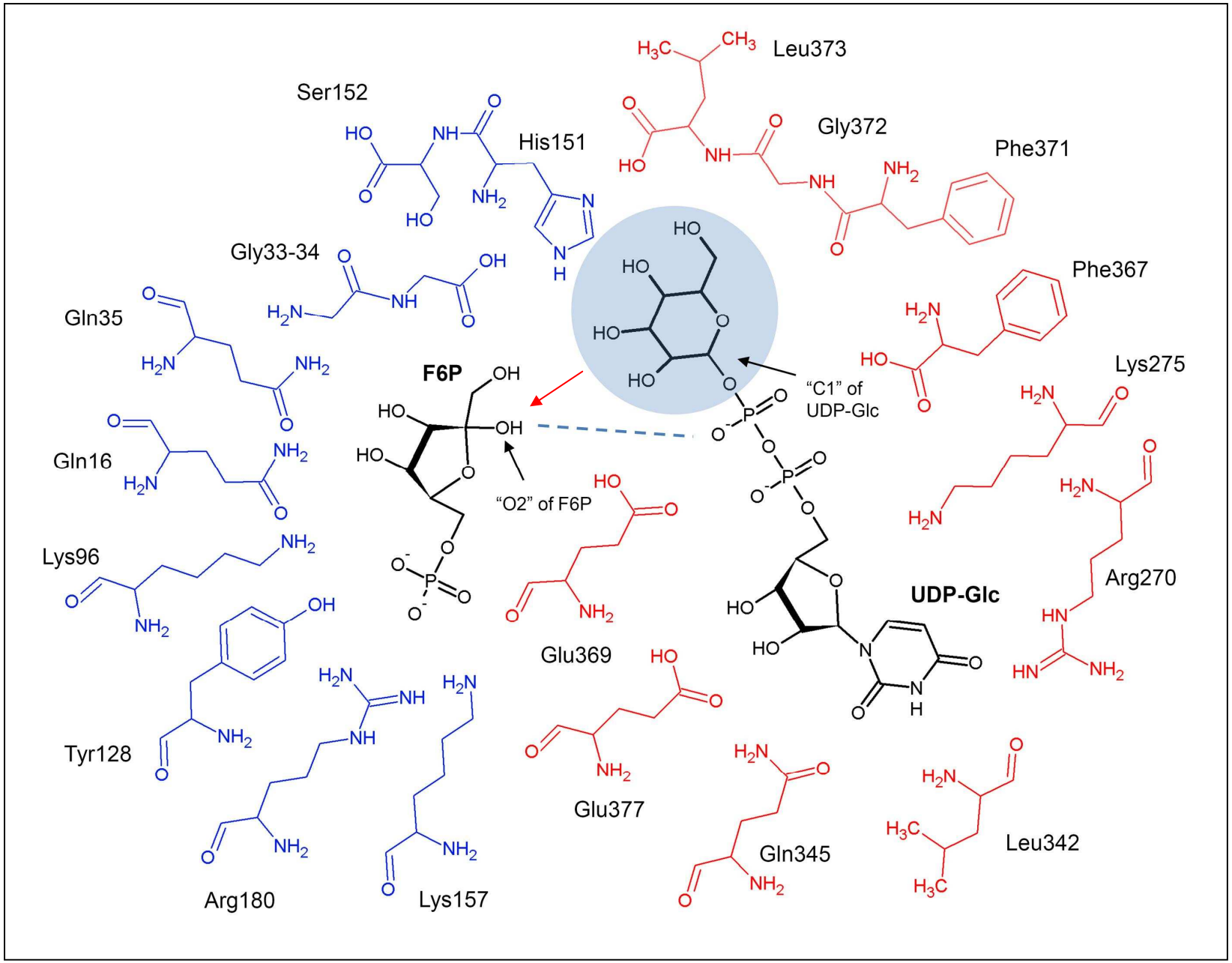

Leu373
Ser152
His151
Gly372
Phe371
Gly33-34
Phe367
Gln35
F6P
"C1" of
UDP-Glc
Lys275
Gln16
"O2" of F6P
Lys96
Arg270
UDP-Glc
Glu369
Tyr128
Glu377
Gln345
Leu342
Arg180
Lys157

Figure 2.24 Schematic diagram of the reaction between F6P and UDP-Glc in the binding cleft of SPS. The A- and B-domain binding residues and two substrates respectively are labeled in blue, red and black. The hypothetical hydrogen bond between O2 of the F6P and the diphosphate group of UDP-Glc is shown as a dotted line. Both the orientation of O2 of F6P and the C1 of UDP-Glc are labeled accordingly. The red arrow indicates the transfer of the glycosyl group (shaded) from UDP-Glc (sugar donor) to F6P (sugar acceptor).

The position of the glycosyl group of S6P in the closed SPS model is believed to be the catalytic reaction centre of SPS. The closed SPS model has revealed several interactions between the glycosyl group of S6P and SPS residues (Figure 2.23). Conserved residues Glu369, Phe371, Gly372, Leu373 (B-domain) and His151 (A-domain) of the closed SPS model are found to be interacting with the glycosyl group of S6P. The corresponding residues of OtsA are also found interacting with the glycosyl group of UDP-Glc in the closed OtsA-UDP-Glc complex structure (PDB code 1UQU, Gibson *et al.*, 2004). The UDP-Glc in this complex structure is constrained to adopt a folded shape by these interactions (Breton *et al.*, 2006, Gibson *et al.*, 2004). Interestingly, the same conformation of a glycosyl group was also observed in UDP-Glc-OtsA complex structure.

In the structure of the SPS-F6P complex, atom O2 of F6P is found to have a strong hydrogen bond (<3.0 Å) with one water molecule. By comparison with the structure of OtsA, this water molecule may get replaced by the phosphate of the incoming donor molecule (Breton *et al.*, 2006). In addition, it is possible that as the two domains close upon binding of a second substrate NDP-Glc, a hydrogen bond is established

between the atom O2 of the F6P and the diphosphate group of NDP-Glc, which is observed in the closed SPS model (Figure 2.24). This hydrogen bond lowers energy barrier, facilitates the formation of a late oxonium-ion like transition-state, as a result of a nucleophilic attack by the deprotonated atom O2 of F6P at the weakened, anomeric C1 of NDP-Glc, leading to the cleavage of NDP-Glc (Breton *et al.*, 2006, Gibson *et al.*, 2002).

In the SPS-F6P and the SPS-S6P complexes, highly conserved His151 from the A-domain of SPS is found to be the only residue that binds to the glycosyl group of the product S6P and has no interaction with F6P (Figure 2.18). Previously, for the OtsA-G6P-UDP complex (Gibson *et al.*, 2002), Gibson *et al.* had proposed a possible interaction of the corresponding His154 with the glycosyl group of UDP-Glc (substrate), which was later confirmed with the OtsA-UDP-Glc complex (Gibson *et al.*, 2004). According to the closed SPS model, while the conserved Gly34 from A-domain interacts with the diphosphate group of UDP at B-domain, His151 remains the only residue from the A-domain interacting with the glycosyl moiety of S6P (or UDP-Glc). Similar to Gly34, His151 may provide a linkage between NDP-Glc and A-domain of SPS, and possibly involved in domain closure upon substrates binding. Most importantly, we propose an active role for the conserved His151 during the transfer of the glycosyl group from NDP-Glc bound to the B-domain to F6P on the A-domain, resulting in the formation of S6P and followed by its release from this domain.

Chapter III

Mechanism of Action and Structure of Fructokinase from *Halothermothrix orenii*

3.1 Introduction

Phosphorylation of free monosaccharides (glucose and fructose) is the initial step of metabolic pathways. Sugar kinases are broadly classified into three superfamilies: the galactokinases, hexokinases and ribokinases. Members of the galactokinase family are involved in diverse pathways, ranging from cholesterol and amino acid synthesis to galactose phosphorylation.

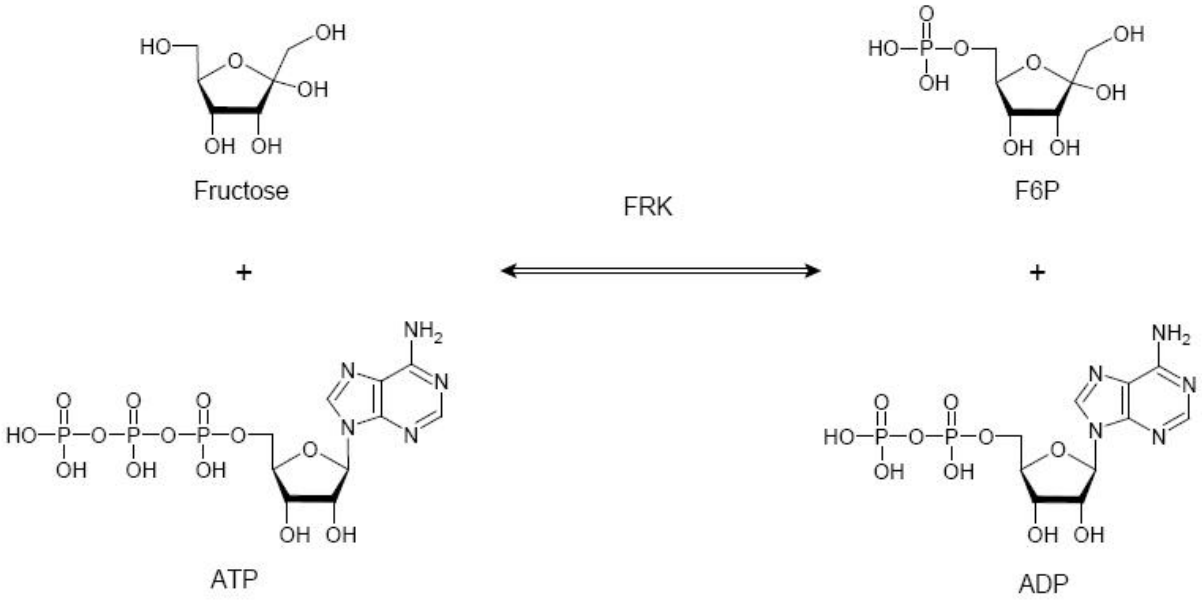


Figure 3.1. A schematic diagram of the reaction involving FRK and Fructose. The synthesis of F6P involves the action of FRK, which catalyzes the phosphorylation of fructose to fructose-6-phosphate (F6P). Phosphate group from ATP to a fructose an acceptor, resulting in the formation of F6P and ADP, a central and regulatory process in sucrose mobilization of plants and bacteria.

Sucrose is the major saccharide in plants; two enzymes are responsible for the phosphorylation of sucrose cleavage products fructose and glucose. Fructokinase (FRK; EC 2.7.1.4) is a ubiquitous, highly specific enzyme that primarily catalyzes the transfer of a phosphate group from adenosine triphosphate (ATP) donor to a saccharide acceptor D-fructose resulting in the formation of D-fructose 6-phosphate (F6P) and ADP (Baker *et al.*, 2001). Hexokinases (Hxk; EC 2.7.1.1) preferentially phosphorylates glucose (Figure 3.1).

FRK was first reported in 1956 although it was only isolated and characterized 20 years later. It belongs to the ribokinase superfamily of sugar kinases and evolutionary tree suggests that family divergence of the fructokinase from the ribokinase ancestor occurred prior to species divergence, thus explaining the high substrate specificity compared with hexokinases. FRK specifically phosphorylates fructose with a Km of 41-220µM, at a pH 8.0 and have much higher affinities for fructose than Hxk (Renz and Stitt, 1993). As fructose phosphorylation by FRK is irreversible and near rate-limiting, it is important for regulating the rate and localization of carbon usage by channelling fructose into a metabolically active state for glycolysis in plants and bacteria (Zhang *et al.*, 2003). This reaction is particularly important in plant tissues where sucrose assimilation and its conversion to starch or other storage sugars are in progress (e.g. tubers, seeds, fruits). In sink tissues where sucrose degradation is mediated by invertase and/or sucrose synthase to produce fructose, fructose must be phosphorylated to maintain the carbon flux to starch or respiration. FRKs are widely reported to have a preference for ATP over other nucleotides as the principle source of phosphate, although the enzyme is also able to utilize GTP or UTP when present in high concentrations, ATP will be the

principle source of phosphate (Chaubron *et al.*, 1995; Mertinez-Barajas *et al.*, 1997). Mg^{2+} is an essential cofactor for this reaction while in some cases, K+ was reported to improve the enzyme's activity (Chaubron *et al.*, 1995).

Sequence alignment of different plant species FRKs revealed significant sequence conservation in the ATP and sugar binding pockets. There are two known isoforms of FRK, both differing in regulation by substrate and cellular location. Analytical ultra-centrifugation studies suggests that FRK1 is associated with the chloroplast (Schnarrenberger *et al.*, 1990). In barley and tomato, FRK1 is constitutively expressed and shown to exhibit little substrate inhibition properties (Baysdorfer *et al.*, 1989; Kanayama *et al.*, 1998).

FRK2 on the other hand, forms the major pool of FRK that is located in the cytosolic fraction. It is involved in stem and root growth as well as storage organ development (Dai *et al.*, 2002b; Davies *et al.*, 2005). Its expression is predominantly sink and source leaves specific (Kanayama *et al.*, 1997; Kanayama *et al.*, 1998). Suppression of FRK2 resulted in stunted growth, reduction in flower number, seeds per fruit, tuber number and size (Odanaka *et al.*, 2002; Davies *et al.*, 2005) but overexpression did not result in elevated levels of tuber growth (Davies *et al.*, 2005). FRK2 is potently inhibited by fructose with Ki values of 1-6mM in barley, tomato, pea and maize. As the expression patterns of FRK2 and sucrose synthase (SS) activity are correlated in several plant species (Schaffer and Petreikov, 1997), and SS exhibits similar inhibition rates of FRK2 by fructose, it has been postulated that FRK2 plays a role in starch production in sink tissues where SS cleaves incoming sucrose. The activity of fructokinase greatly exceeds glucokinase in many tissues and this finding is consistent

with view that sucrose synthase rather than invertase is a major route for sucrose degradation, resulting in a larger amount of fructose being produced over glucose.

In the course of a random sequence analysis of the *Halothermothrix orenii* genome, the FRK open reading frame (ORF) was identified (Mijts *et al.*, 2001). Sequence database searches also revealed a large family of sequences similar to *H. orenii* FRK (HoFRK). HoFRK shared between 30 and 40% sequence identity (50 to 60% similarity) with the top twelve aligned sequences and conserved residues from bacteria and plants. No structural characterization available for FRK in the literature, although several ribokinase structures are known to-date (Sigrell *et al.*, 1998).

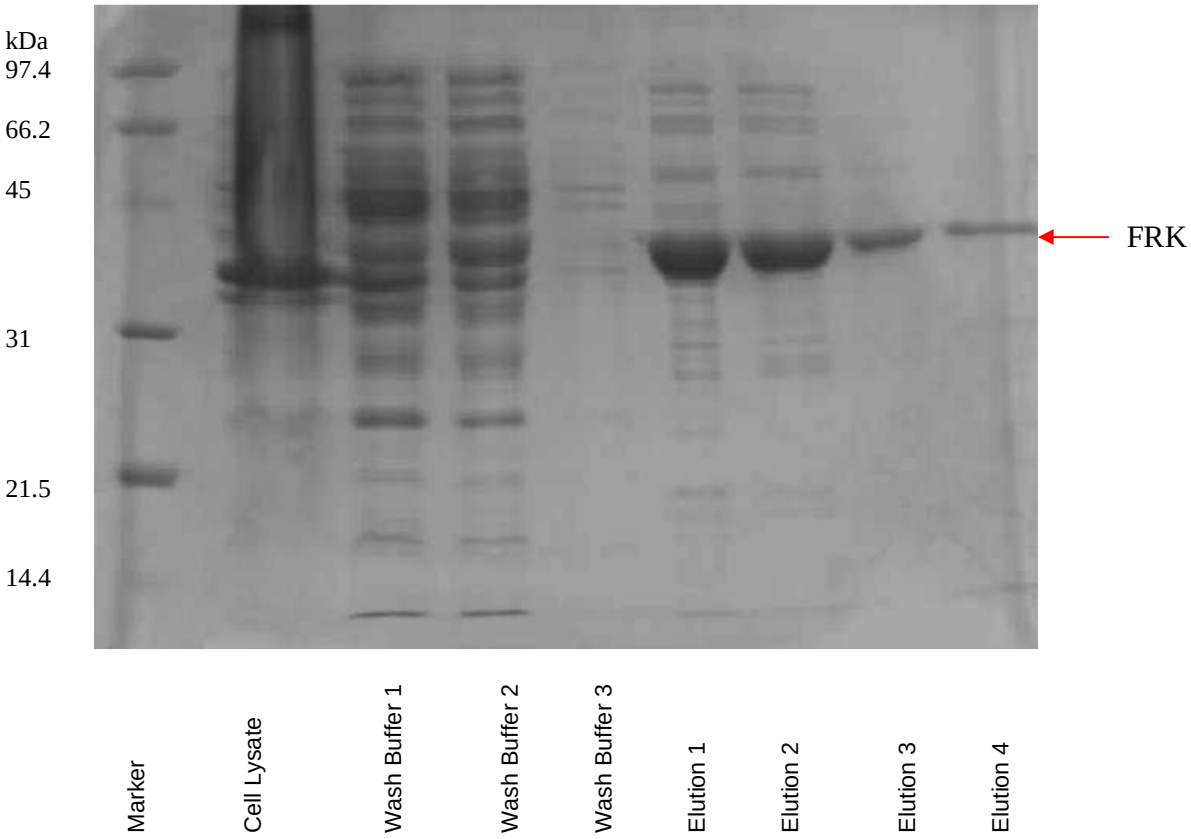
Here we report a 2.8Å resolution crystal structure of *H. orenii* FRK and a proposed mechanism for the phosphorylation of fructose. Comparative analysis revealed a close similarty to plant FRKs and thus refined up to 2.8Å resolution. The report on *H. orenii* FRK provides an insight into their structure and function of FRK from plants with which it shares a close similarity. Based on comparative analysis of FRK structure and combined with literature, we propose a mechanism for phosphorylation of fructose.

3.2 Material and Methods

3.2.1 Cloning, expression and purification.

Primers containing BamHI and KpnI restriction sites at the 5' and 3' ends respectively were used in PCR to amplify the FRK gene. The PCR product was digested by these restriction enzymes, followed by its ligation with the pTrcHisA expression vector (Invitrogen) encoding an N-terminal, non-cleavable His_6 tag (Mijts *et al.*, 2000). The plasmid was transformed into BL21 (DE3) and grown in 1 L of LB broth with

0.1mM Ampicillin at 37°C until it reached an optical density (OD_{600nm}) of about 0.6-0.7. The culture was cooled and induced with 1mM IPTG overnight at 25°C. The *H. orenii* FRK has 327 amino acid residues with a molecular weight of 36.074 kDa. The recombinant *H. orenii* FRK, consisting of a hexahistidine tag and a linker, was expressed as a 40.359 kDa protein. The cells were harvested by centrifugation (9000xg; 30min, 4°C) and resuspended in 30 ml of 20mM Tris-HCl pH 7.5, 200mM NaCl and 10mM imidazole and 1 tablet of EDTA-free Complete™ Protease Inhibitor Cocktail (Roche Diagnostics). Selenomethionine-substituted FRK was expressed using methionine auxotroph *E.coli* DL41 in LeMaster medium supplemented with 25mg/L selenomethionine (SeMet). The cells were lysed by sonication, followed by centrifugation at 11000rpm (Eppendorf 5804R) for 30min. Cell lysate was transferred to a chromatography (affinity) column containing Ni-NTA agarose (Qiagen). 1h of incubation was performed at 25°C with gentle agitation. The non-cleavable His_6-tag SPS was eluted with 500mM imidazole following three wash steps to remove non-specific binding. In the 12.5% SDS-PAGE viewed by Coomassie staining, the purified FRK migrated as a single band (Figure 3.2a) just between the 31kDa and the 45kDa protein ladder (SDS-PAGE Molecular Weight Standard, Low-range by BioRad). The recombinant FRK was further purified using FPLC Hiload 16/60 Superdex200 gel filtration column using AKTA FPLC UPC-900 system (Amersham Biosciences) and eluted at the flow-rate of 0.5ml/min as a single peak (Figure 3.2b) at 80ml in 20mM Tris-HCl pH7.5, 0.2M NaCl and 10mM dithiothreitol (DTT). This was followed by ultrafiltration to bring to a final concentration of the recombinant FRK to 10 mg/ml (Bradford method, Bradford, 1976).


kDa
97.4
66.2
45
31
21.5
14.4
FRK
Marker
Cell Lysate
Wash Buffer 1
Wash Buffer 2
Wash Buffer 3
Elution 1
Elution 2
Elution 3
Elution 4

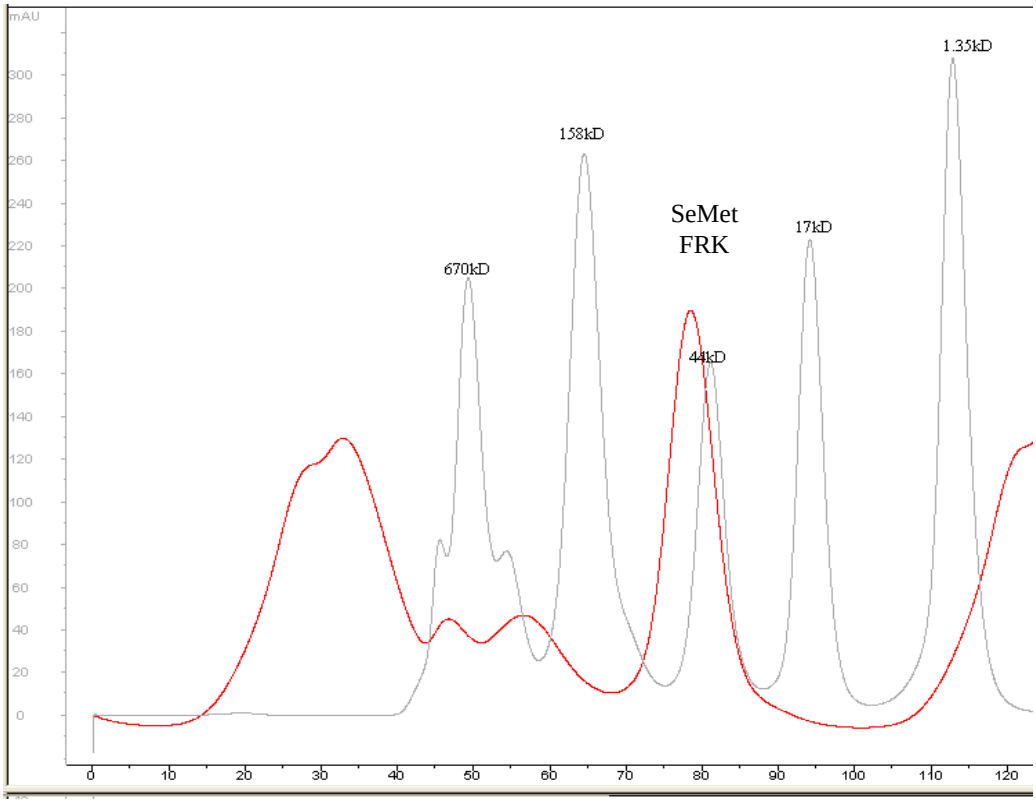


Figure 3.2 Top a) SDS-GEL image of purified FRK. Bottom b) Gel filtration profile of FRK. (a) The purified FRK migrated as a single band between 45 kDa and 31 kDa (Protein Ladder, SDS-PAGE Molecular Weight, Low-range by BioRad) in 12.5% SDS-PAGE, viewed by Coomassie staining. (b) The X-axis indicates the elution volume in mL and the Y-axis indicates the UV absorbance at 280 nm measured in mAU (arbitrary units). The elution profile is for protein injected into FPLC Hiload 16/60 Superdex75 gel filtration column (Amersham Biosciences).

3.2.2 Crystallization.

Crystallization screen was carried out through hanging-drop vapour-diffusion method using Hampton Research (Aliso Viejo, CA, USA) screens as well as by micro batch under-oil technique using JB crystallization screens (Jena Biosciences, Jena, Germany) at room temperature. Initially, apo and SeMet FRK crystals were small in size and appeared after two weeks. After extensive optimization, only a few out of the many crystals that grew were of diffraction quality (Figure 3.3). Obtaining the diffraction quality crystals was the most challenging aspect in this project. The present data set is the best of many data sets collected. As an approach to improve the data quality, we have also attempted to co-crystallize/soak with the substrates. So far no complex was crystallized. The best diffraction quality crystals were obtained from 8% PEG 4000, 0.8M $LiCl_2$ and 0.1M Tris-HCl, pH 8.5 by using micro batch under-oil technique with 2µl of the crystallization solution mixed with 2µl of protein under 15µl of paraffin oil. Native and SeMet crystals diffracted up to 2.8 Å and belonged to space group $P2_1$ with a = 43.85, b=172.84, c =47.01 Å and β=113.29º.

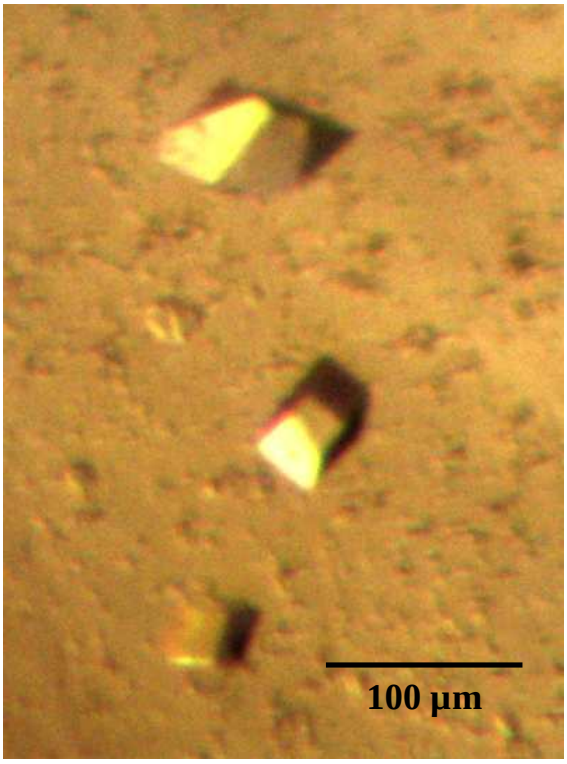


Figure 3.3 Crystal of SeMet FRK. SeMet FRK crystals were obtained by the micro batch under-oil technique.

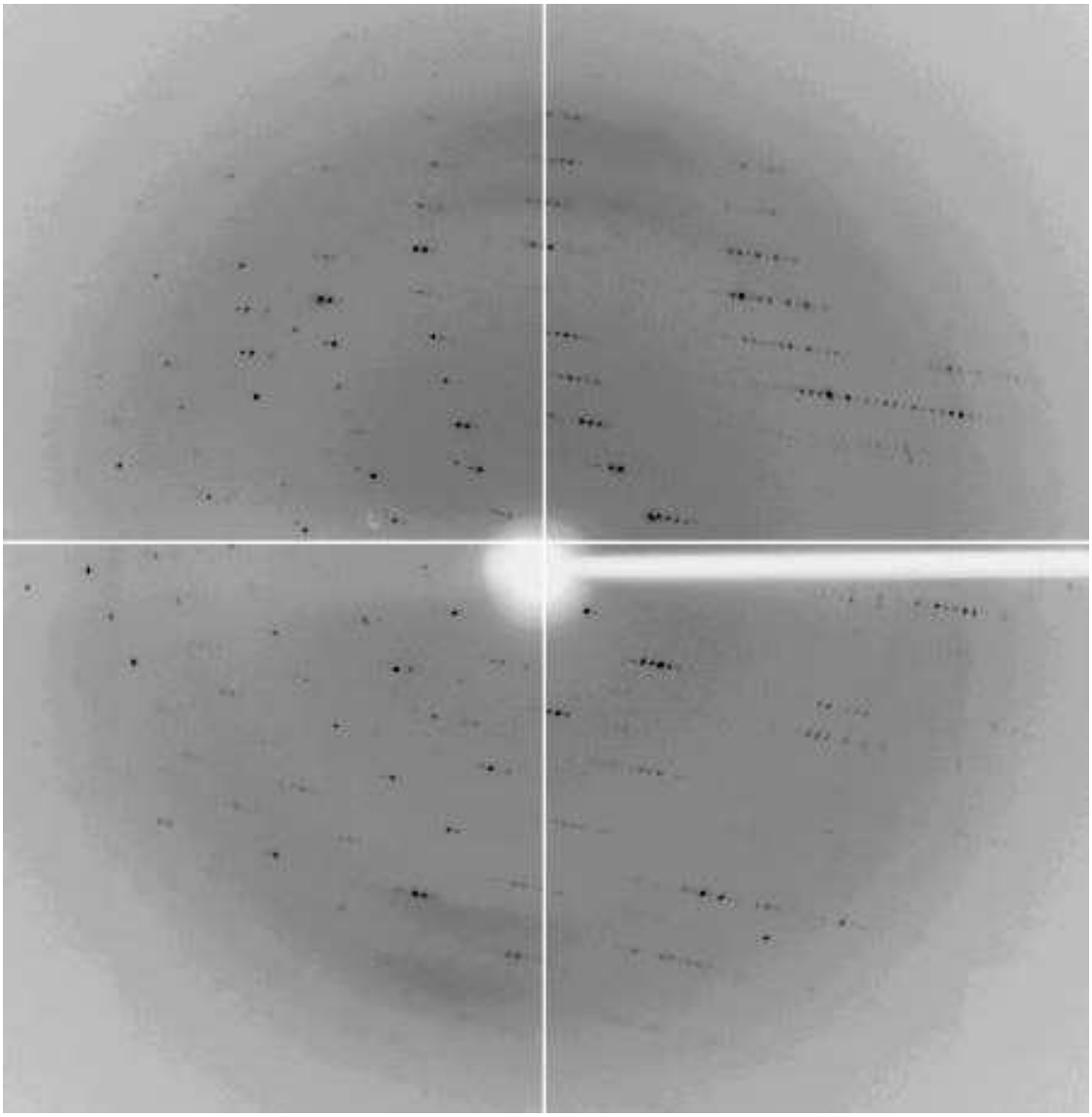


Figure 3.4 Sample diffraction pattern of SeMet FRK crystal. Diffraction pattern collected from ADSC Q210 diffractometer system at X12C beamline (NSLS, BNL) for SeMet FRK crystal.

Table 3.1 Data collection and refinement statistics.

Data set	Peak	Inflection	Remote[a]
Data collection			
Resolution range (Å)	50.0-2.8	50.0-28	50.0-2.8
Wavelength (Å)	0.9790	0.9794	0.9600
Observed reflections > 1	117292	117115	116025
Unique reflections	30770	30424	30193
Completeness (%)	99.9	99.9	99.9
Overall (I/σI)	19.8	18.9	16.7
R_{sym} (%)[b]	4.6	4.5	4.9
Refinement[c] and quality			
Resolution range (Å)			50.0-2.8
R_{work} (no. of reflections)[d]			0.2541
R_{free} (no. of reflections)[e]			0.2880
R.M.S.D. bond lengths (Å)			0.008
R.M.S.D. bond angles			1.59
Average B-factors (Å^2)			
Main-chain			40.37
Side-chain			40.45
Ramachandran plot[f]			
Most favored regions (%)			82.6
Additional allowed regions (%)			15.0
Generously allowed regions (%)			1.5
Disallowed regions (%)			0.9

[a]NCS restraint was kept throughout the refinement of the remote dataset.
[b]$R_{sym}=|I_i-<I>| / |I_i|$ where I_i is the intensity of the ith measurement, and <I> is the mean intensity for that reflection.
[c]For all models, reflections with I>σI was used in the refinement.
[d]$R_{work}=100 \times \Sigma|F_P-F_{P(calc)}|/\Sigma F_P$.
[e]R-free was calculated with approximately 2000 reflections in the test set.
[f]Statistics for the Ramachandran plot from an analysis using PROCHECK (Laskowski et al, 1993).

3.2.4 Data collection, structure solution and refinement.

Crystals were directly taken from the drop, and flash cooled in a N_2 cold stream at 100ºK. The SelMet FRK crystals were diffracted up to 2.8 Å resolution using an R-axis IV++ image plate detector mounted on a RU-H3RHB rotating anode generator (Rigaku Corp., Tokyo, Japan). Synchrotron data were collected at beam lines X12C and X29, NSLS, Brookhaven National Laboratory for the SeMet protein (Figure 3.4). Complete MAD datasets were collected at three wavelengths (Table 1) using Quantum 4-CCD

detector (Area Detector Systems Corp., Poway, CA, USA) to 2.8 Å resolution. Data was processed and scaled using the program HKL2000 (Otwinowski and Minor, 1997)).

3.2.5 Structure solution and refinement.

All four selenium sites in the asymmetric unit were located by the program SOLVE (Terwilliger and Berendzen, 1999). Initial phases were further developed by RESOLVE (Terwilliger, 2000) and improved the overall figure of merit (FOM) to 0.66 which made it possible to build automatically approximately 50% of the residues of one asymmetric unit. The remaining parts of the molecules were built manually using the program O (Jones *et al.*, 1991). Further several cycles of model building alternating with refinement using the program CNS (Brunger *et al.*, 1998) resulted in the final model, with an R-factor of 0.254 (R_{free}=0.288) to 2.8 Å resolution with reflections $I>\sigma I$ was used in the refinement. The final model comprises of 276 residues (Leu22-Ile306) and 114 water molecules. The His tag with the linker, first N terminal 21 residues and the C-terminal 21 residues were not visible in the electron density map. PROCHECK (Laskowski *et al.*, 1993) analysis shows two residues in the disallowed regions of the Ramachandran plot.

3.3 Results and discussion

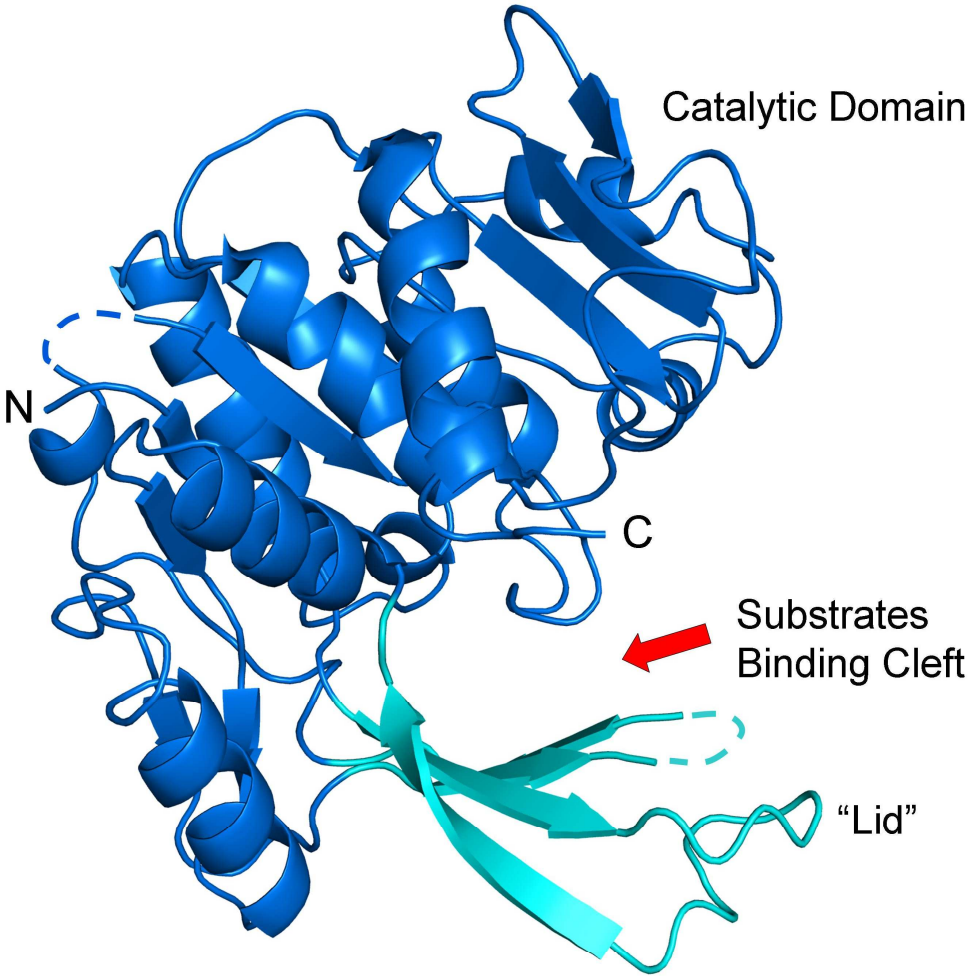
3.3.1 Overall structure.

The crystal structure of recombinant FRK was determined by the MAD method using the synchrotron data set and refined up to 2.8Å resolution (Figure 3.5, Table 3.1). Each FRK monomer consists of residues from Leu22 to Ile306. Neither the N-terminal 21 residues nor the C-terminal 21 residues had interpretable electron density map and were not modelled. FRK crystallized with two molecules in the asymmetric unit and they are related by a 2-fold noncrystallographic symmetry is approximately parallel to the *a*-axis. Interestingly, these two molecules are packed one over other through the lid region β-strands, resulting in a formation of a continuous β-sheet extending across the dimer interface which stabilizes the dimeric structure. Gel filtration chromatography experiment indicated that FRK exists as a dimer in solution which is consistent with the dimeric arrangement observed in the crystal structure.

Each FRK molecule consists of a mixed α/β fold; a characteristic nucleotide binding domain resembles Rossmann fold (Leu22-Ile30, Ser56-Thr108, Ala128-Ile306) (hereafter referred as catalytic domain) and a small β-sheet "lid" (or lid region) (Leu31-Gly55, Thr109-Glu127). The substrate binding cleft is located at the interface between the catalytic domain and the lid region with a dimension of approximately 18 Å wide and 22 Å deep. The catalytic domain has a central core β-sheet consisting of eight mostly parallel β-strands flanked on both sides by eight helices; of which two are 1-2 turn small helices. The lid region from both monomer forms a tilted antiparallel β-sheet consisting of four strands from each monomer and runs from one monomer to other monomer at the dimer interface. This β-sheet maintains the dimeric architecture of FRK, with the dimer

having approximate dimensions of 90 x 40 x 36 Å. The observation of a dimeric FRK in solution as well as in crystal structure suggests a functionally important role for dimerization.

.



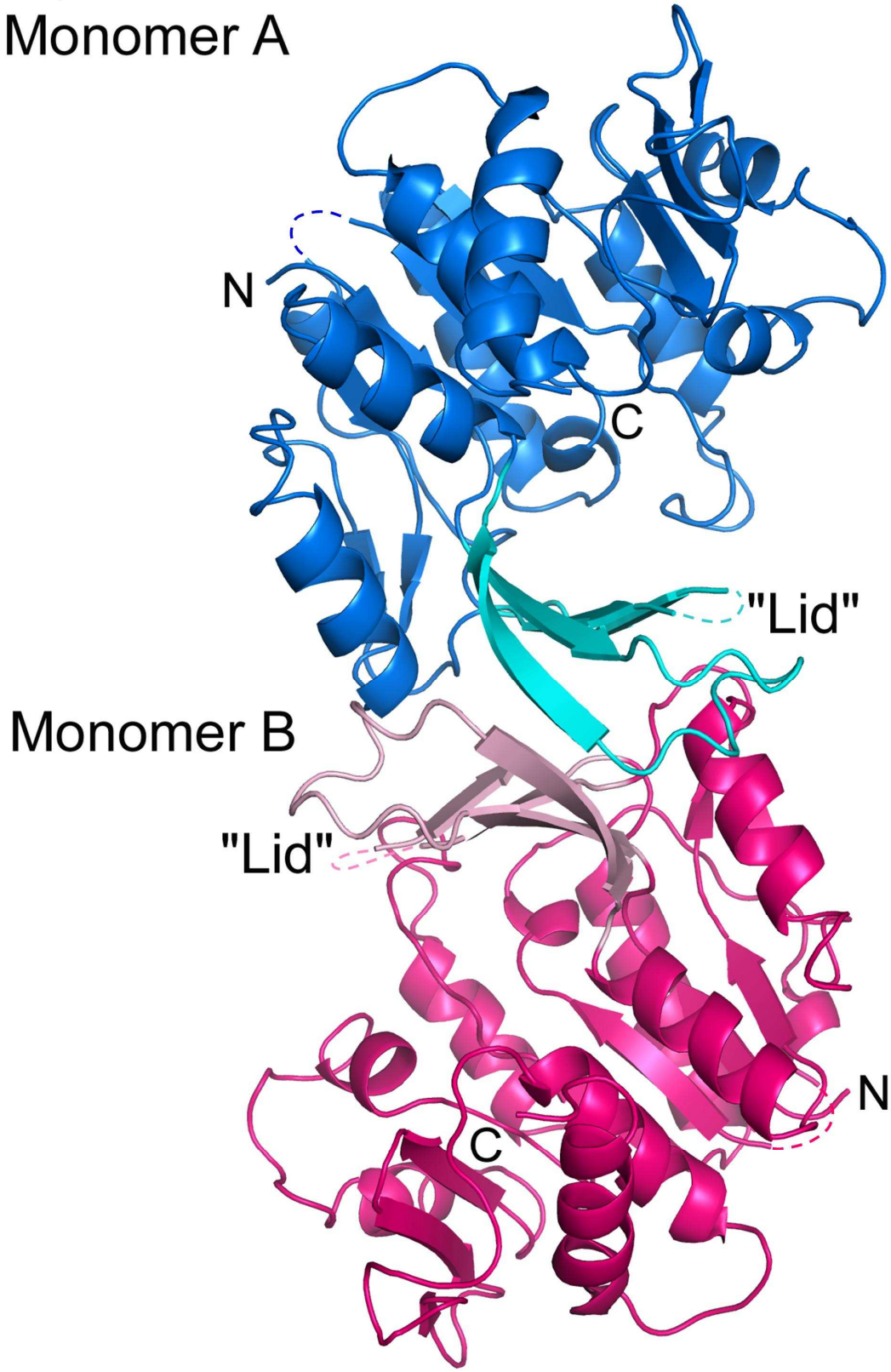

Monomer A
N
C
"Lid"
Monomer B
"Lid"
N
C

Figure 3.5 Crystal structure of *Ho*FRK. ***(a)*** Shows the ribbon representation of the *Ho*FRK monomer. **(b)** Ribbon diagram showing the dimeric *Ho*FRK in the asymmetric unit. The catalytic domain (residues Leu22-Ile30; Ser56-Thr108 and Ala128-Ile306) is depicted in blue and the β-sheet "lid" region (residues Leu31-Gly55 and Thr109-Glu127) in red. The N- and C-terminals are labelled. These figures were prepared using the programs PYMOL (DeLano, 2002).

3.3.2 Sequence and structural similarity

A search for proteins homologous to *H. orenii* FRK (*Ho*FRK) was performed against all bacteria and plant Genbank sequences using BLAST (Altschul *et al.*, 1990). The four bacterial and plant FRK sequences that were significantly most similar to *Ho*FRK were chosen for sequence alignment using ClustalW (Larkin *et al.*, 2007) are shown in Figure 3.6. Overall, *H. orenii* FRK shares between 30 and 40% sequence identity (50 to 60% similarity) with the sequences from bacteria and plants. The bacterial FRK most similar to *H. orenii* FRK belonged to *Petrotoga mobilis*, a thermophilic eubacteria of the family Thermotogaceae. This is followed by FRKs from flavobacteria *Polaribacter dokdonensis, Dokdonia donghaensis and Psychroflexus torques*. Most similar plant FRKs were taxonomically varied. The closest homologous sequence was from *Solanum lycopersicum* (potato), followed by FRK from *Arabidopsis thaliana, Zea mays* (maize), and *Beta vulgaris* (sugar beet) (Figure 3.6).

A search for *Ho*FRK structural homologs was performed using the program DALI (Holm and Sander, 1993). Structures showing overall structural similarity particularly belong to the ribokinase superfamily of proteins, and the most common feature of these proteins is the substrate binding cleft. These structural similarities corresponded to similarities in protein sequences observed via a BLAST search of protein sequences from the PDB. The highest structural similarity is observed between *Ho*FRK and AIR kinase (PDB code 1TZ6) yielding an rmsd of 2.0Å for 262 Cα atoms, with 24% identity. This is followed by KDG kinase (PDB code 1V19; rmsd=2.3Å for 261 Cα atoms; 26% identity) and ribokinase (PDB code 1RKD; rmsd=2.2Å for 252 Cα atoms; 23% identity). In addition, a recently deposited pdb on FRK from *Bacteroides thetaiotaomicron* VPI-5482

(PDB code 2QHP; rmsd=2.2Å for 247 Cα atoms; 17% identity) is structurally related. In several cases the individual domains of FRK and ribokinase homologs superpose well, although the relative disposition of the two domains often varies, especially the β-sheet lid region.

The structure based sequence alignment of *Ho*FRK with the homologs from ribokinase family showed that residues were predominantly conserved in substrate-binding pockets (Figure 3.7). Incidentally, AIR kinase (1TZ6), KDG kinase (1V19) and ribokinase (1RKD) were crystallized as complexes with their respective substrates. KDG kinases and ribokinases utilize ATP as a phosphate donor and complexed with ATP or ADP, we analysed the structure alignment with *Ho*FRK to infer the conserved ATP binding residues of *Ho*FRK. The mode of ATP binding can be classified into two broad categories: phosphate binding and base (A) binding. The phosphate binding residues mainly interact with ATP by direct or water mediated hydrogen bonds, while residues binding to the adenine interact hydrophobically. Although all ATP binding residues are found in the latter half of the protein, conserved residues interacting with adenine are found clustered around Phe263 to Ala267 and Asn299 of *Ho*FRK.

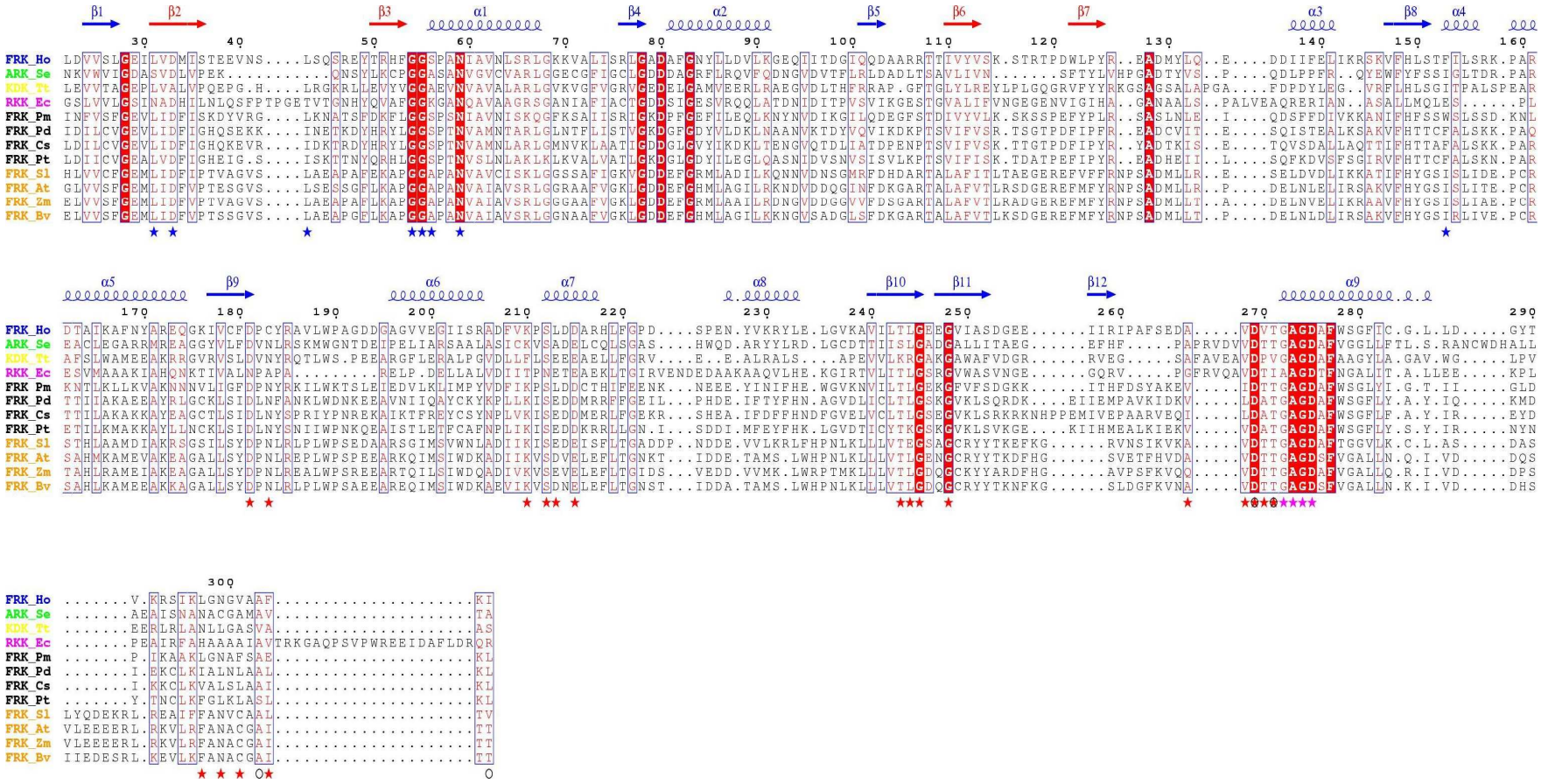

FRK_Ho
ARK_Se
KDK_Tt
RKK_Ec
FRK_Pm
FRK_Pd
FRK_Cs
FRK_Pt
FRK_Sl
FRK_At
FRK_Zm
FRK_Bv

Figure 3.6 Structure based sequence alignment of *Ho*FRK. (a) Top 4 rows: Structure based sequence alignment of *Ho*FRK (blue), ARK (PDB code 1TZ6, green), KDK (PDB code 1V19, yellow) and RK (PDB code 1RKD, magenta). The amino acids are in one-letter codes; the conserved residues are highlighted. Strictly conserved residues are shaded red with semi-conserved residues lettered in red. Secondary structural elements of *Ho*FRK belonging to the α/β domain and the β "lid" are shown in blue and red respectively. This figure was created using the program ESPript (Gouet *et al.*, 1999). Middle 5-8th rows: Sequence alignment of *Ho*FRK (top, blue) with the closest four *Ho*FRK homologs (black) was carried out using ClustalW (Larkin *et al.*, 2007) and ESPript (Gouet *et al.*, 1999). Bottom 9-12th rows: Sequence alignment of *Ho*FRK (top, blue) with the closest four FRK plant homologs (orange). The anion hole motif GAGD is indicated by magenta asterisks. Proposed key substrate binding residues of fructose and ATP are indicated by blue and red asterisks respectively. Proposed K^+ coordinating residues are indicated by open circles. Suffix: FRK_Ho: FRK, *H. orenii;* ARK_Se: Aminoimidazole riboside kinase, *Salmonella enterica* (1TZ6); KDK_Tt: 2-Keto-3-Deoxygluconate Kinase, *Thermus thermophilus* (1V19); RKK_Ec: Ribokinase, *Escherichia coli* (1RKD); FRK_Pm: FRK, *Petrotoga mobilis* SJ95; FRK_Pd: FRK, *Polaribacter dokdonensis*; FRK_Cs: FRK, *Cellulophaga* sp MED134; FRK_Pt: FRK, *Psychroflexus torquis* ATCC 700755; FRK_Sl: FRK, *Solanum lycopersicum*; FRK_At: FRK, *Arabidopsis thaliana*; FRK_Zm: FRK, *Zea mays*; FRK_Bv: FRK, *Beta vulgaris*.

3.3.3 Putative ATP binding pocket

Although we did not obtain the positions of ATP and fructose in the *Ho*FRK structure through crystallization, the binding site of ATP and fructose can be predicted by a comparison with the structures of three known ribokinase family members complexed with ATP/ADP. (Figure 3.8a and 3.8b) The overall architecture of the nucleotide binding site is very similar in *Ho*FRK, AIR kinase (Zhang *et al*, 2004), KDG kinase (Ohshima *et al.*, 2004) and ribokinase (Sigrell *et al*, 1998). The ATP/ADP is predicted to bind to a pocket on the catalytic domain lined up by residues Asp181-Cys183; Lys210-Asp215; Thr243-Gly248; Ala267-Gly274; Leu297; Asn299; Val301 and Phe304. Similarly the fructose binding pocket is predicted to be lined by residues Leu31, Asp33, Leu43, Gly54-Ser56, Asn59, Phe153 and Asp275. It is noteworthy that Leu31 and Asp33 are from the β-sheet lid region. Thr243, Gly274 and Asn299 are the three residues which are totally conserved among all of the analysed sequences. Furthermore, in the superimposed structures these residues are located in the ATP/ADP binding pocket. (Figure 3.8a and 3.8b) The equivalent residues of Thr243 and Gly274 in ribokinase, KDG and AIR kinases were found to interact with the phosphate group of the bound ATP mainly through water mediated hydrogen bonds, while Asn299 interacts with the adenosine base. These residues in *Ho*FRK probably play a role in forming the oxyanion hole to stabilize the intermediate during phosphorylation as they bind to both substrates.

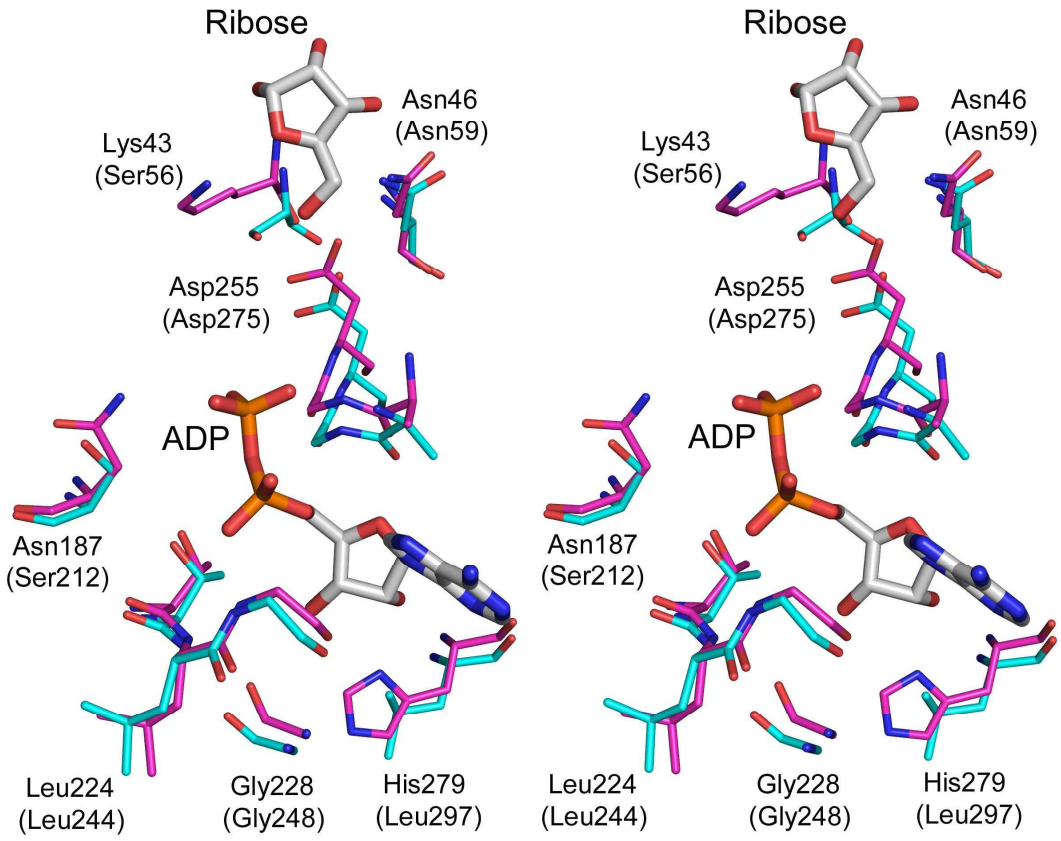
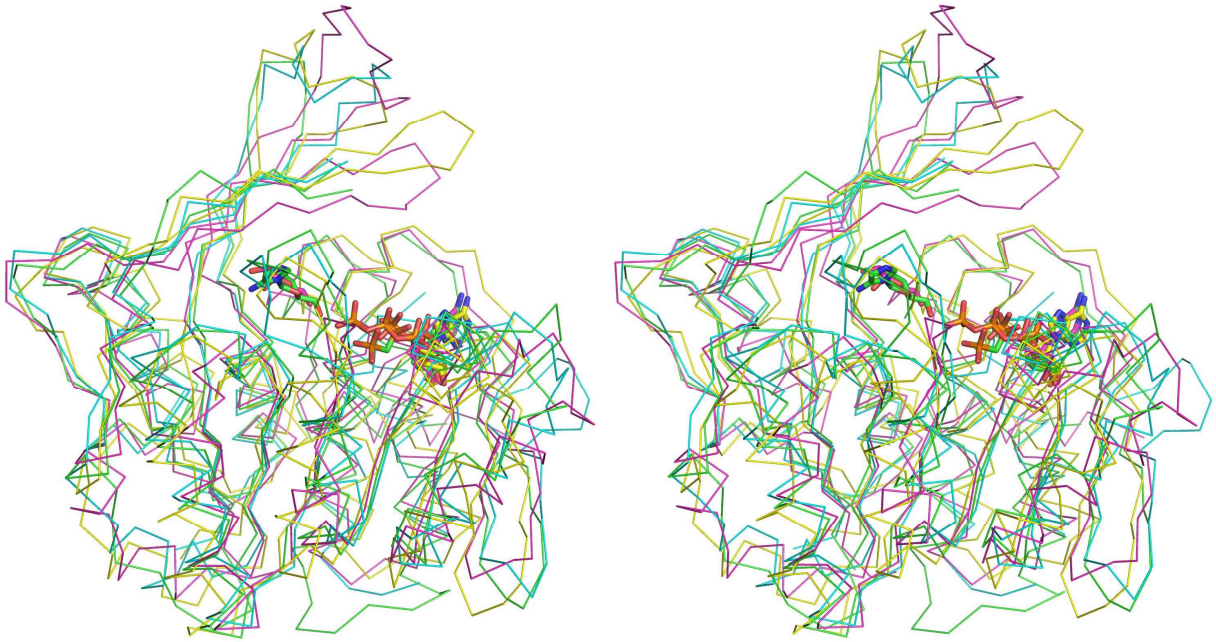


Figure 3.7 Stereo diagram of the conserved, binding residues of RK (magenta; PDB code 1RKD) interacting with both of its ligands ADP (white) and Ribose (white), with the corresponding and conserved residues of *Ho*FRK (cyan) superimposed. For figure clarity, only eleven of the conserved, key residues of RK and *Ho*FRK are shown. Both labelled Leu244 and Asp275 of *Ho*FRK represents Gly243-Gly245 and Gly273-Asp275 respectively shown in the diagram. Residues of *Ho*FRK are labelled in bracket. This figure was prepared using the program PyMOL (DeLano, 2002).



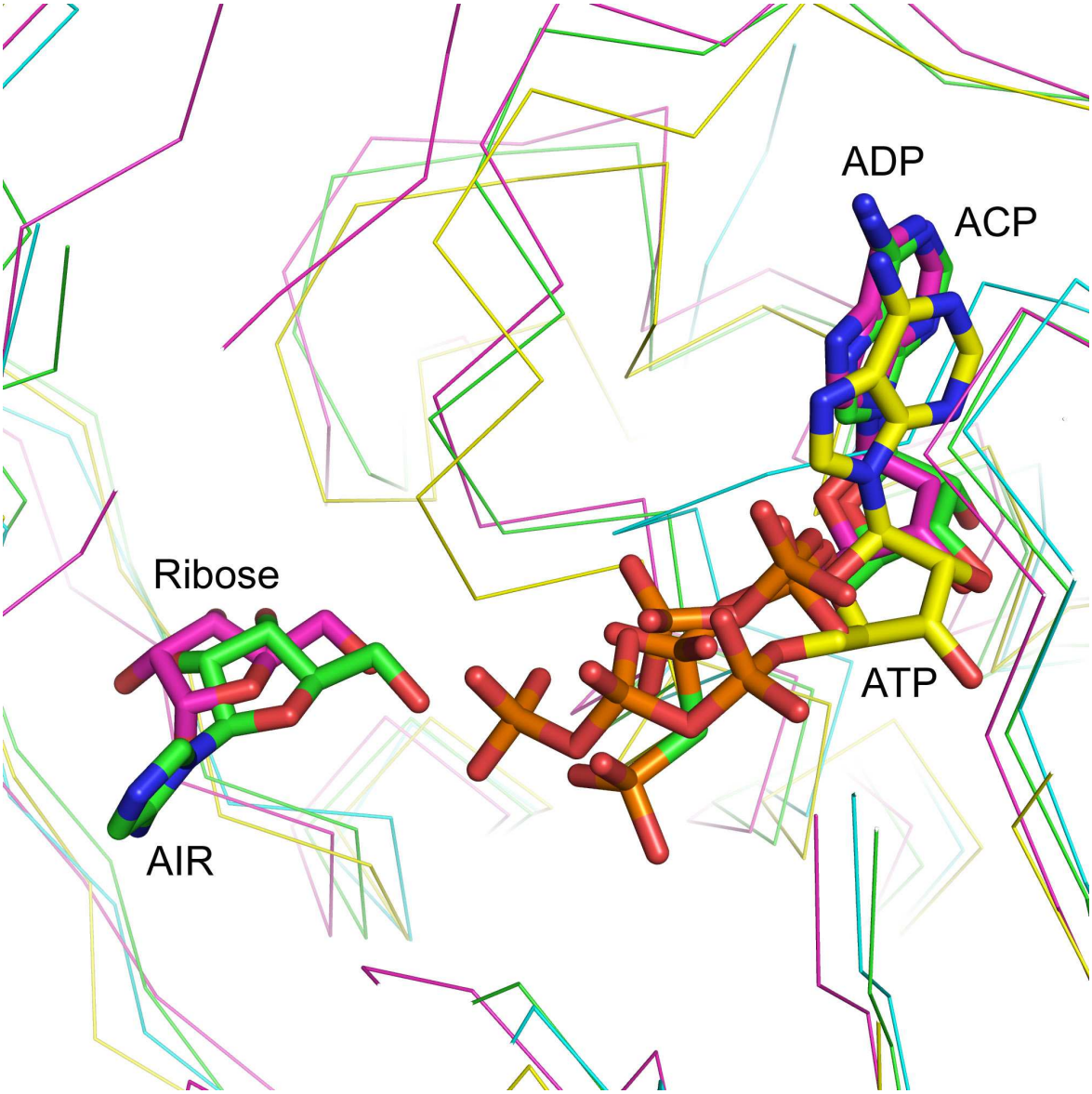


Figure 3.8 (a) Stereo diagram of *Ho*FRK (cyan) and the complex structures of the three ribokinase family members, superimposed on the *Ho*FRK model at the catalytic domain. Coloured lines represent the Cα trace of AIR kinase (1TZ6, green), RDK kinase (1V1B, yellow) and RK (1RKD, magenta). Substrates in the various structures are represented by the stick models. (b) Close-up view of the substrates binding sites of the superimposed model of *Ho*FRK (cyan) and the complex structures of the three

ribokinase family members at the catalytic domain. Substrates in the various structures are represented by the stick models. The colour of the substrates of AIR kinase, RDK kinase and RK were green, yellow and magenta respectively. These figures were produced using PyMOL (DeLano, 2002).

The *Ho*FRK structure revealed a possible movement of the loop region between Ile260 and Val270 to accommodate the incoming substrates. In the present study, the apo *Ho*FRK loop occludes a part of the substrate binding pocket (Figure 3.9). Based on the structural comparisons with the ATP ribokinase complexes and the sequence analysis, both the ATP and fructose interacting residues are established. All the substrate binding residues are located in the well defined binding pocket and these residues are highly conserved (Figure 3.7, 3.8a and 3.8b). The occluding loop however, is not well defined in the electron density map, when compared with the rest of the molecule. It indicated that the binding of the incoming substrate will move and stabilize this loop region.

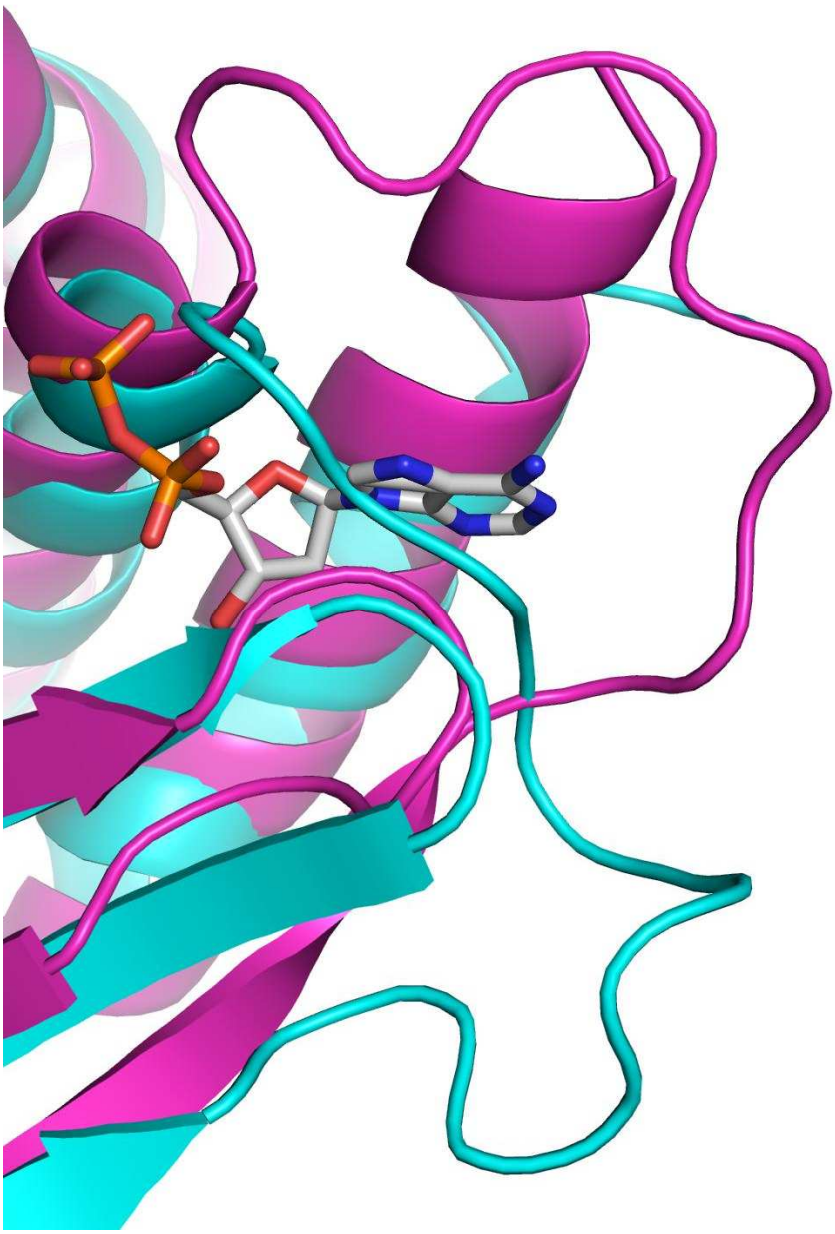


Figure 3.9 Superimposed ribbon diagram of *Ho*FRK (cyan) and RK-ADP (magenta-white) complex structure. The diagram shows a loop region (Ile260-Val270) of *Ho*FRK which occludes a part of the ATP/ADP binding pocket. There is a possible movement of this loop region to open and accommodate the incoming substrates.

3.3.4 Proposed mechanism of action

AIR (Zhang *et al*., 2004), KDG kinases (Ohshima *et al.*, 2004) and ribokinase (RK) are structurally similar to *Ho*FRK with fully formed β-sheet lids and catalytic domains. Figure 3.8a and 3.8b show the superimposition of these homologs with bound substrates. All the substrate binds in the same cleft region. It is clearly evident that RK lid region is found in the closed form and interacting with the substrates. However the lid in RK is the most complex and comprises of a total of 50 residues, which is about 5-10 amino acids more than its structural homologs. The β-sheet lid of *Ho*FRK is a total of 42 residues, comprising of four β-strands and a long loop between residues Ser36 and Glu48. Of these, six residues (Lys115-Pro120) were not observed in the electron density map. Residues Leu31 to Gly55 are from the extreme N-terminal and Gly109 to Glu127 from the middle of the sequence. In comparison to RK, the *Ho*FRK lid region represents a more open form. -As the ribose substrate of RK is structurally most similar to fructose of FRK, predictions based on the structure of RK ribose complex was speculated to be applicable to interpreting FRK substrate binding mechanism. Structure based sequence alignment of FRK with RK showed that of the active site residues, Asp33 (Asp16 in RK), Phe53 (Phe40 in RK) and Gly55 (Gly42 in RK) were highly conserved. Theses residues in RK interact with the hydroxyl groups in ribose sugar through direct and water-mediated hydrogen bonds.

For all members of the ribokinase superfamily, the substrate binding site is surrounded by two conserved N-terminal α-helices and C-terminal β-strands. One of the signature motifs of this family is a Gly-Gly dipeptide. It was found to undergo a conformational switch upon substrate binding to bring the enzyme from open to closed

state relative to the lid region for substrate sequesteration (Schumacher *et al*, 2000). The GG dipeptide is present in all structures where the lid region connects to the catalytic domain. In *Ho*FRK, this Gly-Gly motif is found at position of residues 54-55 and possibly act as a hinge like the others. Gly42 in RK and possibly its equivalent in *Ho*FRK may also interact with a hydroxyl group on its sugar substrate.

The closure of the lid about Gly54-Gly55 is followed by the formation of the anion hole, induced by an essential K^+ ion to activate the enzyme. This K^+ ion is coordinated by the carbonyl oxygen of Asp246, Thr248, Ala287, Ala290 and Gly292 in AIR kinase. Based on the structure alignment, the K^+ ion coordination residues in *Ho*FRK are predicted to be Asp269, Thr271, Ala303, and Ile306 respectively. These residues are found conserved among the homologs. Unfortunately, similar to RK no electron density can be assigned as K^+ ion in *Ho*FRK. Adjacent to the metal ion binding region, the GAGD motif forms anion hole through its main chain nitrogen atoms. This motif is the most highly conserved region among members in the ribokinase superfamily. In *Ho*FRK, this motif is found between Gly272 and Asp275. Figure 3.10 shows the simulated-annealing *Fo-Fc* omit map of part of these residues including Asp275 in FRK. This anion hole helps to neutralize the accumulated negative charge during the phosphorylation of the substrate.

The last residue of the GAGD motif, Asp275 of *Ho*FRK, is found near to the phosphate acceptor hydroxyl group of the sugar ring (Figure 3.11a and 3.11b). In homologs structures, it is proposed to act as a general base to extract the proton from the 5' hydroxyl-group during nucleophilic attack in the first step of the phosphotransferase reaction and we speculate that this may be applicable for FRKs (Matthews *et al*, 1998;

Schumacher *et al*, 2000; Sigrell, 2000). Subsequently, the negatively charged hydroxyl group performs nucleophilic attack on the γ-phosphate group of bound ATP. The intermediate formed is stabilized by the anion hole. This intermediate eventually decomposes into ADP and fructose-6-phosphate by a S_N2 reaction.

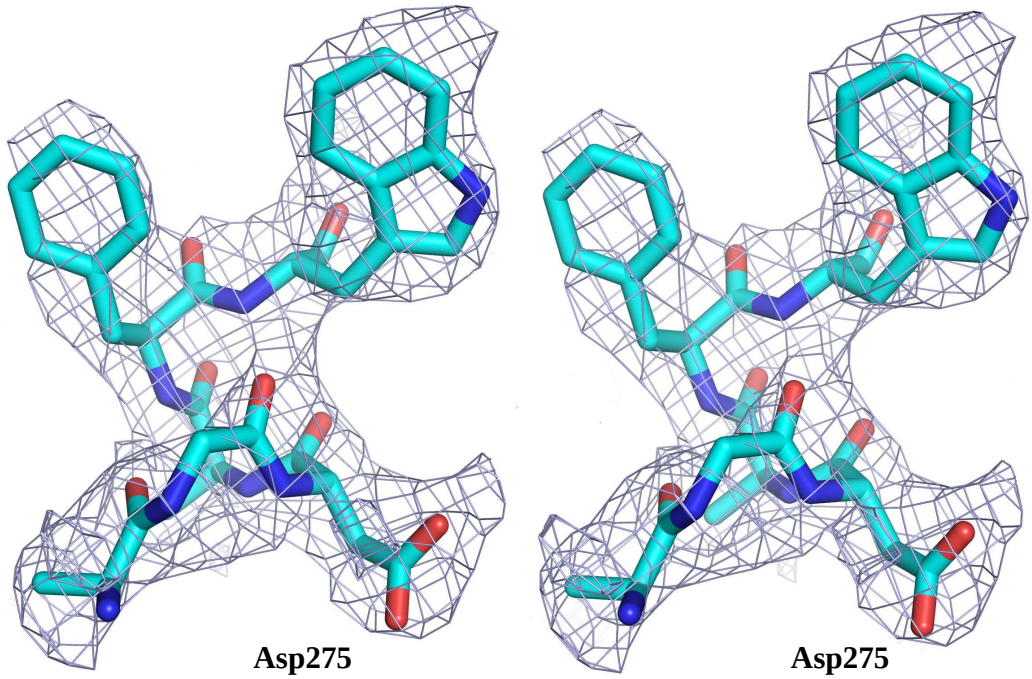
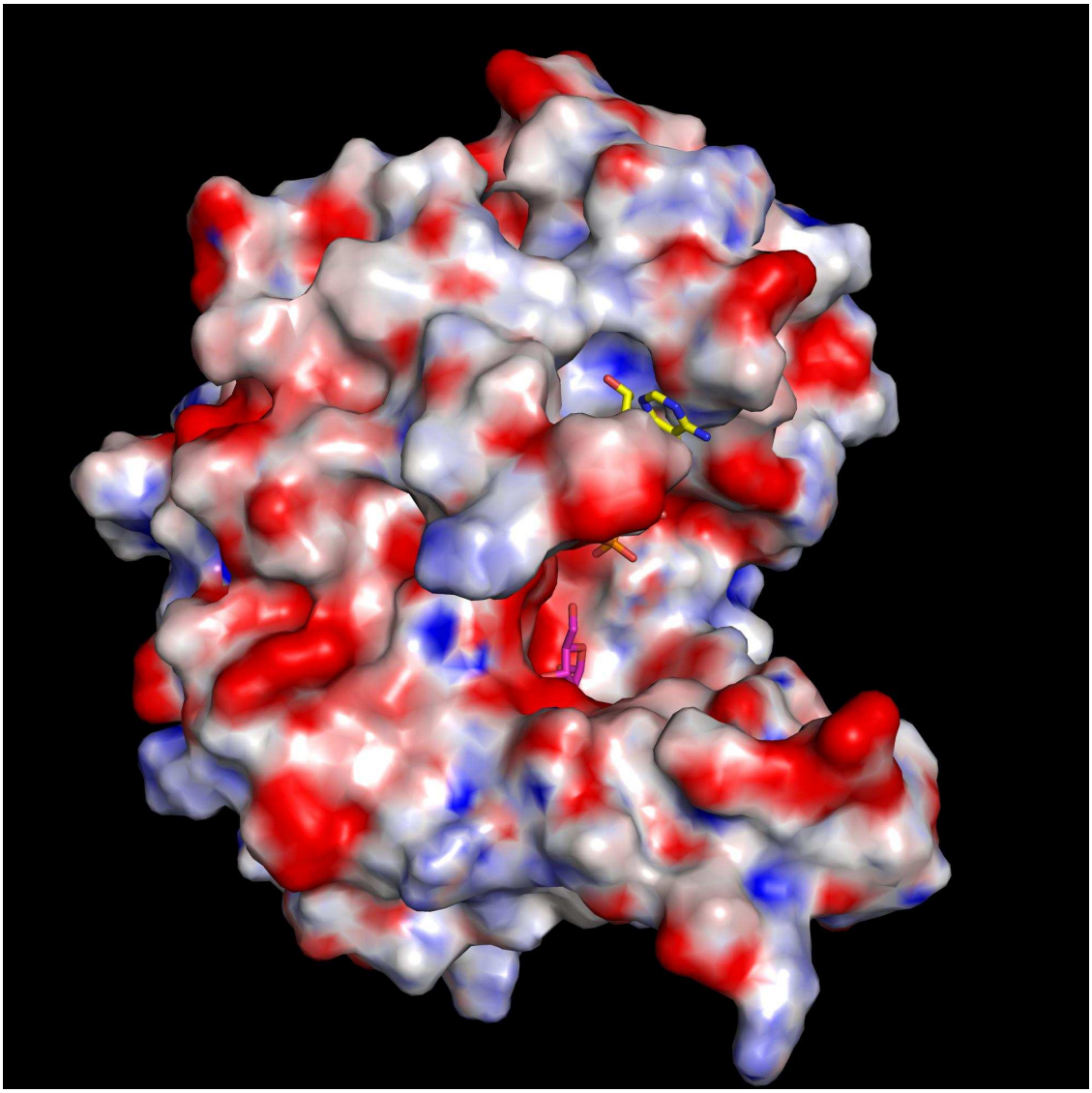


Figure 3.10 Stereo diagram of simulated-annealing *Fo-Fc* omit map of residues in *Ho*FRK. The map contoured at a level of 2.0σ. All atoms within 2 Å of Ala273 to Trp278 were omitted prior to refinement and map calculation. Atoms are shown in cyan (C), blue (N), red (O) and orange (P). This figure was prepared using the program PyMOL (DeLano, 2002).



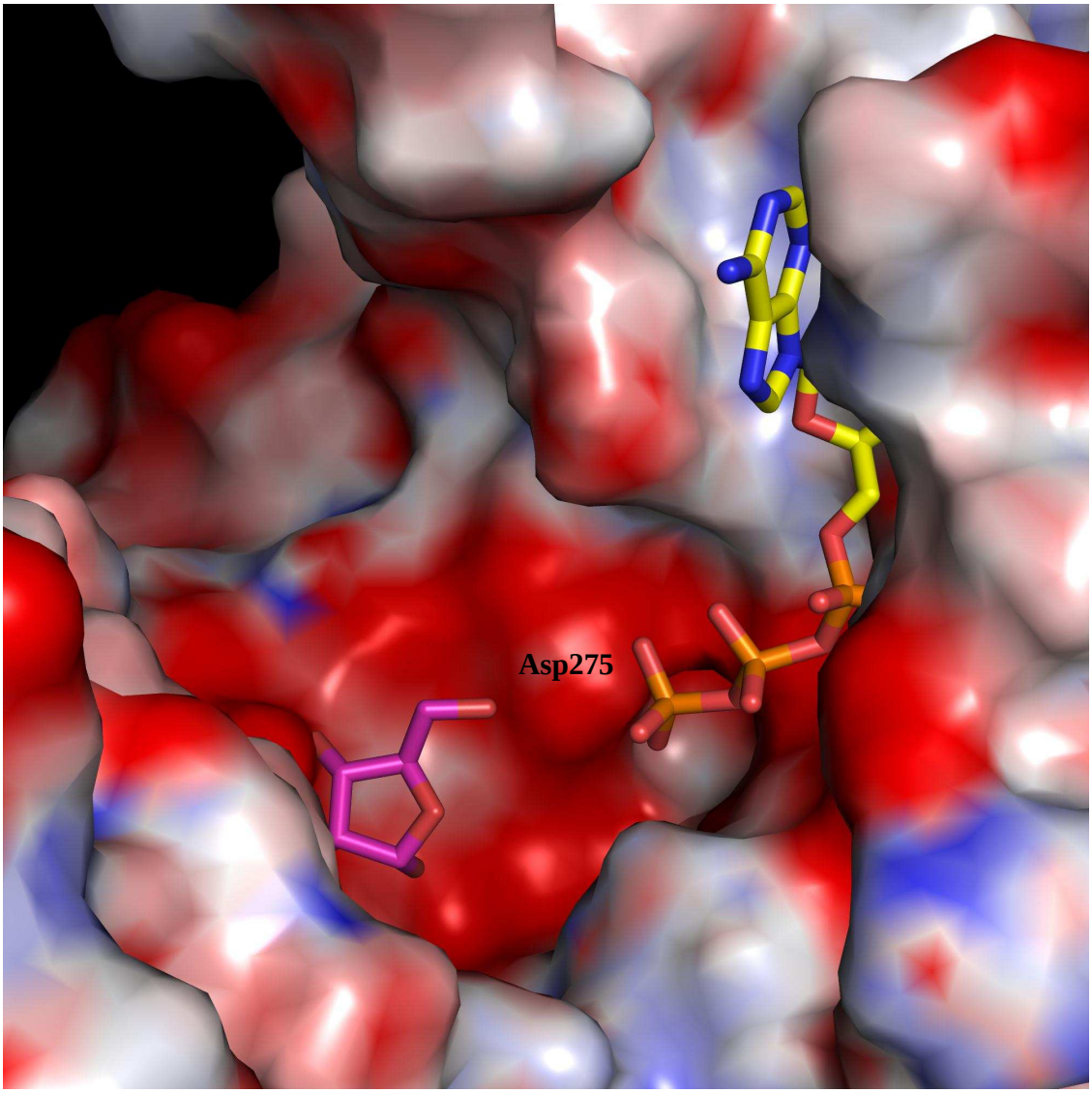


Figure 3.11 (a) Molecular surface of *Ho*FRK showing the distinct domain and lid structural features separated by a large substrate binding cleft. The inferred fructose (ribose) and ATP molecule is shown in the cleft region. (b) Close-up view of the inferred fructose (ribose) and ATP binding site. Atoms are shown white (C), red (O) and orange (P). These figures were produced using PyMOL (DeLano, 2002) using electrostatic potential generated by APBS (Baker *et al.*, 2001).

Chapter IV

Conclusions and Future Directions

The crystal structures of SPS and FRK from *H. orenii* were determined and have been thoroughly described in this thesis. In addition, their mechanisms of action were proposed based on these structures combined with bioinformatics analyses, ITC data, enzyme-substrate/product complexes in the case of SPS and inferred complex models in the case of FRK. The elucidation of their structures and mechanisms are significant in these family of enzymes. These structures are the first unique structures of their respective enzymes to be characterized structurally. SPS and FRK from the plant source was shown to be very difficult for purification and characterization. In order to understand the mechanism of the plant enzyme, a closest homolog was taken from the bacterial system. *H.orenii* SPS and FRK exhibits close sequence homology with their plant counterparts. Thus our findings on the structure and mechanism can be easily extended to describe plant SPS and FRK enzymes. Present demonstrations on *H. orenii* enzymes represent valid models for their plant homologs.

The availability of both apo- and complexed SPS structures contribute invaluable insight to its catalytic mechanism. It is the first enzyme of its family to be structurally characterized as apo as well as with a bound substrate/product. Our study uncovered the importance of His151 for its role in domain closure and the transferring the glucose moiety of UDP Glu from B-domain to A-domain.

SPS has been implicated in food productivity and stress response. As a continuation of this project, in the future, we will determine the structure of the complex with both substrates. In addition, the structure based mutagenesis will be performed on its catalytic site to select for transgenic high yielding crops with a greater resistance to osmotic fluctuations. We are also interested to study the full length plant SPS structure

comprising of the N-terminal domain, catalytic domain (similar to *Ho*SPS, present work) and the SPP like domain will also be determined, as well as their independent domains to widen our understanding of the catalytic mechanism of SPS in plants

This thesis also reported the structure and the proposed mechanism of FRK for the first time. The crystallization of HoFRK is the most challenging part of this project. Although we have attempted to determine the crystal structres of the complexes, no enzyme-substrate complex was trapped in the crystal. However, a comparative study with other members of the ribokinase family demonstrated that FRK adopts a similar mechanism as the other members of this group using the highly conserved GAGD motif which forms an anion hole during catalysis. Furthermore, Asp275 acts to remove a proton from fructose and the fuuctose hydroxyl-FRK intermediate attacks ATP nucleophillically and decomposes into ADP and F6P.

For future studies, the next immediate procedure would be therefore to confirm these propositions by obtaining the crystal structures of FRK in complex with the substrates ATP and/or fructose. These studies will enhance our understanding on the mechanism of action of FRK. Further this will confirm or refute the abovementioned conclusions drawn based on the comparative study. This will conclude FRK mechanism in plants.

References

Altschul, SF, Madden, TL, Schaffer, AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of SS and its potential role in synthesis of cellulose and callose in plants. Proc Natl Acad Sci USA. 92:9353-9357.

Baker NA, Sept D, Joseph S, Holst MJ and McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA. 98:10037-10041.

Baysdorfer DF, Kremer and RC Sicher (1989) Partial purification and characterization of fructokinase activity from barley leaves. J. Plant Physiol. 134:156–161.

Bradford, MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-254.

Breton C, Snajdrova L, Jeanneau C, Koca J and Imberty A (2006) Structures and mechanisms of glycosyltransferases. Glycobiology. 16:29R-37R.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse- Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T and Warren GL (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Crystallogr D. 54:905-921.

Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA and Alzari PM (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. EMBO J. 23:3196-3205.

Campobasso N, Matthews II, Begley TP, Ealick SE (2000) Crystal structure of 4-methyl-5-b-hydroxyethylthiazole kinase from Bacillus subtilis at 1.5A resolution. Biochemistry 39:7868-7877.

Carlson SJ, Chourey PS (1996) Evidence for plasma membraneassociated forms of sucrose synthase in maize. Mol Gen Genet. 252:303-310.

Castleden CK, Aoki N, Gillespie VJ, MacRae EA, Quick WP, Buchner P, Foyer CH, Furbank RT, and Lunn JE (2004) Evolution and Function of the Sucrose-Phosphate Synthase Gene Families in Wheat and Other Grasses. Plant Physiol. 135:1753–1764.

Cayol JL, Ollivier B, Patel BK, Prensier G, Guezennec J and Garcia JL (1994) Isolation and characterization of Halothermothrix orenii gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. Int J Syst Bacteriol. 44:534-540.

Chaubron F, Harris N, Ross HA, Davies HV (1995) Partial purification and characterization of fructokinase from developing taproots of sugar beets (Beta vulgaris). Plant Sci. 110:181-186.

Cheng G, Bennett EM, Begley TP, Ealick SE (2002) Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase at 2.3A resolution. Structure 10:225-235.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG and Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31:3497-3500.

Cid E, Gomis RR, Geremia RA, Guinovart JJ and Ferrer JC (2000) Identification of Two Essential Glutamic Acid Residues in Glycogen Synthase. J Biol Chem. 43:33614-33621.

Coutinho PM and Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. "Recent Advances in Carbohydrate Bioengineering", H.J. Gilbert, G. Davies, B. Henrissat and B. Svensson eds., The Royal Society of Chemistry, Cambridge, pp. 3-12.

Cumino A, Curatti L, Giarrocco L, and Salerno GL (2002) Sucrose metabolism: Anabaena sucrose-phosphate synthase and sucrose-phosphate phosphatase define minimal functional domains shuffled during evolution. FEBS Lett. 517:19-23.

Curatti L, Folco E, Desplats P, Abratti G, Limones V, Herrera-Estrella L and Salerno G (1998) Sucrose-Phosphate Synthase from Synechocystis sp. Strain PCC 6803: Identification of the spsA Gene and Characterization of the Enzyme Expressed in Escherichia coli. J Bacteriol. 180:6776-6779.

Dai N, German MA, Matsevitz T, Hanel R, Swartzberg D, Yeselson Y, Petreikov M, Schaffer AA, Granot D (2002) LeFRK2, the gene encoding the major fructokinase in tomato fruits is not required for starch biosynthesis in developing fruits. Plant Sci. 162:423-430.

Davies HV, Shepherd LVT, Burrell MM, Carrari F, Urbanczyk-Wochniak E, Leisse A, Hancock RD, Taylor M, Viola R, Ross H, McRae D, Willmitzer L, Fernie AR (2005) Modulation of Fructokinase Activity of Potato (Solanum tuberosum) Results in Substantial Shifts in Tuber Metabolism. Plant Cell Physiol. 46:1103-1115.

DeLano WL (2002) The PyMOL Molecular Graphics System. Delano Scientific, San Carlos, CA, USA. http://www.pymol.org

Doiron B, Cuif MH, Chen R, Kahn A (1996) Transcriptional glucose signaling through the glucose response element is mediated by the pentose phosphate pathway. J Biol Chem. 271:5321–5324.

Etxeberria E and Gonzalez P (2003) Evidence for a tonoplast-associated form of sucrose synthase and its potential involvement in sucrose mobilization from the vacuole. J Exp Bot. 54:1407-1414.

Fieulaine S, Lunn JE, Borel F and Ferrer JL (2005) The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. Plant Cell 17:2049-2058.

Frickey T and Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20:3702-3704.

Gancedo JM (1998) Yeast carbon catabolite repression. Microbiol Mol Biol. 62:334–361.

Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R and Davies GJ (2002) Insights Into Trehalose Synthesis Provided by the Structure of the Retaining Glycosyltransferase OtsA. Chem Biol. 9:1337-1346.

Gibson RP, Tarling CA, Roberts S, Withers SG and Davies GJ (2004) The Donor Subsite of Trehalose-6-phosphate Synthase. J Biol Chem. 279:1950-1955.

Gouet P, Courcelle E, Stuart DI and Metoz F (1999) ESPript: multiple sequence alignments in PostScript. Bioinformatics 15:305-308.

Greenway H, Munns R (1980) Mechanisms of salt tolerance in nonhalophytes. Annu Rev Plant Physiol. 31:149-190.

Hite D, Outlaw WH Jr. and Tarczynski MC (1993) Elevated levels of both sucrose-phosphate synthase and sucrose synthase in Vicia guard cells indicate cell-specific carbohydrate interconversions. Plant Physiol. 101:1217-1221.

Holm L and Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol. 233:123-138.

Horcajada C, Guinovart JJ, Fita I and Ferrer JC (2006) Crystal structure of an archaeal glycogen synthase: insights into oligomerization and substrate binding of eukaryotic glycogen synthases. J Biol Chem. 281:2923-2931.

Hubac C, Vieira Da Silva (1980) Indicateurs metaboliques de contraintes mesologique. Physiol Veg. 18:45-53.

Huber JL, Huber SC, and Nielsen TH (1989) Protein phosphorylation as a mechanism for regulation of spinach leaf sucrose-phosphate synthase activity. Arch Biochem Biophys. 270:681-690.

Huynh F, Tan TC, Swaminathan K, and Patel BK (2005) Expression, purification and preliminary crystallographic analysis of sucrose phosphate synthase (SPS) from Halothermothrix orenii. Acta Crystallogr F. 61:116-117.

Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H, Wakagi T (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyper thermophilic archeon. Structure 9:205-214.

Jones TA, Zou JY, Cowan SW and Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A. 47:110-119.

Kanayama Y, Dai N, Granot D, Petreikov M, Schaffer A, Bennett A (1997) Divergent fructokinase genes are differientially expressed in tomato. Plant Physiol. 113:1379-1384.

Kanayama Y, Granot D, Dai N, Petreikov M, Schaffer A, Bennett AB (1998) Tomato fructokinases exhibit differential expression and substrate regulation. Plant Physiol. 117:85–90.

Kaytor EN, Shih HM, Towle HC (1997) Carbohydrate regulation of hepatic gene expression. Evidence against a role for the upstream stimulatory factor. J Biol Chem. 272:7525–7531.

Keller F, Frehner M, Wiemken A (1988) Sucrose synthase, a cytosolic enzyme in protoplasts of Jerusalem artichoke tubers (*Helianthus tuberosus* L.). Plant Physiol. 88:239-241.

Koch KE (1996) Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol. 47:509–540.

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins. 37:228-41.

Kraulis PJ (1991) Molscript -- a program to produce both detailed and schematic plots of protein structures. J Appl Cryst. 24:946-950.

Kumar S, Tamura K and Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. 5:150-163.

Kurowski, M.A. and Bujnicki, J.M. (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res. 31:3305-3307.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) ClustalW2 and ClustalX version 2. Bioinformatics 23:2947-2948.

Laskowski RA, McArthur, MW, Moss DS and Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst. 26, 282–291.

Lee MH, Yang CC, Wang HL, and Lee PD (2003) Regulation of sucrose phosphate synthase of the sweet potato callus is related to illumination and osmotic stress. Bot Bull Acad Sin. 44, 257-265.

Levitt J (1980) responses of plants to environmental stress vols I, II. Academic Press, London New York.

Li MH, Kwok F, Chang WR, Lau CK, Zhang JP, Lo SCL, Jiang T, Liang DC (2002) Crystal structure of brain pyridoxyl kinase, a novel member of the ribokinase superfamily. J Bio Chem. 277:46385-46390.

Liu J and Mushegian A (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci. 12:1418-1431.

Lunn JE (2002) Evolution of sucrose synthesis. Plant Physiol. 128:1490-1500.

Lunn JE and MacRae E (2003) New complexities in the synthesis of sucrose. Curr Opin Plant Biol. 6:208–214.

Lunn JE, Price GD and Furbank RT (1999) Cloning and expression of a prokaryotic sucrose-phosphate synthase gene from the cyanobacterium Synechocystis sp. PCC 6803. Plant Mol Biol. 40:297-305.

MacGregor EA (2002) Possible structure and active site residues of starch, glycogen, and sucrose synthases. J Prot Chem. 21:297-306.

Martinez-Barajas E, Luethy MH, Randall DD (1997) Molecular cloning and analysis of fructokinase expression in tomato (Lycopersicon esculentum Mill.) Plant Sci. 125:13-20.

Matthews II, Erion MD, Ealick SE (1998) Structure of human adenosine kinase. Biochemistry 37:15607-15620.

Merritt EA and Bacon DJ (1997) Raster3D Version 2: photorealistic molecular graphics. Methods of Enzymology. 277:505-524.

Mijts BN, and Patel BK (2001) Random sequence analysis of genomic DNA of an anaerobic, thermophilic, halophilic bacterium, Halothermothrix orenii. Extremophiles 5:61-69.

Muller W, Wegmann K (1978) Sucrose biosynthesis in Dunaliella I. Isolation and properties of sucrose phosphate synthetase. Planta 141:159-163.

Munn R, Brady CJ, Barlow EWR (1979) Solute accumulation in the apex and leaves of wheat during water stress. Aust J Plant Physiol. 6:379-389.

Nicholls A, Sharp KA and Honig B (1991) Protein folding and association - insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct Funct Genet. 11:281-296.

Odanaka S, Bennett AB, Kanayama Y (2002) Distinct physiological roles of fructokinase isozymes revealed by gene-specific suppression of FRK1 and FRK2 expression in tomato. Plant Physiol. 129:1119-1126.

Ohshima N, Inagaki E, Yasuike K, Takio K, Tahirov TH (2004) Structure of Thermus thermophilus 2-keto-3-deoxygluconate kinase: evidence for recognition of an open chain substrate. J Mol Biol. 340:477-489.

Otwinowski Z and Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276:307–326.

Pego JV, Kortstee AJ, Huijser C, Smeekens SCM (2000) Photosynthesis, sugars and the regulation of gene expression. J Exp Bot. 51:407–416.

Pego JV and Smeekens SCM (2000) Plant fructokinases: a sweet family get together. Trends Plant Sci 5:531-536.

Porchia AC, Curratti L and Salerno GL (1999) Sucrose metabolism in cyanobacteria: sucrose synthase from Anabaena sp. strain PCC 7119 is remarkably different from the plant enzymes with respect to substrate affinity and amino-terminal sequence. Planta 210:34-40.

Porchia AC and Salerno GL (1996) Sucrose biosynthesis in a prokaryotic organism: Presence of two sucrose-phosphate synthases in Anabaena with remarkable differences compared with the plant enzymes. Proc Natl Acad Sci USA. 93:13600-13604.

Renz A Stitt M (1993) Substrate specificity and product inhibition of different forms of fructokinases and hexokinases in developing potato tubers. Planta 190:166–175.

Ronne H (1995) Glucose repression in fungi. Trends Genet 11:12–17.

Schaffer AA, Petreikov M (1997) Inhibition of fructokinase and sucrose synthase by cytosolic levels of fructose in young tomato fruit undergoing transient starch synthesis. Physiol Plant 101:800-806.

Schnarrenberger C (1990) Characterization and compartmentation in green leaves, of hexokinases of different specificities of glucose, fructose, and mannose and for nucleoside triphosphates. Planta 181:249–255.

Schumacher MA, Scott DM, Matthews II, Ealick SE, Roos DS, Ullman B,Brennan RG (2000) Crystal structures of Toxoplasma gondii adenosine kinase reveal a novel catalytic mechanism ad prodrug binding. J Mol Biol. 298:875-893.

Sigrell JA, Cameron AD, Jones TA, Mowbray SL (1998) Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8A resolution: insights into a new family of kinase structures. Structure 6:183-193.

Terwilliger TC and Berendzen J (1999) Automated MAD and MIR structure solution. Acta Crystallogr D. 55:849-861.

Terwilliger TC (2000) Maximum likelihood density modification. Acta Crystallogr D. 56:965-972.

Ullman CG and Perkins SJ (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J. 326:929-942.

Wegmann K (1979) Biochemical adaptation of Dunaliellia tertiolecta to salinity and temperature changes. Ber Dtsch Bot Ges. 92:43-54.

Wegmann K, Ben Amotz A, Avron M (1980) Effect of temperature on glycerol retention in halotolerant algae Dunaliellia and Asteromonas. Plant Physiol. 66:1196-1197.

Wrabl JO and Grishin NV (2001) Homology Between O-linked GlcNAc Transferases and Proteins of the Glycogen Phosphorylase Superfamily. J Mol Biol. 314:365-374.

Zhang SR, Nichols SE, Dong JG (2003) Cloning and characterization of two fructokinases from maize. Plant Sci. 165:1051-1058.

Zhang Y, Dougherty M, Downs DM, Ealick SE (2004) Crystal structure of an aminoimidazole riboside kinase from Salmonella enteric: implications for evolution of the ribokinase superfamily. Structure 12:1809-1821.